Journal of Scientometric Research, 2026; 15(3):1017-1031.
https://www.jscires.org

Research Article

Emerging Research Trends in Large Language Models with a Focus on Lung Cancer: Bibliometric Analysis

Mukesh Kumar[1,2], Rammohan Mallipeddi[1], Pradeep Reddy Challa[3], Vikasdeep Singh Mann[4]

[1]Department of Artificial Intelligence, School of Electronics Engineering, Kyungpook National University, Daegu, REPUBLIC OF KOREA.
[2]Advanced Centre of Research & Innovation (ACRI), School of Advance Computing (SCA), CGC University, Mohali, Punjab, INDIA.
[3]GRT Institute of Engineering and Technology, Mahalakshmi Nagar, Srinivasapuram, Perungalathur, Tamil Nadu, INDIA.
[4]Department of Mechanical Engineering, School of Engineering and Technology, CGC University, Mohali, Punjab, INDIA.

ABSTRACT

Large Language Models (LLMs) are advancing the use of Artificial Intelligence (AI) in healthcare and are at the forefront of technology use for medical diagnosis and clinical decision support. With lung cancer being one of the leading causes of cancer deaths worldwide and an area of major growth for the use of LLM technology, the use of LLMs in radiology and pathology report interpretation, clinical text mining, and automating clinical decision support is increasing rapidly. Although there have been increasing numbers of empirical research and review articles examining LLM use in lung cancer, there have not been any comprehensive bibliometric assessments examining how this emerging area of research has evolved. To address this gap in knowledge, this study conducts a systematic bibliometric analysis to evaluate the global research trends, intellectual structure, and patterns of collaboration of the use of LLM technology in lung cancer research. The study analysed Scopus-indexed publications through the year 2026 using VOSviewer by evaluating the growth of publications, the type of documents published, the most influential journals, the most productive authors, the most productive institutions, the contributing countries, the impact of citations, and keyword co-occurrence networks. Bibliometric analyses revealed that after 2023, there have been considerable publications contributing to LLM use in lung cancer due to the recent clinical applications of new and advanced Generative Pre-Trained Transformers (GPTs) and Vision-Language Models (VLMs). Themes consistently emerging from quantitative analyses of the research highlight that LLMs are being utilized in medical diagnosis, radiology and pathology reporting, natural language processing, and clinical decision support systems. The bibliometric analysis also demonstrates that the United States and China are the leading countries publishing on this topic, with other countries contributing also, and that the leading journals of publication for research are interdisciplinary journals in digital medicine and biomedical informatics. Overall, the results of the bibliometric analysis provide important information about the future direction of LLM-related research in lung cancer, the emerging hotspots of research, and the network of researchers collaborating on LLM use within the context of lung cancer treatment and care.

Keywords: Bibliometric Analysis, Large Language Models, Lung Cancer, Medical Diagnosis, Natural Language Processing, Scientometric Analysis.

Correspondence:
Mukesh Kumar
Advanced Centre of Research & Innovation (ACRI) and School of Advance Computing, CGC University, Mohali, Punjab, INDIA.
Email: mukeshthakur1021982@gmail.com

Received: 04-06-2026;
Revised: 28-07-2026;
Accepted: 16-09-2026.

INTRODUCTION

Worldwide, Lung Cancer is one of the most diagnosed and deadly forms of cancer, and it accounts for a large part of the global cancer burden in terms of morbidity and mortality. Despite improvements made in imaging, pathology, and treatment options, lung cancer remains one of the hardest types of cancer to find at an early stage and diagnose accurately. AI has gained traction within oncology in recent years, taking on more roles within the oncology workflow, making it easier to improve diagnostic accuracy, automate clinical documentation, and support ongoing clinical decision-making processes. One example of AI is the development of LLMs that are capable of processing and interpreting vast amounts of unstructured medical text including radiology reports, pathology reports, and medical records (Shi *et al*., 2025; Tripathi *et al*., 2025; Liu *et al*., 2026; Güneş *et al*., 2025). While there are many AI techniques available to assist with these types of tasks, large language models are an emerging (highly transformational) technology that can be applied to many aspects of healthcare-related applications due to their ability to process and interpret large amounts of unstructured data. Transformer-based architecture language models (for example, GPT-3 model) have been used successfully in healthcare to

DOI: 10.5530/jscires.20260166

Copyright Information :
Copyright Author (s) 2026 Distributed under Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

accomplish many NPL-related tasks such as extracting clinical data, answering clinical questions, summarizing reports, and providing clinical decision support (Duan *et al.*, 2025; Fung *et al.*, 2025; Luo *et al.*, 2025; Ghorbian & Ghobaei-Arani, 2025; Qiu *et al.*, 2025). In lung cancer research, LLMs have been used to extract the smoking history from clinical notes (Luo *et al.*, 2025), assist in TNM staging using information obtained from radiology and pathology reports (Ismayilov *et al.*, 2025; Ruan *et al.*, 2025; Wiest *et al.*, 2025), Support Multidisciplinary Team (MDT) decision making processes (Liu *et al.*, 2026), and automate the interpretation and reporting of lung cancer screening exams using vision-language models (Nguyen *et al.*, 2025; Yang *et al.*, 2025). The use of LLMs for these applications demonstrates the emerging importance of using large language-based technologies to increase the efficiency of, and consistency of diagnostic and overall clinical workflow.

In recent years, the combination of LLM's and the multi-modal and vision language frameworks have gained significant traction, particularly in the context of lung cancer diagnosis and prognosis assessments. Various multi-center benchmarking studies have evaluated the practices of LLM's within the context of Clinical Decision Support Systems (CDSS) associated with Lung Cancer Screening (Duan *et al.*, 2025), while an alternative series of studies have highlighted the ability of LLM's to generate structured digital clinical information from unstructured pathology and radiology reports (Tripathi *et al.*, 2025; Ruan *et al.*, 2025; Hao *et al.*, 2025). Additionally, the systematic and narrative review processes have put forward several opportunities and challenges related to the use of LLM's within oncology, particularly regarding reliability, interpretability, and clinical safety (Yuan *et al.*, 2025; Yang *et al.*, 2025; Kaba, 2025). These recent developments illustrate the rapidly maturing methodologies that utilise LLM's as essential tools in lung cancer research. Although there has been considerable empirical literature published on LLM's, current published data focuses primarily on technical performance assessment, model benchmarking, and systematic reviews of application-defined models (Yuan *et al.*, 2025; Yang *et al.*, 2025; Chang *et al.*, 2025). Although all such publications provide valuable evidence regarding the feasibility and efficacy of LLM's, they do not address the broader landscape of scientific investigation, such as publication growth patterns, influential contributors to the field, collaboration networks between authors and thematic evolution of lung cancer research. While bibliometric and Scientometric analyses are established methods of exploring trends in scientific research, as well as the structure of knowledge and discovering new areas of research, there has been limited comprehensive bibliometric investigations of LLM's within lung cancer research thus far.

- To fill this gap in current research, this work is a bibliometric analysis of LLMs publications for lung cancer using the SCOPUS database as the primary source. The present study will have three main objectives:
- To analyse the historical growth of publications in this area.
- To identify the most prolific journals, authors, institutions, and countries publishing works in these fields.
- To analyse and identify dominant and developing themes of research using citation analysis and keyword co-occurrence networks.

In addition, this study will use VOSviewer software to visualise and perform a network analysis of the data created through this process, and will provide a systematic summarisation of the worldwide trends in research and collaboration in this area. The findings of this study are expected to provide insights to clinicians and researchers and will ultimately benefit have continued development around the use of AI for the diagnosis and clinical decision making in the treatment of lung cancer.

METHODOLOGY

Materials

According to standard bibliometric research procedures, Scopus is the database in which the data for this study were obtained, Scopus is recognised as providing many high-quality peer-reviewed journals across many different disciplines. Scopus' indexing of journals, conference proceedings and review articles in AI and medical research made it a good choice for this study. The extraction date of 30 January 2026 was chosen to capture the most current publications at the time of writing. The search was conducted in Scopus for the **TITLE-ABS-KEY** to capture studies contained in the title of the study, the abstract of the study, and in author supplied keywords. "Large Language Model" OR "LLM" AND "Lung Cancer" were used to perform the search. No limitations were placed on the year of publication, type of document or language to provide for a broad spectrum of studies from the area of interest. All documents recorded in Scopus which met the search criteria were included in the primary dataset. Table 1 shows search strategy and data selection criteria for retrieving publications on large language models and lung cancer from the Scopus database

The inclusion criterion covered all articles pertaining to Large Language Models and Lung Cancer in the Scopus database if the articles were not retracted; therefore, retracted articles were removed to maintain dataset integrity and reliability. After applying inclusion and exclusion criteria, there were 346 documents available for library metrics and network analysis.

Methods

The bibliometric analysis of this study was done using a free software called VOSviewer (version 1.6.20) commonly utilized for drawing and visualizing bibliometric networks. VOSviewer can create networks based on citation, co-authoring, keywords that are co-occurring, and bibliographic coupling, therefore, these networks lead to an in-depth exploration of the intellectual structure of research in any scientific field and the trends in research of that field.

