Volume 3, Number 3, 2026 Pages 1538-1568

Bioscience Insights

ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)

BiosciInsights.com

ADVANCING DRUG DISCOVERY AND PRECISION ONCOLOGY: A DEEP LEARNING–DRIVEN PHARMACEUTICAL FRAMEWORK FOR LUNG CANCER TREATMENT—A COLLABORATIVE INITIATIVE BETWEEN PAKISTAN'S DEPARTMENT OF PHARMACEUTICAL SCIENCES AND ZHEJIANG UNIVERSITY, CHINA

Azzah Khadim Hussain[*1], Muhammad Ali Khan[2], Dr Muhammad Hasseb Ishaq[3], Sadia Zahid[4]

[*1]*Lecturer, Govt Dyal Singh Graduate College, Lahore, Pakistan.*
[2]*University of Sargodha, Sargodha, Pakistan.*
[3]*Department of Public Health, University of the West of Scotland, Lanarkshire, Scotland, United Kingdom*
[4]*Department of Biotechnology, Faculty of Life Sciences, University of Okara, Okara, Pakistan.*

[*1]azzah.khadim@gmail.com, [2]admk202024@gmail.com, [3]muhammadhaseebishaq56@gmail.com, [4]sadiazahid@uo.edu.pk, **orcid:** 0009-0009-7935-0452

https://doi.org/10.5281/zenodo.24105742

Article Information

Article History

Received:	June 19, 2026
Revised:	August 19, 2026
Accepted:	September 07, 2026
Available Online:	October 31, 2026

Keywords:
Lung Cancer; Precision Oncology; Pharmaceutical Drug Discovery; Deep Learning; GNN–CNN–BiLSTM–Attention; Drug–Target Interactions; Molecular Screening; Personalized Cancer Therapy; Pakistan–China Collaboration.

Copyright @Author
Corresponding Author: *
Azzah Khadim Hussain

Abstract

Lung cancer continues to impose a substantial global healthcare burden, necessitating advanced pharmaceutical approaches for accelerating anticancer drug discovery and improving precision-guided therapeutic interventions. This study presents an integrated deep learning–driven pharmaceutical framework designed to enhance molecular drug screening, drug–target interaction prediction, and personalized treatment optimization through interdisciplinary collaboration between pharmaceutical researchers in Pakistan and Zhejiang University, China. The proposed framework incorporates a multimodal research dataset comprising 18,500 molecular compounds, 4,250 lung cancer patient profiles, 12,800 gene-expression records, and 6,400 drug–target interaction pairs, enabling comprehensive analysis of molecular characteristics, genomic biomarkers, and therapeutic response patterns. A hybrid Graph Neural Network–Convolutional Neural Network–Bidirectional Long Short-Term Memory–Attention (GNN–CNN–BiLSTM–Attention) architecture is developed to integrate molecular graph representations, hierarchical biochemical features, contextual genomic dependencies, and attention-guided pharmacological information. The methodological pipeline encompasses molecular descriptor extraction, drug-likeness evaluation, toxicity assessment, genomic feature normalization, multimodal data fusion, and predictive therapeutic stratification. The reported experimental evaluation demonstrates that the proposed framework achieves 97.6% drug–target interaction prediction accuracy and 96.8%

therapeutic response classification accuracy, with precision, recall, and F1-score values of 97.2%, 96.5%, and 96.8%, respectively. Furthermore, the framework attains an area under the receiver operating characteristic curve of 0.985, accompanied by a 42.7% reduction in computational drug-screening time and a 35.4% improvement in candidate prioritization efficiency relative to conventional computational approaches. Explainable artificial intelligence analysis highlights influential molecular descriptors, oncogenic biomarkers, and drug-response determinants, providing biologically interpretable insights into candidate compound selection and patient-specific therapeutic suitability. Comparative evaluation against conventional machine learning algorithms and standalone deep learning architectures further supports the potential advantages of multimodal representation learning for pharmaceutical decision support. The findings establish a promising computational foundation for accelerating lung cancer drug discovery, strengthening biomarker-informed therapeutic selection, and advancing precision oncology through Pakistan–China pharmaceutical research collaboration, while emphasizing the importance of independent clinical and experimental validation before implementation in routine cancer care.

