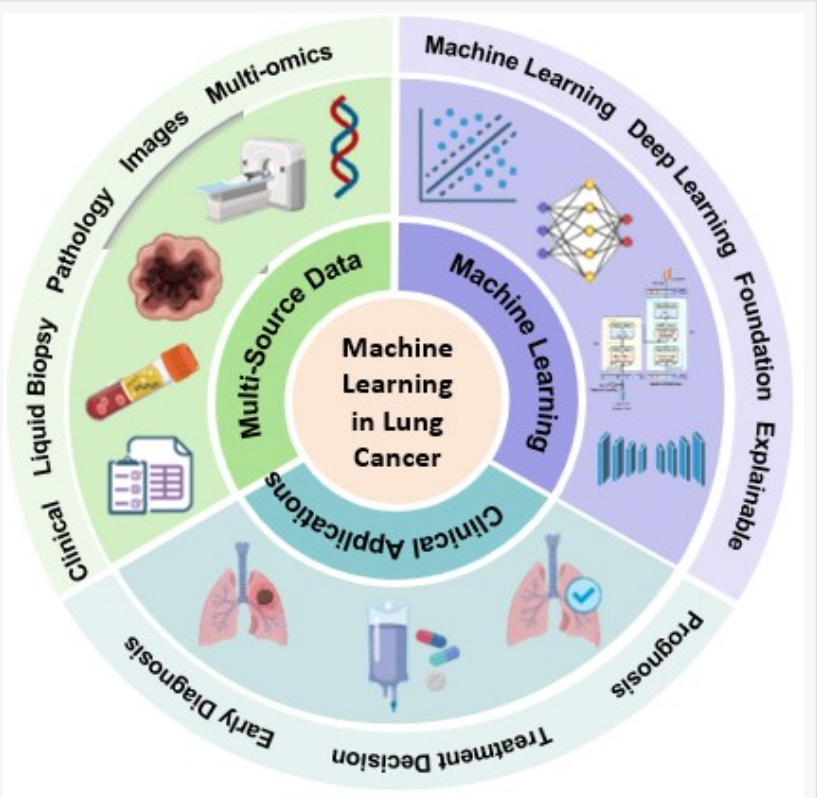
Multi-Source Data-Driven Machine Learning Reshapes the Diagnosis and Treatment of Lung Cancer

A comprehensive review of multi-source data-driven machine learning for lung cancer diagnosis, treatment, and prognosis



The integrated framework of data, model, and application for machine learning in lung cancer management

CHENGDU, CHINA, August 17, 2026 /EINPresswire.com/ -- Section 1: Background

Lung cancer remains one of the malignancies with the highest incidence and mortality worldwide. Clinical practice has long been plagued by core dilemmas, including insufficient sensitivity in early screening, lack of personalized treatment regimens, and limited accuracy in prognostic evaluation, which severely restrict the improvement of patient survival rates. Traditional lung cancer diagnosis and treatment rely heavily on empirical judgment, which can hardly address the high heterogeneity of tumors and the complex evolution of the disease course. With the continuous accumulation of medical resources such as medical imaging, omics detection, liquid biopsy, digital pathology, and electronic health records, lung cancer management has entered a new era driven by multi-source data. Machine learning, with its powerful capabilities in data mining and pattern recognition, can extract latent patterns from complex, heterogeneous, and multi-dimensional medical data. It transforms morphological features, molecular characteristics, pathological structures, blood biomarkers, and clinical information into quantitative evidence for diagnosis, treatment decision-making, and prognosis assessment, thus becoming a key technology to break through the bottlenecks of lung cancer care.

Against this background, a joint team from Shanghai Jiao Tong University, Qilu Hospital of Shandong University, and other institutions systematically analyzed the characteristics of multi-source lung cancer data and the evolution of machine learning technologies. Based on the data-model-application framework, the team comprehensively elaborated the innovative applications of multi-source data-driven machine learning in early screening, diagnosis, treatment optimization, and prognosis evaluation of lung cancer. The review reveals the inherent logic that data characteristics determine model selection and model performance supports clinical value, providing systematic theoretical support and technical references for the deep integration of artificial intelligence and lung cancer clinical decision-making, as well as the implementation of precision medicine.