The current study has two main elements of bibliometric analysis: descriptive analysis and network analysis. The descriptive analysis gives an overview of the fundamental features of retrieved publications, such as the trend of annual publication, type of document, title of source, and number of citations. These measures provide information regarding the growth and maturity of research in large language models and lung cancer. The network analysis portion of this study pertains to the relationships and interactions between various bibliometric units. Citation analysis demonstrated what documents and authors are considered influential in the field of study based on how many times they are cited. Co-authorship analysis assessed collaboration patterns between authors, institutions, and countries therefore, illustrating the composition of research partnerships and collaborative networks. Keyword co-occurrence analysis using keywords generated by authors was completed to establish the identified major themes, hot spots, and evolution of themes related to the application of large language models for lung cancer. Finally, bibliographic coupling analysis was conducted to determine how similar documents and organizations were, based on shared references, therefore allowing for the identification of closely related streams of research and areas of institutional research focus.

RESULTS AND DISCUSSION

Scopus Database Analysis

The analysis of the Scopus database provides an overall, descriptive overview of the publication characteristics associated with large language model research in relation to lung cancer. The analysis is based on the data extracted from the Scopus database, obtained on January 30, 2026; there were a total of 346 documents identified and analysed. The analysis also examines key bibliometric indicators, such as the annual publication trends, types of documents published, and the most productive source titles for documenting the growth and structure of the research area, to provide examples of the historical development and the research momentum associated with using large language model applications for lung cancer. The analysis of the annual publication data provides evidence of the temporal development of the research and demonstrates the development trajectory of large language model applications for lung cancer; the analysis of the document type provides evidence of the nature of academic contributions and shows the extent to which the research area is primarily driven by original research, review articles or conference presentations; and the analysis of sources identifies the leading and most productive source titles and, thus, establishes the major sources and interdisciplinary characteristics for this research area.

Year wise Documents Published

Table 2 shows the yearly number of papers published regarding large-scale language models and lung cancer between 2015 and 2026, as recorded in the Scopus database. The data indicates that during the beginning of the study duration, there was an insignificant volume of research using language models in conjunction with lung cancer; with less than five papers being published each year from 2015 to 2018. Therefore, the initial phase represents research on language models/lung cancer at an earlier stage than later's phase of width for use. Guy, for example, continued to have a relatively small milestone during 2019-2021, indicating a rise in interest toward the use of natural language processing methods to analyse clinical text/data and for reported medical outcomes.

A substantial surge in publication output is evident from 2023 onwards, with 15 documents published in 2023, followed by a sharp increase to 122 publications in 2024 and 144 publications in 2025, marking the most productive years in the observed period. The advancing and ongoing use of superior quality large language models, including transformer architecture and models in a series of general-purpose text prediction (GPT-series), is correlated with the increase of their application across radiology, pathology, and clinical decision support for lung cancer. In 2026 alone, 12 publications were indexed at time of data extraction; therefore, publication activity will continue to occur and will likely occur at an increasing rate (Figure 1).

Table 1: Search strategy and data selection criteria for retrieving publications on large language models and lung cancer from the Scopus database (data extracted on 30 January 2026).

Item No.	Parameter	Specification
1	Write exactly like the reference paper	347
2	Database	Scopus
3	Date of data extraction	30-01-2026
4	Search strategy	TITLE-ABS-KEY
5	Keywords used	Large Language Model, Lung Cancer
6	Inclusion criteria	All article included in Scopus
7	Exclusion criteria	Retracted Article
8	Final number of documents	346

Overall, the year-wise publication trend demonstrates a clear transition from exploratory research to accelerated growth and maturation of the field. The increasing volume of publications highlights the rising academic and clinical interest in leveraging large language models for lung cancer diagnosis and management, and indicates strong potential for continued research expansion soon.

Type of papers published

The document type distribution pertaining LLMs research, with an emphasis on lung cancer, is demonstrated by Table 3. Research articles represent the largest proportion of publications (a total of 212), suggesting that most of this research is based on original and empirical/manuscript methodologies. The predominance of articles is indicative of how heavily the development, examination and validation of large language model approaches used for lung cancer diagnosis, reporting and clinical decisions are prioritised.

The second most common group of publication types are conference papers (60 total), demonstrating that early and quickly disseminated research results and novel technology in this new area of research are continuing to grow in number. With 33 review articles in total, the field has likely matured as reviews typically appear when primary research in a particular area reaches a level of maturity; thus, synthesis of current work in that area is required. Another way of indicating that work in this new area remains ongoing and there is interest in collecting and summarising recent developments that researchers have presented at scholarly conferences is the total of ten review articles presented at these conferences (19 total documents) (Figure 2).

The other document types (editorial (12), letter (5), note (3) and book chapter (1)) represent an insignificant percentage of the total number of publications. However, these document types can provide expert commentary, initial findings, or specific discussions regarding uses and challenges associated with utilising large language models in relation to lung cancer. Overall, the distribution of document types illustrates both the rapid increase in research exploring large language models and the gradual consolidation of this work into a growing evidence base around lung cancer research.

Journal-wise Documents Published

Table 4 shows that NPJ Digital Medicine is the most active publisher of research papers dealing with LLMs in the context of lung cancer (10 publications). As such, NPJ Digital Medicine plays a key role in disseminating research at the intersection of clinical medicine and AI. The increased focus on technologies

Table 2: Year-wise distribution of publications on LLMs with a focus on lung cancer (2015-2026).

Year of Publication	Total Documents	Year of Publication	Total Documents
2026	12	2020	3
2025	144	2019	5
2024	122	2018	1
2023	15	2017	2
2022	11	2016	3
2021	12	2015	2

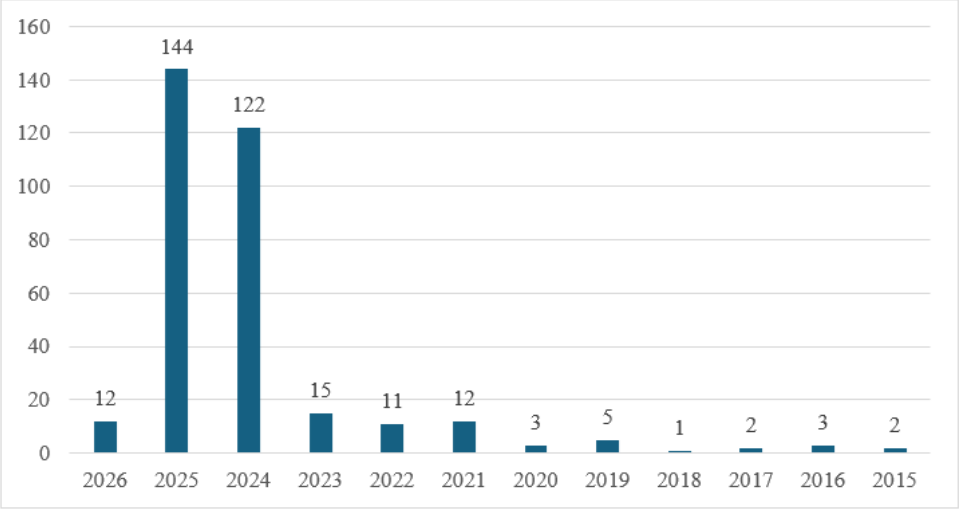


Figure 1: Annual publication trend of studies on LLMs in lung cancer research (2015-2026), based on Scopus data extracted on 30 January 2026.

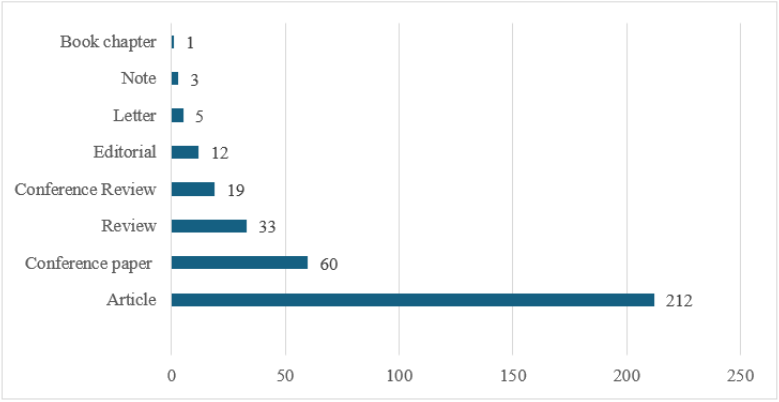


Figure 2: Distribution of document types in LLMs-based lung cancer research.

Table 3: Distribution of document types in large language model-based lung cancer research indexed in Scopus.

Documents Type	Documents Number	Documents Type	Documents Number
Article	212	Editorial	12
Conference paper	60	Letter	5
Review	33	Note	3
Conference Review	19	Book chapter	1

Table 4: Most productive source titles publishing research on LLMs in lung cancer.

Sl. No.	Source	Documents
1.	Npj Digital Medicine	10
2.	JCO Clinical Cancer Informatics	9
3.	Radiology	9
4.	Journal of Biomedical Informatics	7
5.	Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence, Lecture Notes in Bioinformatics	7
6.	Cancers	6
7.	Cancer	5
8.	Computers in Biology and Medicine	5

related to digital health, AI-assisted diagnostics, and translational medical research in NPJ Digital Medicine further supports this statement. JCO Clinical Cancer Informatics and Radiology rank second in terms of total number of publications, each having 9 total publications. JCO Clinical Cancer Informatics highlights the increasing role of informatics/computational methods in oncology practice. Radiology exemplifies how widely large language models are applied to interpret radiology reports, screen patients, and guide the management of lung cancer through diagnostic workflows.

The Journal of Biomedical Informatics and Lecture Notes in Computer Science, along with their AI and bioinformatics subseries, have collectively produced a total of 7 articles, showing the multi-disciplinary nature of this research area, as it relates to both computer science and biomedical informatics, as well as potential clinical applications. In addition, oncology related journals such as Cancers (6 articles), and Cancer (5 articles) are showing increased acceptance for the use of LLMs methodologies within established cancer research fields (Figure 3).