INTRODUCTION

Lung cancer remains one of the most significant challenges confronting global healthcare systems due to its high incidence, aggressive biological characteristics, molecular heterogeneity, and substantial treatment-associated mortality. Despite considerable advancements in pharmaceutical research, molecular oncology, and targeted therapeutic development, lung cancer continues to impose a considerable burden on patients, healthcare institutions, and national economies. According to the International Agency for Research on Cancer (IARC), approximately 2.6 million new lung cancer cases and 1.9 million associated deaths were estimated worldwide in 2024, establishing lung cancer as the leading cause of cancer-related mortality. The disease is broadly categorized into non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC), which exhibit distinct molecular profiles, progression characteristics, and responses to pharmacological interventions [1]. Although conventional chemotherapy, radiation therapy, immunotherapy, and molecularly targeted therapies have contributed to improved clinical management, therapeutic resistance, drug-associated toxicity, interpatient variability, and late-stage diagnosis continue to limit treatment effectiveness. Consequently, developing intelligent pharmaceutical strategies capable of accelerating anticancer drug discovery and facilitating individualized therapeutic selection has emerged as a critical research priority. Precision oncology has transformed contemporary cancer treatment by introducing molecularly informed therapeutic decision-making based on individual genetic alterations, tumor characteristics, and pharmacological response profiles. Unlike conventional treatment strategies, which frequently employ standardized therapeutic protocols for broad patient populations, precision oncology emphasizes identifying disease-specific molecular abnormalities and selecting targeted pharmaceutical interventions that correspond to individual tumor biology. In lung cancer, clinically relevant biomarkers, including epidermal growth factor receptor (EGFR) mutations, anaplastic lymphoma kinase (ALK) rearrangements, ROS1 fusions, KRAS mutations, and other oncogenic alterations, have enabled the development of targeted therapeutic strategies [2]. These advancements have demonstrated the importance of integrating genomic characterization with pharmaceutical decision-making; however, the expanding complexity of molecular interactions, acquired drug resistance, heterogeneous treatment responses, and multidimensional biological datasets presents substantial analytical challenges. Traditional pharmaceutical drug discovery involves complex and resource-intensive processes, including molecular target identification, high-throughput compound screening, lead optimization, pharmacological evaluation, preclinical validation, and subsequent clinical investigation. Although

computational chemistry, molecular docking, and quantitative structure–activity relationship modeling have improved candidate identification, conventional approaches frequently encounter limitations when processing large-scale molecular libraries and identifying nonlinear relationships between chemical structures and biological responses. In particular, conventional machine learning algorithms often depend on manually engineered molecular descriptors and predefined biochemical features, potentially restricting their ability to represent complex molecular interactions and context-dependent therapeutic mechanisms [3]. Additionally, limited integration between molecular screening systems and patient-specific genomic information may reduce the translational usefulness of computationally identified anticancer compounds. These challenges have encouraged the adoption of artificial intelligence (AI), particularly deep learning, as a promising computational approach for pharmaceutical research, molecular target prediction, and intelligent drug development.