Prof. Jinhong Guo, doctoral supervisor, Director of the Institute of Medical Multimodal Sensing, Shanghai Jiao Tong University.

Section 2: Summary of Review Content

The joint research teams from the Institute of Medical Multimodal Sensing at the School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, the Research Center for Internet of Medical Things at Qilu Hospital of Shandong University, and the Research Center of Shaoxing Keqiao Laboratory Medicine, Chongqing Medical University have published a review article in Volume 2 of Intelligent Opto-Electronics journal on June 29, 2026, entitled Multi-Source Data-Driven Machine Learning for Lung Cancer: Diagnosis, Treatment, and Prognosis. Centered on the main line of data type–model adaptation–clinical application, this review systematically summarizes the dimensional characteristics, clinical value, and complementary mechanisms of five core data sources (imaging, multi-omics, liquid biopsy, pathology, and clinical data). It compares the technical features, applicable scenarios, and advantages/disadvantages of traditional machine learning, deep learning, multimodal fusion, foundation models, and interpretable models, and systematically analyzes the latest progress of multi-source data-driven machine learning in lung cancer diagnosis, treatment, and prognosis.

At the data level, the review details the feature dimensions and clinical value of multi-modal data such as imaging, multi-omics, pathology, liquid biopsy, and clinical data, revealing their complementarity and synergistic effects in lung cancer care.

At the model level, it compares the technical characteristics and applicable scenarios of mainstream machine learning paradigms and clarifies the optimal selection strategies for different data types and clinical tasks.

At the application level, it deeply discusses the advances and challenges of multi-source data-driven machine learning in improving early diagnosis accuracy, optimizing personalized treatment, and realizing dynamic prognostic risk stratification, offering strong AI support for precise clinical decision-making.

This review innovatively proposes a complete logical chain of data characteristics–model adaptation–clinical value, defines optimal model selection strategies in different scenarios, and puts forward systematic solutions to key challenges including insufficient data standardization, bottlenecks in multimodal fusion, lack of algorithm interpretability, and delayed clinical translation, pointing out a clear path for the field to move from laboratory research to clinical application.

Section 3: Outlooks

Multi-source data-driven machine learning is driving a profound transformation of lung cancer care from experience-driven to data-driven, and from extensive intervention to precise stratification. Future efforts should focus on data standardization, adaptive multimodal fusion, improved interpretability, and prospective clinical validation to accelerate clinical translation. The ultimate goal is to build a full-cycle closed-loop diagnosis and treatment system with precise early screening, personalized therapy, and dynamic prognosis, thereby significantly improving the survival rate and quality of life of lung cancer patients.

Section 4: Introduction to the Research Group

This research was jointly completed by the team led by Prof. Jinhong Guo from the School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Qilu Hospital of Shandong University, and Chongqing Medical University. The team focuses on interdisciplinary research of multimodal medical sensing, medical artificial intelligence, and medical-industrial translation, especially the clinical translation of machine learning in biomedicine. The team has published more than 200 papers in journals including Nature Electronics, Advanced Materials, Angewandte Chemie, and IEEE Transactions series. It has authored 3 monographs, filed over 100 national invention patents (70+ granted), and obtained more than 30 medical device registration certificates. The team has won two First Prizes of Science and Technology Progress Award of Chinese Optical Engineering Society and led more than 10 national, provincial, and industrial projects. Committed to clinical demand, the team promotes the application of multi-source data fusion, deep learning, and interpretable AI in tumor screening, precision treatment, and prognosis evaluation, building an interdisciplinary innovation platform to boost the high-quality development of precision and smart healthcare.

Reference

Title of original paper: Multi-source data-driven machine learning for lung cancer: diagnosis, treatment, and prognosis
Journal: Intelligent Opto-Electronics
DOI: https://doi.org/10.67704/ioe.2026.260006

Funding Information: This work was supported by the National Key R&D Program of China (No. 2023YFF0724300).

Siyi Ma
Institute of Optics and Electronics, Chinese Academy of Sci
+86 19136186296
oej_publishing@ioe.ac.cn