Ultimately, Computers in Biology and Medicine (5 publications) indicates how computational intelligence methods have been incorporated into biological and medical research. In addition, the journal distribution shows that research done on large language models related to lung cancer is published in various interdisciplinary and domain-specific journals. The convergence of AI and biomedical research in this field is driving innovation at a fast pace.

Top Papers with highest Citations

The publications contained in Table 5 are those that are the most referenced in the literature according to the available data and show how those studies have played a vital role in the establishment of research into LLMs and lung cancer. For each of these papers, there are several underlying themes that correspond to various areas of research and include clinical meta-analyses, digital pathology/ foundation model, radiology applications, and LLM-assisted clinical text interpretation. The inclusion of both classical oncology and more recent AI-related studies reflects the multidisciplinary approach and progressive intellectual basis that characterizes this field.

The increasing influence of LLMs in the clinical oncology workflow and knowledge extraction is shown through the highly cited literature on ChatGPT LLMs, Digital Pathology, and LLMs decision support applications. The primary evidence for supporting the adoption of LLMs into lung cancer research is

drawn from these influential publications on LLM methodology and clinical practice advancements.

Country-Wise Research Productivity and Citation Impact

The data in Table 6, identify a country-level assessment of research being conducted on large language models in lung cancer and show publication productivity and citation impacts. Research findings for this area indicate that the United States maintains the highest number of publications and citations; thereby, indicating that the US has an established research infrastructure and the early adoption of AI within this area (Oncology). There is also noteworthy publication and citation activity by both China and

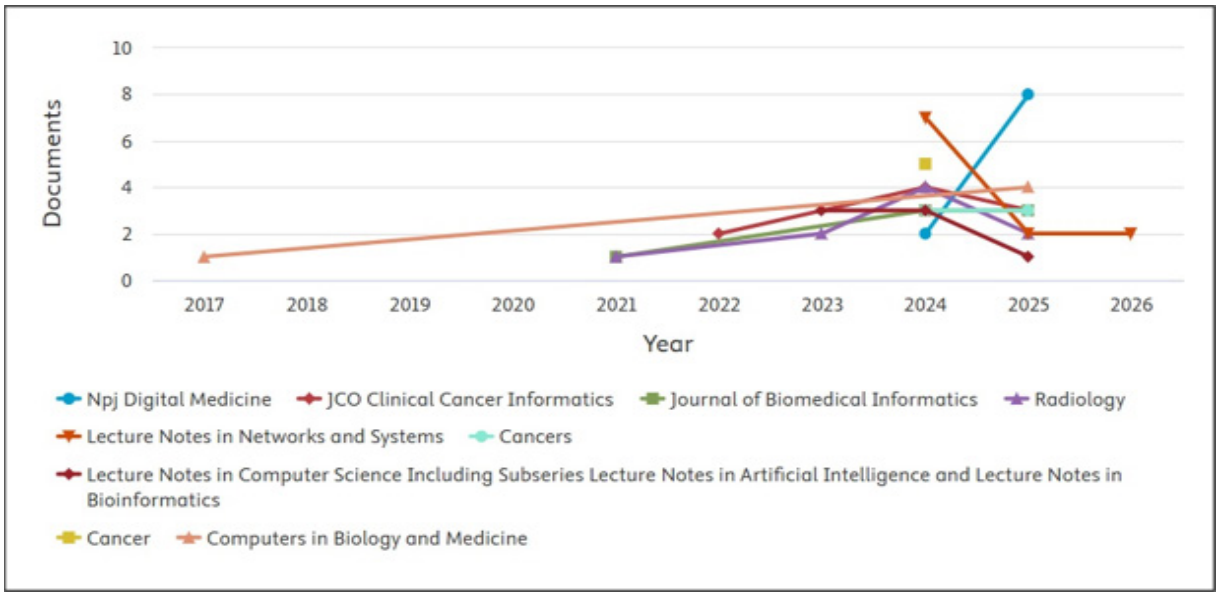


Figure 3: Most productive journals publishing research on LLMs in lung cancer.

Table 5: Top highly cited publications in LLMs-based lung cancer research retrieved from the Scopus database, highlighting influential studies shaping the intellectual foundation of the field.

Sl. No.	Title of the research articles	Year	Source title	Cited
1.	Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients	2021	BMC Infectious Diseases	648
2.	A whole-slide foundation model for digital pathology from real-world data	2024	Nature	375
3.	Benefits of adding drug to a single-agent or a 2-agent chemotherapy regimen in advanced non-small-cell lung cancer: A meta-analysis	2004	JAMA Network	354
4.	PET-CT for assessing mediastinal lymph node involvement in patients with suspected resectable non-small cell lung cancer	2014	Cochrane Database of Systematic Reviews	244
5.	Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential	2023	Visual Computing for Industry, Biomedicine, and Art	233
6.	Association of Survival and Immune-Related Biomarkers with Immunotherapy in Patients with Non-Small Cell Lung Cancer: A Meta-analysis and Individual Patient-Level Analysis	2019	JAMA Network Open	228
7.	How AI Responds to Common Lung Cancer Questions: ChatGPT versus Google Bard	2023	Radiology	222
8.	Second malignancies after radiotherapy for prostate cancer: Systematic review and meta-analysis	2016	BMJ	220
9.	Potential of ChatGPT and GPT-4 for Data Mining of Free-Text CT Reports on Lung Cancer	2023	Radiology	192
10.	Leveraging Large Language Models for Decision Support in Personalized Oncology	2023	JAMA Network	188

Table 6: Country-wise distribution of publications and citation impact in LLMs-based lung cancer research indexed in the Scopus database.

Sl. No.	Country	Documents	Citations	S.N.	Country	Documents	Citations
1.	United States	125	3119	11	Italy	8	40
2.	China	68	759	12	Netherlands	8	27
3.	United Kingdom	31	805	13	Taiwan	8	31
4.	India	26	137	14	France	7	482
5.	Germany	22	504	15	Japan	7	155
6.	Canada	18	412	16	Pakistan	7	27
7.	South Korea	11	148	17	Austria	6	54
8.	Australia	9	43	18	Spain	6	262
9.	Hong Kong	9	72	19	Switzerland	6	33
10.	Turkey	9	35	20	Brazil	5	14

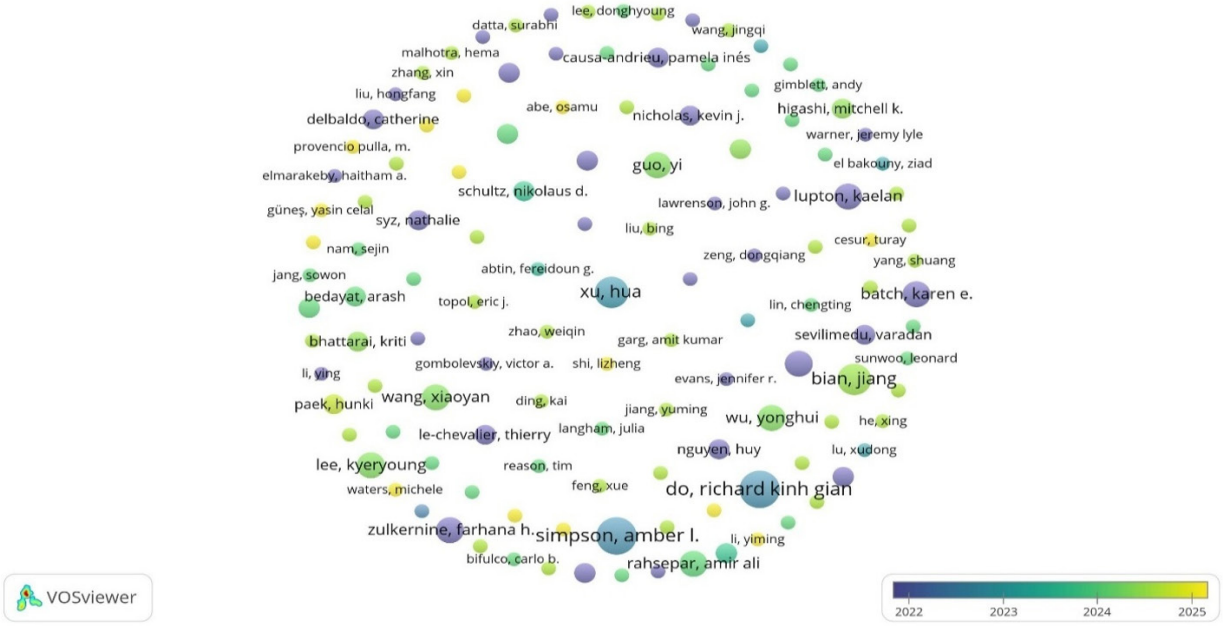


Figure 4: Author-wise analysis in LLM research related to lung cancer generated using VOSviewer. Node size represents the number of publications; Minimum number of documents of an author=2, Minimum number of citations of an author=2, Total authors identified: 2114, Authors meeting the thresholds: 126.

the United Kingdom, demonstrating the increased global interest in this interdisciplinary area of research.

Countries in Europe, such as Germany, France and Spain, have a high number of citations per article - although they typically have lower article output (i.e., they produce few articles). This information suggests the presence of few scientific papers that are highly influential on this journal. Countries, such as India, Pakistan and Brazil are also providing more researchers who are contributing to the field and reflecting the increasing diversity of research across borders. This analysis also highlights the international collaboration required to advance research in lung cancer using large language models and documents a disparity between the geographic locations of productivity and the influence of each country on LLM-based lung cancer research.