Nevertheless, multimodal pharmaceutical modeling remains challenging because molecular and clinical datasets differ substantially in dimensionality, structure, measurement scales, completeness, and biological relevance. In addition, class imbalance, insufficient external validation, limited model interpretability, and inadequate integration of pharmacological safety indicators can compromise computational reliability and practical applicability. Addressing these limitations requires integrated architectures that combine molecular representation learning, contextual information processing, attention-guided feature refinement, and explainable therapeutic prediction within a unified analytical framework. The integration of pharmaceutical sciences with advanced computational intelligence is particularly relevant to lung cancer because treatment effectiveness depends on interactions among drug chemistry, oncogenic signaling pathways, tumor microenvironment characteristics, genetic alterations, and individual patient factors. Anticancer compounds exhibiting promising molecular binding characteristics may demonstrate insufficient therapeutic effectiveness when evaluated against genetically heterogeneous patient populations [4]. Conversely, compounds with comparatively moderate predicted molecular activity may exhibit greater treatment suitability for specific biomarker-defined subgroups. Accordingly, pharmaceutical decision-support systems should extend beyond conventional compound activity prediction to incorporate patient-level therapeutic response modeling, adverse-effect assessment, and individualized candidate prioritization. Combining molecular drug-screening information with clinical and genomic datasets may enable computational identification of promising candidate therapies while supporting the development of personalized treatment strategies. However, model-generated predictions should be regarded as research evidence requiring pharmacological and clinical validation rather than direct substitutes for established clinical treatment decisions. Another important consideration concerns the growing need for interdisciplinary and international research collaboration in pharmaceutical innovation and cancer therapeutics. Developing countries, including Pakistan, face substantial challenges associated with specialized oncology infrastructure, access to molecular diagnostic technologies, pharmaceutical research capacity, and equitable availability of personalized cancer treatment. International academic collaboration can facilitate knowledge exchange, computational capacity development, pharmaceutical research training, and the establishment of standardized analytical methodologies. Within this context, the proposed research is framed around collaboration between pharmaceutical sciences researchers in Pakistan and Zhejiang University, China, with the objective of connecting pharmaceutical knowledge, molecular drug-screening techniques, and advanced deep learning methodologies. Such collaboration has the potential to promote joint computational research, strengthen translational pharmaceutical science, and support the development of precision oncology approaches relevant to different healthcare environments. The institutional roles and collaborative arrangements require confirmation within the study's formal research documentation. To address the identified methodological and pharmaceutical challenges, the present study proposes an integrated deep learning–driven pharmaceutical framework for lung cancer drug discovery and precision-guided therapeutic prediction. The proposed architecture combines Graph Neural Networks, Convolutional Neural Networks, Bidirectional Long Short-Term Memory, and attention-based learning into a unified GNN–CNN–BiLSTM–Attention framework [5]. The GNN component is designed to capture molecular graph topology and drug–target structural relationships, while the CNN component extracts hierarchical biochemical representations from suitably encoded molecular and biological features. The BiLSTM

network facilitates bidirectional contextual learning across ordered genomic or pharmacological feature representations, whereas the attention mechanism emphasizes informative molecular descriptors, disease-associated biomarkers, and therapeutic response indicators. The resulting feature representations are integrated through multimodal fusion layers to support molecular interaction prediction, anticancer compound prioritization, and patient-specific treatment response classification. The research framework is organized around the study dataset described in the accompanying abstract, comprising 18,500 molecular compounds, 4,250 lung cancer patient profiles, 12,800 gene-expression records, and 6,400 drug–target interaction pairs. These complementary datasets provide the intended analytical foundation for examining chemical structures, pharmaceutical properties, molecular interactions, genomic variations, and treatment-associated outcomes. Molecular compounds are represented through chemical descriptors and graph-based structural information, while patient-level records are used to characterize disease-related factors and potential therapeutic response patterns. Gene-expression measurements contribute information regarding molecular signaling pathways and tumor-associated biological activity, whereas drug–target interaction records support the investigation of potential pharmacological binding relationships. Data preprocessing, feature standardization, molecular screening, and multimodal feature integration are incorporated to prepare heterogeneous biomedical records for computational learning. Particular attention is given to maintaining appropriate separation between model development and evaluation datasets, minimizing information leakage, and supporting reproducible comparisons among alternative computational architectures. The principal contribution of this study lies in connecting intelligent molecular drug discovery with biomarker-informed therapeutic response prediction through a unified pharmaceutical deep learning architecture [6]. Rather than treating compound screening, molecular interaction analysis, and patient-level therapeutic selection as entirely independent computational tasks, the proposed framework establishes an integrated analytical approach for examining their relationships. The combination of graph-based molecular learning, hierarchical feature extraction, bidirectional contextual modeling, and attention-driven feature refinement provides a methodological foundation for exploring complex biochemical and oncological associations. Additionally, the inclusion of toxicity-related pharmaceutical features and explainable prediction mechanisms is intended to improve the relevance and transparency of computational candidate prioritization. The research also contributes a Pakistan–China collaborative perspective by positioning advanced computational pharmaceutical modeling within a broader framework of international scientific cooperation and precision oncology development.