Bibliometric Analysis Using VOSviewer

A bibliometric analysis was conducted through VOSviewer to further analyze the intellectual structure, collaboration processes, and thematic changes in research relating to large language models and lung cancer. VOSviewer shows relationships between articles, authors, institutions, countries, and research themes as opposed to descriptive analyses, which have focused on the volume and source of publication.

Using VOSviewer, several network-based analyses were performed (i.e., citation analysis, co-authorship analysis, keyword co-occurrence analysis, and bibliographic coupling). These techniques help identify the most influential documents/authors, develop and identify collaborative research networks, and identify both dominant and emergent themes in the field. Each node in the networks created by VOSviewer represents a

bibliometric unit (document, author, keyword, organisation, or country), while links between nodes represent the strength of the relationship based on citations, co-authors, co-occurrence, and shared references. The size of each node represents the relative contribution of that unit (i.e., number of documents, citations, or keyword occurrences), while the distance between nodes represents how closely related those items are related. Clusters of items that are closely related are shown as different colours and represent groups of items that fit together as coherent research subdomains. The bibliometric analysis assists in providing additional understanding of the overall research landscape through visual representations of the research landscape; it allows for a more systematic understanding of the knowledge structure, the collaborative process of research, and research highlights of the repetitive language model lung cancer literature.

Citation Analysis

Authors wise Analysis

A co-authorship network of authors' collaborative patterns has been created using the program VOSviewer for the purpose of analyzing collaborative patterns among authors. During the process of creating the maps for the network, threshold values were utilized that would provide a clear and relevant visualization of the network. Thresholds for the VOSviewer maps were selected such that the minimum number of documents that a particular author had published would be 2, and the minimum number of citations that an author would have would also be a count of 2. The thresholds for document and citation counts were established in an attempt to include only those authors who have had a significant level of contribution and academia impact on the overall research area. Out of the total number of authors that were identified for the study (2114), only 126 authors met the thresholds for inclusion in the co-authorship network analysis. The authors that did not meet the established thresholds were omitted from the visualization to assist in reducing the overall complexity of the network and improve overall interpretability of the network. The resulting co-authorship map illustrates the collaborative relationships between all of the active authors contributing to high-level research using large language models focused on lung cancer, where an author is represented as a node, and all of the co-published works are represented by lines that connect the authors in the visual network (Figure 4).

The number of documents written by each author is shown in node size, while the frequency of co-authored documents connects the nodes. Networks identify clusters of authors as indications of collaboration, identifying the collaborative communities and research partnerships in this emerging area.

Country wise Analysis

A co-authorship network was created using VOSviewer to analyze worldwide research collaboration on large language model related research specific to lung cancer. A threshold value was used to create the map so that only those countries that had a sufficient level of research output and citation impact were included. In this case, each country must have 2 documents and 2 citations meet the minimum level of research output and citation impact per country. Of the 56 total countries in the original analysis, there were a total of 34 countries representing co-authorship at the time the co-authorship network was created. The 22 other countries (those countries that did not meet the threshold) were removed from the network so that the visual representation of the network would be less cluttered and provide for a better means of interpreting the structure of the international collaboration on lung cancer research. In the co-authorship network that was created, each node represents a country, while the links

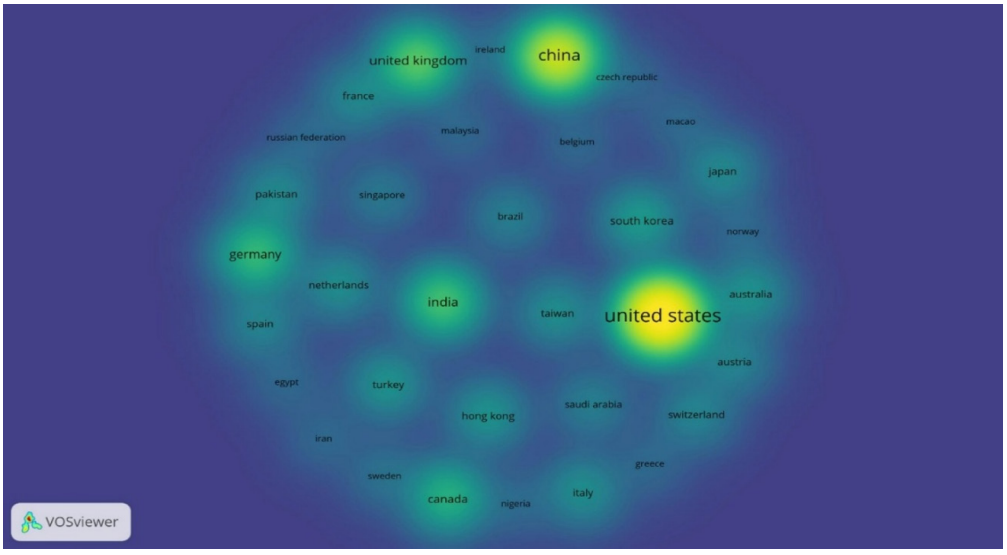


Figure 5: Country-wise analysis in LLM research related to lung cancer generated using VOSviewer. Node size represents the number of publications; Minimum number of documents of a country=2, Minimum number of citations of a country=2. Of the 56 countries, 34 meet the thresholds.

between each of these nodes show the existence of international collaboration through publications that are co-authored (Figure 5).

The size of each node in the co-authorship network is indicative of the number of documents that are co-authored, and the strength of the link between each of the nodes indicates the strength (number of linked nodes) of the collaboration. The clusters of countries within the co-authorship network are also indicative of countries with a strong collaborative tie, showing regional and international collaboration in large language model related research on lung cancer. This analysis of large language model related lung cancer research therefore can be used to show the distribution of research activity and the extent of collaboration across countries.

Document-wise Analysis

The first step in mapping the impact of the literature is to conduct a document level citation analysis of the literature related to large language model research in lung cancer, using VOSviewer. The mapping of the literature was completed using a set of established criteria to determine whether or not documents had a significant impact on academia. The minimum required citation threshold was set at 10 citations, and subsequently the analysis was used to identify documents with high citation counts and therefore high academic impact within the domain of large language model based lung cancer research. Based on this threshold, out of 346 documents identified within the domain, 90 documents met the minimum citation requirement, and were included in the citation network visualisation. The documents that did not meet the citation threshold were excluded resulting in a citation network for the remaining documents that had a lower density of documents and provided an easier pathway for researchers to interpret the citation network structure. The network includes nodes representing individual documents and linked nodes representing citation linkages (Figure 6).

The size of each node represents the number of times the document was cited, and therefore the larger the node, the greater the citation count. The clustering of the nodes represents groups of closely related documents that have contributed to the foundation of large language model based lung cancer research. Based on this citation analysis, researchers can determine the most influential research contributions and identify key studies that have been instrumental to the growth and evolution of this rapidly emerging area of research.

Keywords co-occurrence Analysis

A keyword co-occurrence analysis (using VOSviewer) was performed on research concerning LLMs and lung cancer based on author's keywords to identify thematic structure related to research efforts using LLMs about lung cancer. Keyword co-occurrence analysis is an important bibliometric tool to identify core themes of research and emerging hotspots by looking at how many times the keywords co-occurred within the published works. The network visualization from the keyword co-occurrence results indicates that lung cancer is a central or highly connected keyword and importantly represents a key theme in how the research area was developed. The keywords that are closely related to lung cancer are primarily related to medical diagnosis, radiology, pathology, and clinical decision support, therefore reflecting the primary areas where LLMs will be applied

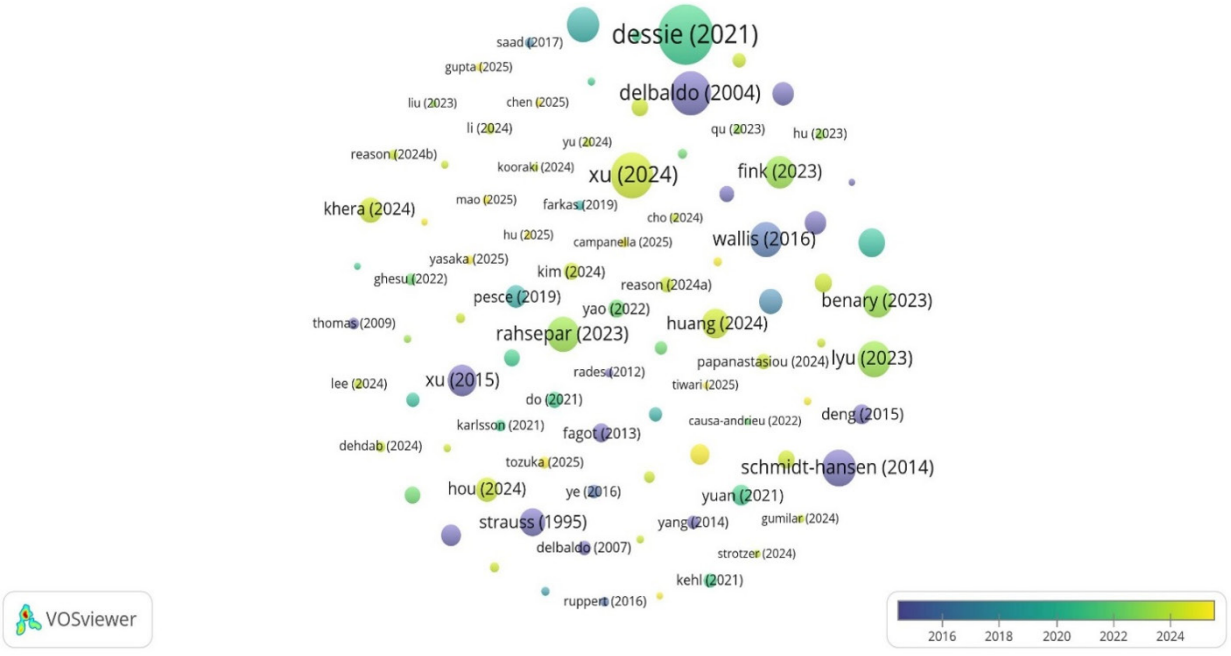


Figure 6: Document citation network of highly cited publications (minimum citations per document=10) in LLM-based lung cancer research. Node size corresponds to citation frequency; clusters represent closely related research streams.

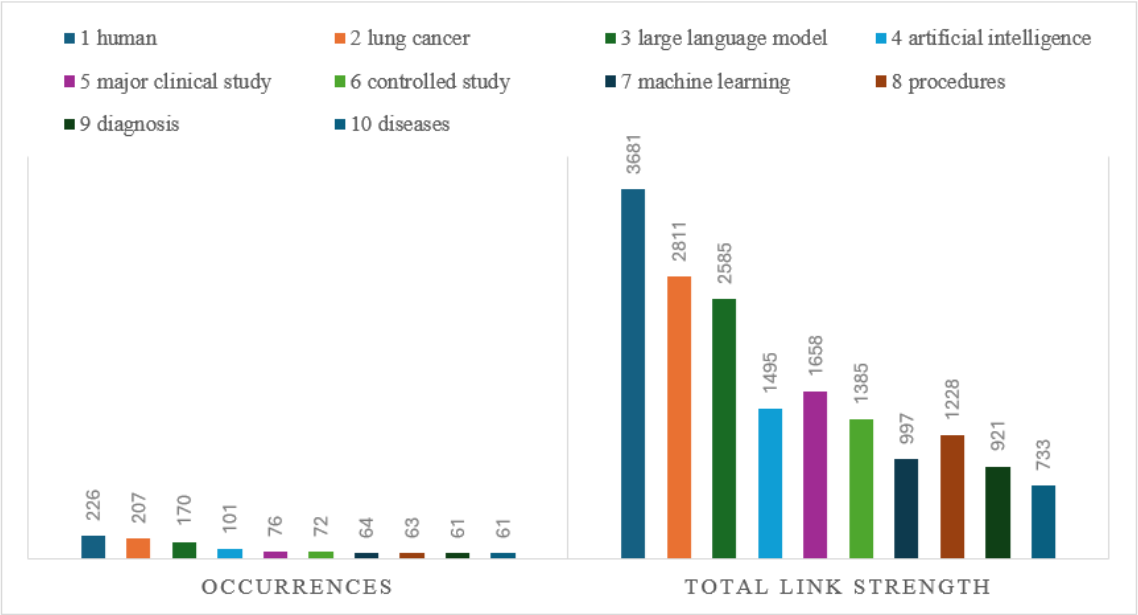


Figure 7: Author keyword co-occurrence network focusing on "lung cancer" in LLM research. Visualization generated using VOSviewer (version 1.6.20); node size corresponds to keyword frequency, link strength reflects co-occurrence intensity, and colored clusters represent thematic research structures.

within the disciplined. The high levels of co-occurrence among these keywords suggest that LLMs will frequently be used in lung cancer to assist with diagnostic interpretation, staging, screening, and extraction of clinically relevant data from unstructured medical data sources (Figure 7).

Keywords co-occurrence with Lung Cancer

The analysis of the multiple keyword clusters found throughout the data helped to create distinct subthemes for researching. The first cluster has many diagnostic and screening keywords (i.e. "imaging report," "radiological evaluation," and "early detection"). The second cluster contains a significant number of keywords related to clinical text analysis and natural language processing, indicating that LLMs can assist in processing pathology reports, electronic health records, and clinical narratives. Other clusters include decision support systems, risk stratification, and multimodal integration which indicate that LLM applications are expanding beyond diagnostic activities (Figure 8).

The analysis of keyword co-occurrences indicates an interdisciplinary area of research in LLMs focused on lung cancer, integrating areas of research within oncology, medical imaging, biomedical informatics, and artificial intelligence. The high level of thematic connectivity within lung cancer establishes it as a prominent clinical challenge driving the creation and use of new LLM-based solutions. The analysis provides useful insights regarding known research themes as well as possible future studies that would expand the use of LLMs around lung cancer care.

Keywords co-occurrence with Large Language Model

To better explore the thematic composition of research involving large language models in lung cancer studies, a keyword co-occurrence analysis was performed by using VOSviewer with a focus on "large language model". This analysis indicates the conceptual relationship of large language models to other research topics, thus illustrating the technological and methodological emphasis of the body of literature. The co-occurrence network displays that "large language model" has a significant relationship with keywords associated with natural language processing, clinical text mining, and medical report analysis; this indicates that LLM's primary use is to process/unlock meaning from unstructured clinical data. Other significant keywords include the interpretation of radiology and pathology reports, information extraction, and clinical documentation and these reflect the predominant use of LLMs to automate and improve diagnostic workflow (Figure 9).

The keyword clusters in this network show distinct directions of research. One cluster is dedicated to model development, including many kinds of models related to transformers, GPT, and benchmarking studies. A second cluster is devoted to applying those models clinically, with a special focus on lung cancer diagnosis, staging and screening support. The third cluster shows a developing interest in multimodal and integration of vision and language, suggesting that researchers are increasingly looking to combine text and image data to improve the accuracy of diagnoses and support clinical decisions. Thus, from the analysis of co-occurring keywords, we see that potential research on large language models in lung cancer has transitioned from

exploratory methodology studies into clinically focused work. The strong connection between LLMs and clinical diagnostic tasks indicates their growing relevance to oncological research. This structure of themes supports the development of LLMs into a key enabling technology for lung cancer diagnosis and precision medicine, as well as outlining future directions for research; specifically, the artificial intelligence group believes future directions include enhanced multimodal integration and clinical deployment strategies.

Organization wise Bibliographic Coupling Analysis

A bibliography-based connection analysis using VOSviewer was performed to analyze the relationship of institutions conducting studies related to a LLM and lung cancer. Bibliographic coupling, which defines a pair of institutions to have a level of similarity between them, was calculated using the number of common references that each institution has in their published documents. This identifies institutions that are contributing to similar research areas or themes, and/or have related foundations of knowledge. The resulting network of connections between the institutions (or nodes), using the volume of common references as an indicator of the degree of bibliographic coupling, produces visual information that displays the absolute relationship of the institutions to one another based on their degree of bibliographic coupling. Institutions that exhibit a higher level of bibliographic coupling are found closer to one another in the visualization, indicating a higher level of thematic similarity in their research productivity. The application of threshold values to the bibliographic coupling output ensures the resulting network visualization is precise and can be useful for interpreting the bibliographic coupling result (Figure 10).

The analysis of this network shows the emergence of multiple clusters that include various types of organizations conducting research into using large language models for lung cancer. For example, the majority of organizations included in the larger clusters are either academic medical centers, advanced research universities, or interdisciplinary laboratories focused on biomedical informatics, oncology, and artificial intelligence. The existence of multiple clusters indicates the interdisciplinary character of this research area, as cross-institutional collaboration exists between many different types of organizations (e.g., medical schools, cancer research facilities, computer science). Furthermore, based upon the strength of bibliographic coupling by institution, the research on large language models in lung cancer is being conducted by a collaborative network of institutions rather than by independent researchers. Additionally, bibliographic coupling patterns illustrate that there is a convergence of expertise across the areas of oncology, medical imaging and artificial intelligence at the level of scientific institutions. Overall, results of this research provide useful information about the key areas