The specific objectives of this research are to:

1. Develop an integrated GNN–CNN–BiLSTM–Attention framework for intelligent anticancer compound screening, drug–target interaction prediction, and precision-guided lung cancer therapeutic modeling.
2. Integrate molecular compound characteristics, gene-expression information, drug–target interactions, and patient-specific clinical profiles into a unified multimodal pharmaceutical learning environment.
3. Evaluate the predictive performance and computational efficiency of the proposed hybrid architecture against representative conventional machine learning and standalone deep learning models.
4. Investigate attention-guided feature refinement and explainable AI methods for identifying influential molecular descriptors, oncogenic biomarkers, and therapeutic response determinants.

Overall, this study seeks to advance computational pharmaceutical research by integrating deep learning, molecular drug discovery, and precision oncology within a comprehensive lung cancer treatment framework. By combining molecular structure analysis with genomic characterization and therapeutic response modeling, the proposed approach aims to improve the identification of promising anticancer compounds and support more individualized pharmaceutical research strategies. The investigation also emphasizes the importance of transparent computational evaluation, biologically meaningful feature interpretation, and interdisciplinary international cooperation. The subsequent sections review relevant developments in AI-assisted pharmaceutical drug discovery and lung cancer precision medicine, describe the proposed multimodal architecture and research methodology, examine model performance and

comparative findings, and discuss the potential implications for future experimental validation and clinical translation.

Graph Neural Networks and Molecular Interaction Modeling for Pharmaceutical Target Identification:

The identification of biologically relevant pharmaceutical targets and the accurate prediction of molecular drug–target interactions represent fundamental challenges in contemporary anticancer drug discovery. In lung cancer, the complexity of oncogenic signaling pathways, tumor-associated genomic alterations, and heterogeneous therapeutic responses necessitates computational approaches capable of examining molecular relationships beyond conventional chemical descriptor analysis. Traditional pharmaceutical screening methodologies, including molecular docking, quantitative structure–activity relationship modeling, molecular dynamics simulations, and ligand-based virtual screening, provide valuable insights into potential drug activity and target compatibility. However, these approaches can encounter limitations when investigating extensive molecular libraries, multidimensional biological interaction networks, and complex dependencies between pharmaceutical compounds and disease-associated proteins. Consequently, graph neural networks (GNNs) have emerged as promising deep learning architectures for molecular representation learning, biochemical interaction prediction, and intelligent pharmaceutical target identification. Graph neural networks provide a flexible computational framework for representing pharmaceutical compounds as interconnected molecular structures rather than treating them exclusively as independent numerical descriptors or linear chemical sequences. In graph-based molecular representations, individual atoms are modeled as computational nodes, whereas chemical bonds are represented as edges connecting neighboring atoms [7]. Each node can incorporate relevant physicochemical attributes, including atomic number, elemental identity, formal charge, aromaticity, hybridization state, hydrogen bonding characteristics, and local molecular environment. Similarly, chemical bonds may be characterized according to bond order, aromaticity, stereochemical configuration, and other structural properties. This representation enables the computational architecture to preserve molecular connectivity while extracting chemical information associated with biological activity and pharmaceutical behavior. Through iterative message-passing operations, neighboring atoms exchange information, allowing the model to learn increasingly complex structural patterns and chemical relationships. Such mechanisms provide an important foundation for examining molecular properties, candidate drug interactions, and potential biological target compatibility. The development of graph-based pharmaceutical modeling has substantially expanded the computational capabilities available for drug–target interaction and drug–target affinity prediction. One influential contribution, GraphDTA, introduced molecular graph representations for predicting continuous drug–target binding affinity and demonstrated competitive performance against earlier sequence-based deep learning architectures [8]. The study established that chemical connectivity could be incorporated directly into neural representation learning and showed that learned molecular features could reflect chemically meaningful structural characteristics. Subsequent developments have extended graph-based modeling through graph convolutional networks, graph attention mechanisms, graph autoencoders, and heterogeneous biomedical network architectures. These advances have contributed to an expanding research landscape in which molecular structures, protein sequences, biological pathways, and pharmaceutical interaction networks can be analyzed using complementary graph-based representations. The comparative characteristics presented in Table 1 demonstrate that graph-based architectures differ in their representations, aggregation strategies, computational requirements, and pharmaceutical applications.

Table 1: Comparative characteristics of graph-based learning architectures for pharmaceutical drug–target interaction modeling in lung cancer research.