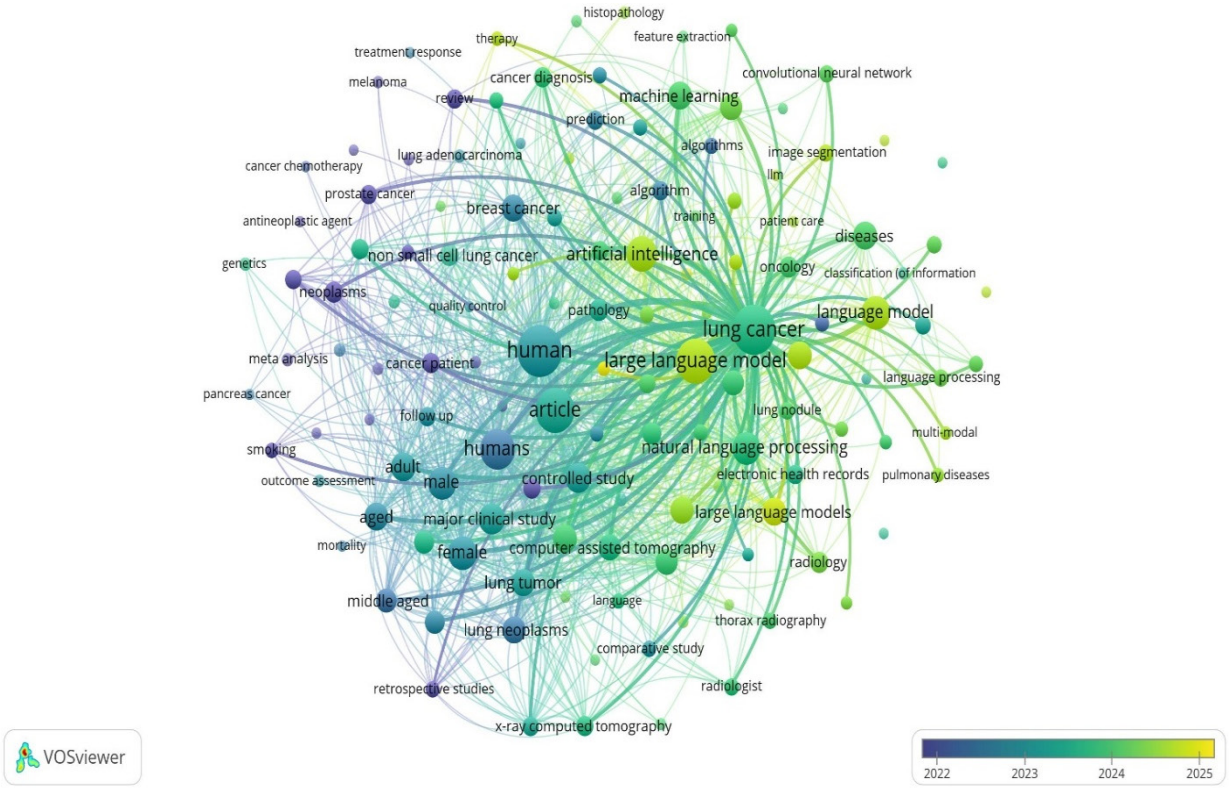


Figure 8: Keyword co-occurrence network centered on "lung cancer" generated using VOSviewer. Minimum number of occurrences of a keyword=12. Of the 3,788 identified keywords, 133 met the threshold and were included in the visualization. Node size represents keyword frequency, link strength reflects co-occurrence relationships, and clusters indicate major thematic research subdomains.

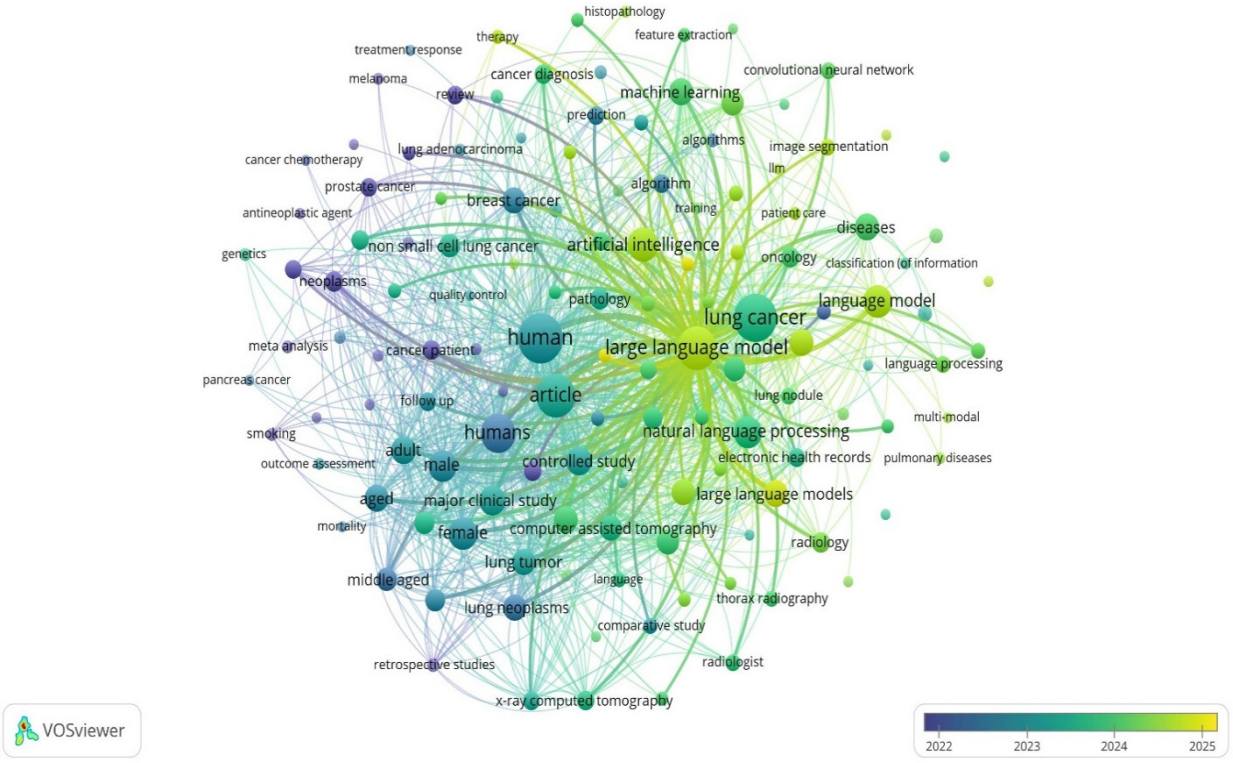


Figure 9: Keyword co-occurrence network centered on LLM generated using VOSviewer. Minimum number of occurrences of a keyword=12. Of the 3,788 identified keywords, 133 met the threshold and were included in the visualization. Node size represents keyword frequency, link strength reflects co-occurrence intensity, and clusters illustrate dominant methodological and clinical research themes.

of investigation with regard to institutional connections involved while developing this new area of research.

DISCUSSION

In this study we provide an overview of the bibliometric data for large language models and lung cancer research, including their structural changes over time; the degree to which exploration/research is clustered together thematically; and how collaborative patterns have developed in the context of this rapidly growing field. To achieve a comprehensive analysis of the development of large language models to date, we conducted a combination of descriptive statistical analyses of research outputs and network-based analysis of the same outputs using VOSviewer. This analysis provides an overall picture of how large language model-related research started as exploratory and has transitioned into large-scale, application-based, research-driven studies in recent years as indicated by the year-on-year data. As indicated by the analysis of year of publication, research output began to significantly increase after 2023 in line with the continuing proliferation of advanced transformer-based and GPT-3 series large language models and the increasing amount of clinical text and imaging data available for artificial intelligence-based analysis. The predominance of research articles and conference papers indicates that empirical validation, system development and real-world clinical applications are the primary driving forces behind the field, while the increasing number of review articles suggests that there is now a growing need to integrate the existing literature and provide a critical examination of the existing body of knowledge.

The journal analysis further demonstrates the interdisciplinary characteristics of this area of research. The prominence of journals including NPJ Digital Medicine, JCO Clinical Cancer Informatics and Radiology shows that artificial intelligence, biomedical informatics, and clinical oncology are converging. Publications with both medicine and computer science focus are examples of research related to the use of large language models for lung cancer within different scholarly communities that intersect with each other through cross-domain transfer of knowledge.

Bibliometric network studies give a richer view of the collaborative and intellectual fabric of the field. From a keyword co-occurrence perspective, lung cancer-focused LLM work is located thematically within the areas of medical diagnosis, interpreting radiology/pathology reports, using Natural Language Processing (NLP), and the provision of clinical decision support. Co-occurrence patterns indicate a slow evolution of lung cancer LLM research from methodological innovation toward clinically integrated and multimodal applications. The clustering of vision-language models and multimodal integration points to an increased emphasis on combining the textual and image-based data to support greater accuracy in diagnosis and enhanced clinical usefulness of the results. Strong partnerships between highly inter-dependent authors and institutions provide evidence of a small number of intellectually and organizationally central groups. Co-citation

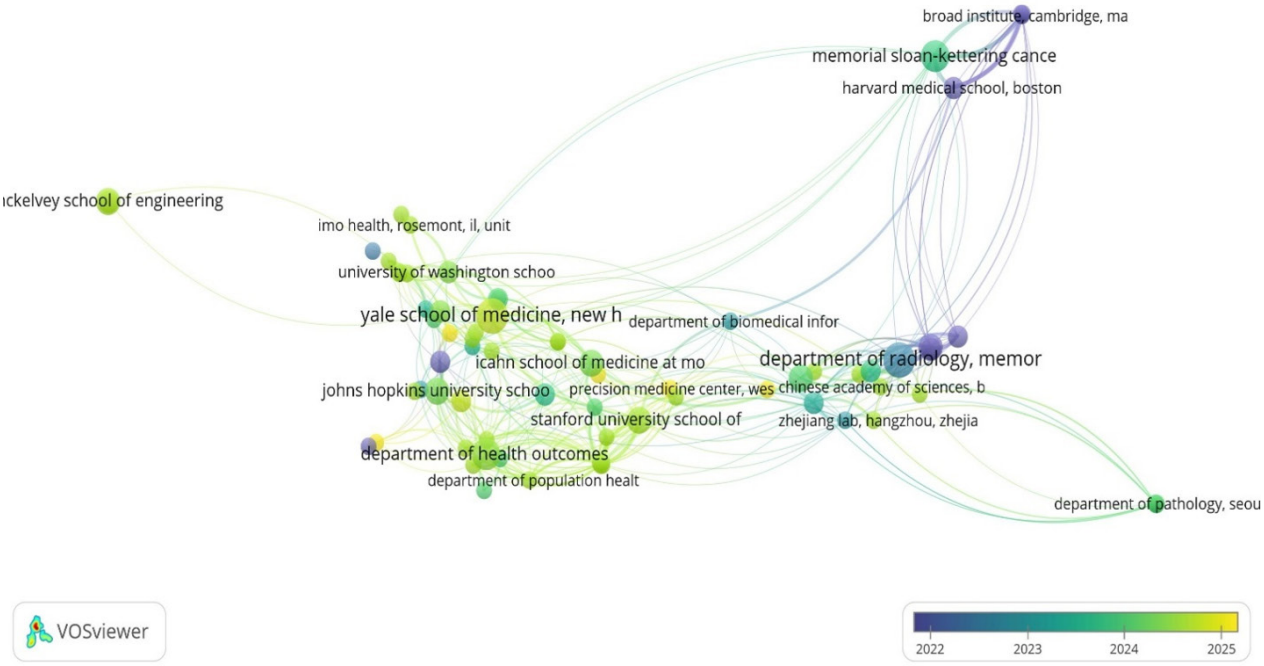


Figure 10: Organization-wise bibliographic coupling network in LLM-based lung cancer research generated using VOSviewer. Minimum number of documents per organization=2; minimum number of citations per organization=2. Of the 1100 identified organizations, 77 met the threshold criteria and were included in the visualization. Node size represents publication volume, link strength reflects similarity based on shared references, and clusters indicate institution-level research communities.

analyses provide insight into how key influencers in the natural language processing, medical informatics, and oncology literature form the theoretical foundation of the field. Similarly, bibliographic coupling of organizations indicates that much of the research is conducted in research-intensive Universities and academic medical centres, demonstrating the need for inter-disciplinary academic environments to advance LLM-based lung cancer research. Collaboration patterns at the country level further indicate the increasing inter-national character of the co-operative relationship between countries, which confirms the global significance of AI-enhanced diagnostic technologies for cancer.

Taken together, the findings suggest that LLM research in lung cancer has transitioned into a phase of accelerated growth and consolidation. The field is characterized by strong interdisciplinarity, concentrated intellectual leadership, and a clear movement toward clinically actionable application. However, the observed concentration around specific regions, institutions, and application areas also indicates potential gaps, particularly in terms of broader geographic representation, external clinical validation, and long-term outcome-focused studies.

CONCLUSION

This bibliometric analysis provides a detailed overview of the world's research trends in large language models focusing on lung cancer. The analysis utilized a systematic approach to collecting data from the Scopus database along with the VOSviewer software to visualize networks to evaluate the growth of publications, types of documents published, the leading journals, the leading authors, institutional collaboration, and the thematic clusters in research. The bibliometric analysis demonstrates a growing trend of research in this area, especially after 2023, which is attributed to advances in architecture of large language models and their growing role in the integration of clinical oncology workflow. Additionally, it is evident that most of the research in this area has been produced from an interdisciplinary perspective, specifically around the themes of diagnosis of disease, interpretation of radiology and pathology reports, and assisting medical professionals with clinical decision-making; whereas, most publications originate from research-intensive institutions and interdisciplinary journals. Overall, efforts made by influential authors and institutions have contributed to shaping the intellectual structure of this area of study, while keyword networks and citation networks continue to move toward multimodal/clinically integrated AI applications. Collectively, this analysis indicates that large language models are becoming a foundational technology for lung cancer research and precision oncology.

Although this research contributes to the understanding of its subject, three shortcomings exist within it that describe some future directions of research. First, the database used to complete the analysis was limited to the Scopus database; future bibliometric studies could expand use of other databases that include Web of Science or PubMed to gain a more complete view of the literature. Second, as this area of research is rapidly changing, periodic updates to the dataset are needed to include recently published literature and emerging trends in the field.

Future studies could also include attributes related to the clinical impact and translational readiness of the application of large language models to lung cancer care, as well as compare the use of large language models between cancer types and examine ethical and regulatory concerns and assess their applications in a clinical setting. Finally, if greater emphasis is placed on explainability, robustness, and equitable participation on a global basis in the development and use of large language models, the responsible development of this technology could be guided. In summary, this study represents a significant contribution to our understanding of the evolution of large language model research in lung cancer; it is anticipated that the results will help assist researchers, clinicians, and decision-makers with identifying research areas of interest, increasing collaboration, and providing direction for the future evolution of innovation in the integration of artificial intelligence and cancer care.

ABBREVIATIONS

LLM: Large Language Model; **AI:** Artificial Intelligence; **GPT:** Generative Pre-trained Transformer; **VLM:** Vision-Language Model; **NLP:** Natural Language Processing; **CDSS:** Clinical Decision Support System; **MDT:** Multidisciplinary Team; **TNM:** Tumor-Node-Metastasis staging system.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Ancusa, V. M., Trusculescu, A. A., Constantinescu, A., Burducescu, A., Fira-Mladinescu, O., Manolescu, D. L., *et al.* (2025). Temporal trends and patient stratification in lung cancer: A comprehensive clustering analysis from Timis County, Romania. *Cancers*, *17*(14), 2305.

Chang T. Y., Gou, Q., Leyi Zhao, L., Zhou, T., Hongyu Chen, H., Yang, D., Ju, H., Smith, K. E., Sun, C., Pan, J., Huang, Y., He, X., Zhang, X., Xu, D., Xu, J., Bian, J., & Chen, A., (2025), From image to report: Automating lung cancer screening interpretation and reporting with vision-language models. *Journal of Biomedical Informatics,* https://doi.org/10.1016/j.jbi.2025.104931

Chang, Y. C., Hsiao, S. H., Yeh, W. C., Hsing, Y. C., Wang, C. C., & Chen, C. Y. (2025). Extracting critical clinical indicators and survival prediction of lung cancer from pathology reports using large language models. *Computers in Biology and Medicine,* 195, https://doi.org/10.1016/j.compbiomed.2025.110621

Chao, P. J., Chang, C. H., Wu, J. J., Liu, Y. H., Shiau, J., Shih, H. H., *et al.* (2024). Improving prediction of complications post-proton therapy in lung cancer using large language models and meta-analysis. *Cancer Control,* 31, https://doi.org/10.1177/10732748241286749

Cho, H., Yoo, S., Kim, B., Jang, S., Sunwoo, L., Kim, S., *et al*. (2024). Extracting lung cancer staging descriptors from pathology reports: A generative language model approach. *Journal of Biomedical Informatics*, 157, https://doi.org/10.1016/j.jbi.2024.104720

Choi, E., Luo, S. J., Ding, V. Y., Wu, J. T., Kumar, A. V., Wampfler, J., *et al.* (2024). Risk model-based management for second primary lung cancer among lung cancer survivors through a validated risk prediction model. *Cancer*, *130*(5), 770-780.

Du, H., Zheng, Y., Jing, B., Zhao, Y., Kou, G., Liu, G., *et al.,* (2025), Causal discovery through synergizing large language model and data-driven reasoning. *In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Vol. 2, pp. 543-554, Association for Computing Machinery*, https://doi.org/10.1145/3711896.373687

Duan, Z., Huang, X., Lu, R., Xu, W., Liu, H., Geng, Y., et al. (2025). Multi-center benchmarking of large language models for clinical decision support in lung cancer screening. *Cell Reports Medicine*, 6(12), https://doi.org/10.1016/j.xcrm.2025.102465

Fung, M. M., Tang, E. H., Wu, T., Luk, Y., Au, I. C., Liu, X., *et al.* (2025). Developing a named entity framework for thyroid cancer staging and risk level classification using large language models. *NPJ Digital Medicine*, *8*(1), 134.

Garg, A., Gupta, S., Vats, S., Handa, P., & Goel, N. (2024). Prospect of large language models and natural language processing for lung cancer diagnosis: A systematic review. *Expert Systems*, *41*(11), e13697.

Ghorbian, M., & Ghobaei-Arani, M. (2025). A comprehensive review of LLM applications for lung cancer diagnosis and treatment: Classification, challenges, and future directions. *Journal of Big Data*, *12*(1), 249.

Güneş, Y. C., Cesur, T., Çamur, E., & Karabekmez, L. G. (2025). Evaluating text and visual diagnostic capabilities of large language models on questions related to the Breast Imaging Reporting and Data System Atlas 5th edition. *Diagnostic and Interventional Radiology*, *31*(2), 111.

Hackl, W. O., Neururer, S. B., Richter, S., Taha, H., Muehlboeck, H., Hickmann, C., et al. (2025). Development of a synthetic oncology pathology dataset for large language model evaluation in medical text classification. *In dHealth* 2025 (pp. 215-220). IOS Press, https://doi.org/10.3233/SHTI250191

Hao, Y., Qiu, Z., Holmes, J., Löckenhoff, C. E., Liu, W., Ghassemi, M., & Kalantari, S. (2025). Large language model integrations in cancer decision-making: A systematic review and meta-analysis. *NPJ Digital Medicine*, *8*(1), 450.

Ismayilov, R., Altundag, O., Gencoglu, E. A., Aktas, A., Alparslan, S., Ozcicek, A., *et al.* (2025). Artificial intelligence for immunotherapy response assessment in lung cancer using PET/CT reports. *Japanese Journal of Radiology*, *43*(12), 2042-2050.

Jiang, Q., Wang, X., Jiang, X. A., Sun, H., Chen, Q., Cheng, Y., *et al.* (2025). A narrative review of preoperative CT for predicting spread through air spaces of lung cancer. *Translational Lung Cancer Research*, *14*(4), 1471.

Kaba, E. (2025). Retrieval-augmented generation for answering Breast Imaging Reporting and Data System (BI-RADS)-related questions with large language models. *Diagnostic and Interventional Radiology*. Advance online publication. 31(6), 566-567

Kim, A., Miller, D. G., Lee, V., Simhal, A. K., Meng, Y. J., Book, A. J., *et al*. (2025). Leveraging large language models on personal computers for automated identification of lung cancer progression in radiology reports. https://doi.org/10.1200/OP.2025.21.10_suppl.624

Kim, K., Lee, Y., Park, D., Eo, T., Youn, D., Lee, H., *et al.* (2024). LLM-guided multi-modal multiple instances learning for 5-year overall survival prediction of lung cancer. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 239-249). *Springer Nature Switzerland.* https://doi.org/10.1007/978-3-031-72384-1_23

Kumsa, F. A., Brett, C. L., Hashtarkhani, S., Rashid, R., Chinthala, L., Zink, J. A., *et al.* (2025). Leveraging centralized health system data management and large language model-based data preprocessing to identify predictors for radiation therapy interruption. JCO Clinical Cancer Informatics, 9, e2500218, https://doi.org/10.1200/CCI-25-00218

La Rosa, M., Fiannaca, A., Mendolia, I., La Paglia, L., & Urso, A. (2025). GL4SDA: Predicting snoRNA-disease associations using GNNs and LLM embeddings. *Computational and Structural Biotechnology Journal*, 27, 1023-1033. https://doi.org/10.1016/j.csbj.2025.03.014

Li, Y., Li, F., Hong, N., Li, M., Roberts, K., Cui, L., *et al.* (2025). A comparative study of recent large language models on generating hospital discharge summaries for lung cancer patients. *Journal of Biomedical Informatics*, https://doi.org/10.1016/j.jbi.2025.104867

Liu, J., Yang, M., Bi, Y., Zhang, J., Yang, Y., Li, Y., *et al*. (2025). Large language models enable tumour-type classification and localisation of cancers of unknown primary from genomic data. *Cell Reports Medicine*, 6(9). https://doi.org/10.1016/j.xcrm.2025.102332

Liu, Y., Wang, F., Wang, P., Zhou, Z., Wang, H., Li, J., *et al.* (2026). AI-MDT: An automatic and intelligent multidisciplinary team consultations platform for lung cancer diagnosis. *Journal of Cancer Research and Clinical Oncology*, *152*(1), 32.

Luo, I., Graber-Naidich, A., Zhang, M., Kaushik, R., Nieda, G. M., Chen, T., *et al.* (2025). Leveraging large language models to extract smoking history from clinical notes for lung cancer surveillance. *NPJ Digital Medicine*, *8*(1), 731.

Luo, Y., Hooshangnejad, H., Ngwa, W., & Ding, K. (2025). Opportunities and challenges in lung cancer care in the era of large language models and vision language models. *Translational Lung Cancer Research*, *14*(5), 1830.

Martell, M. B., Linton-Reid, K., Chen, M., & Aboagye, E. O. (2025). Radiomics for lung cancer diagnosis, management, and future prospects. *Clinical Radiology,* 86, https://doi.org/10.1016/j.crad.2025.106926

Matsuo, H., Nishio, M., Matsunaga, T., Fujimoto, K., & Murakami, T. (2024). Exploring multilingual large language models for enhanced TNM classification of radiology report in lung cancer staging. *Cancers*, *16*(21), 3621.

Moore, C. L., Socrates, V., Hesami, M., Denkewicz, R. P., Cavallo, J. J., Venkatesh, A. K., & Taylor, R. A. (2025). Using natural language processing to identify emergency department patients with incidental lung nodules requiring follow-up. *Academic Emergency Medicine*, *32*(3), 274-283.

Mou, Y., Lehmkuhl, J., Sauerbrunn, N., Köchel, A., Panse, J., Truh, D., et al. (2024). Improving the quality of unstructured cancer data using large language models: A German oncological case study. *In Digital health and informatics innovations for sustainable health care systems* (pp. 685-689). IOS Press, https://doi.org/10.3233/SHTI240507

Nguyen, D., Rahsepar, A. A., Chalian, H., Hassani, C., Kim, G. H. J., & Bedayat, A. (2025). Revisiting LLMs and lung cancer questions: How AI responds to common lung cancer questions two years later. *Academic Radiology.* https://doi.org/10.1016/j.acra.2025.08.056

Onishi, S., Shibata, D., Tsujikawa, M., Kubo, M., Nakagawa, A., Ishii, R., *et al.* (2025). Fine-grained prompting in large language models for accurate and efficient TNM staging from radiology reports. In 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1-6). IEEE. https://doi.org/10.1109/EMBC58623.2025.11252841

Park, H. J., Huh, J. Y., Chae, G., & Choi, M. G. (2024). Extraction of clinical data on major pulmonary diseases from unstructured radiologic reports using a large language model. *PLoS ONE*, *19*(11), e0314136.

Qiu, Z., Jiang, A., Qi, C., Gan, W., Zhu, L., Mou, W., ... & Luo, P. (2025). Temporal evolution of large language models (LLMs) in oncology. *Journal of Translational Medicine*, *23*(1), 1-17.

Reddy, C. L., & Prasad, M. V. D. (2025). Automatic lung cancer detection from CT image using optimized Robust Deformed Convolutional Neural Network with TriHorn-Net. *Expert Systems with Applications,* 276, 127124, https://doi.org/10.1016/j.eswa.2025.127124

Ruan, Q., Shi, J., Dai, Y., Yang, P., Zhu, N., & Wang, S. (2025). Performance of large language models in complex anesthesia decision-making: A comparative study of four LLMs in high-risk patients. *Journal of Medical Systems*, *49*(1), 1-12.

Schimmelpfennig, L. E., Bhattarai, K., Oh, I. Y., Lever, J., Griffith, O. L., Griffith, M., ... & Abrams, Z. B. (2025). CeRTS: Certainty retrieval token search in large language model clinical information extraction. *Journal of Biomedical Informatics*, 104900.

Shi, N., Liu, Z., Wan, Z., Yang, G., Shi, Y., Wang, P., & Liu, X. (2025). A vision-language model-based approach for lung cancer diagnosis using lossless 3D CT images: Evaluation of GPT-4.1 and GPT-4o for patient-level malignancy assessment. *Health Information Science and Systems*, *14*(1), 16.

Tariq, A., Trivedi, S., Urooj, A., Ramasamy, G., Fathizadeh, S., Stib, M., *et al.* (2025). Patient-centric summarization of radiology findings using two-step training of large language models. *ACM Transactions on Computing for Healthcare*, *6*(2), 1-15.

Timilsina, M., Buosi, S., Torrente, M., Provencio, M., Cobo, M., Rodríguez Abreu, D.,... & Nováček, V. (2025). Large language model vs. traditional machine learning: Evaluating predictive models for early detection of tumor relapse. 283, 127641

Tozuka, R., Johno, H., Amakawa, A., Sato, J., Muto, M., Seki, S., *et al.* (2025). Application of NotebookLM, a large language model with retrieval-augmented generation, for lung cancer staging. *Japanese Journal of Radiology*, *43*(4), 706-712.

Tripathi, A., Waqas, A., Venkatesan, K., Ullah, E., Khan, A., Khalil, F., *et al.* (2025). Using consensus-based reasoning and large language models to extract structured data from surgical pathology reports. *Laboratory Investigation*, Article 104272.

Wang, H., & Bai, Z. (2025). Advances in CT-based artificial intelligence techniques for lung nodule diagnosis. *Chinese Journal of Clinical Research*, *38*(5).

Wang, Q., Wang, Z., Li, M., Ni, X., Tan, R., Zhang, W., ... & Liu, C. (2025). A feasibility study of automating radiotherapy planning with large language model agents. *Physics in Medicine & Biology*, *70*(7), 075007.

Wiest, I. C., Wolf, F., Leßmann, M. E., van Treeck, M., Ferber, D., Zhu, J., ... & Kather, J. N. (2025). A software pipeline for medical information extraction with large language models, open source and suitable for oncology. *npj Precision Oncology*, *9*(1), 313.

Yan, Y., Yang, Y., Tong, Z., Wang, Y., Yang, F., Pan, Z., ... & Li, Y. (2025). Adaptive multi-view learning method for enhanced drug repurposing using chemical-induced transcriptional profiles, knowledge graphs, and large language models. *Journal of Pharmaceutical Analysis*, 15(6), 101275.

Yang, J., Cai, X., Dai, X., & Xie, C. (2025). Assessing the performance of ChatGPT-4 and ChatGPT-4o in lung cancer diagnoses. *Journal of Translational Medicine*, *23*(1), 346.

Yang, S., Liang, M., Mehta, H. J., Salloum, R. G., Braithwaite, D., Wu, Y., *et al.* (2025). Patient and nodule characteristics associated with adherence to lung cancer screening in a large integrated healthcare system. *Scientific Reports*, *15*(1), 29172.

Yang, S., Yang, X., Lyu, T., Huang, J. L., Chen, A., He, X., *et al.* (2024). Extracting pulmonary nodules and nodule characteristics from radiology reports of lung cancer screening patients using transformer models. *Journal of Healthcare Informatics Research*, *8*(3), 463-477.

Yuan, Y., Zhang, G., Gu, Y., Hao, S., Huang, C., Xie, H., *et al.* (2025). Artificial intelligence-assisted machine learning models for predicting lung cancer survival. *Asia-Pacific Journal of Oncology Nursing*, *12*, 100680.

Cite this article: Kumar M, Mallipeddi R, Challa PR, Mann VS. Emerging Research Trends in Large Language Models with a Focus on Lung Cancer: Bibliometric Analysis. J Scientometric Res. 2026;15(3):1017-31.