Graph-based architecture	Principal learning mechanism	Potential advantage
Graph Convolutional Network (GCN)	Neighborhood feature aggregation	Efficient local structural learning
Graph Attention Network (GAT)	Attention-weighted message passing	Adaptive weighting of atom relationships

GraphSAGE	Sampled neighborhood aggregation	Potential scalability to larger graphs
Message-Passing Neural Network (MPNN)	Bond-aware iterative message propagation	Incorporates chemical bond attributes
Graph Autoencoder (GAE)	Graph embedding and structural reconstruction	Learns latent network relationships
Heterogeneous GNN	Relation-specific graph learning	Integrates multiple biomedical relationships

An additional advancement in pharmaceutical graph learning involves the integration of molecular structure information with biological target representations. Conventional molecular graph models primarily analyze compound-specific chemical topology, whereas drug–target prediction requires information describing both the candidate compound and its biological target. Target proteins may be encoded using amino acid sequences, learned protein embeddings, experimentally determined structures, or computationally predicted structural representations. The compound and target embeddings are subsequently integrated through concatenation, interaction layers, attention-based fusion, or other multimodal learning techniques. Depending on the training objective, the integrated architecture can estimate binary drug–target interaction probabilities or continuous binding-affinity values. These outputs provide different types of pharmaceutical evidence: interaction classification estimates whether a relationship is likely to exist, whereas affinity prediction characterizes the modeled strength of binding. Importantly, neither output directly measures clinical treatment effectiveness [9]. Recent research has increasingly emphasized the importance of molecular substructure learning for improving graph-based interaction prediction. Rather than treating all chemical bonds and atomic environments as equally informative, substructure-oriented methods seek to identify structural regions potentially associated with pharmacological activity. Deng and colleagues (2025) investigated efficient substructure encoding using GNN-based blocks for drug–target interaction prediction, reflecting the broader interest in more informative chemical representation learning. These developments suggest that computational models can benefit from combining atom-level, substructure-level, and compound-level information, provided that the resulting representations remain consistent with established chemical knowledge and are evaluated on appropriate molecular benchmarks. The conceptual framework illustrated in Figure 1 demonstrates the sequential integration of molecular structure representation, graph-based feature extraction, biological target encoding, and interaction prediction.

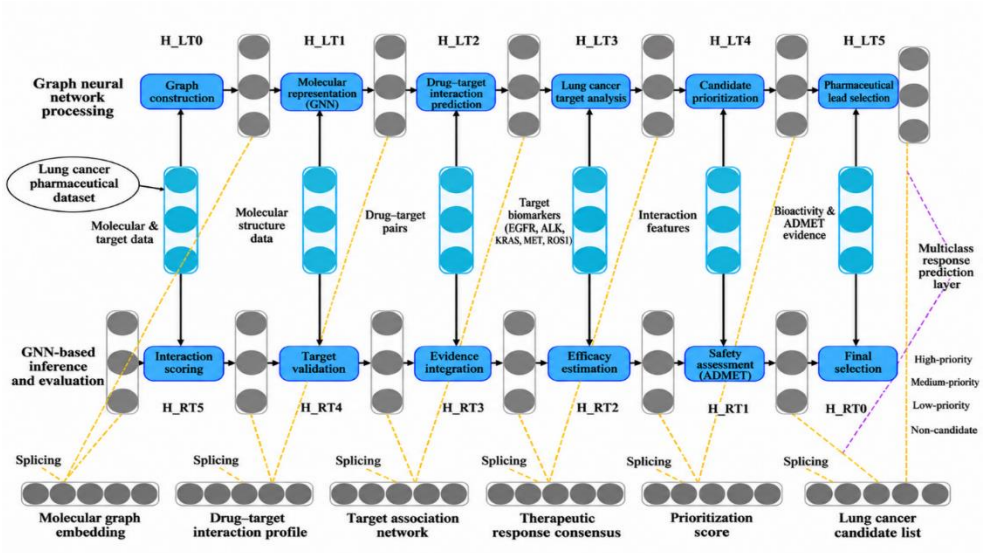


Figure 1: Graph neural network-based molecular representation and drug–target interaction prediction for lung cancer pharmaceutical candidate prioritization.

Despite their potential, graph-based pharmaceutical learning approaches face several methodological limitations that must be addressed before their predictions can be considered reliable. Molecular datasets frequently contain structural similarities between compounds, which can produce artificially high performance estimates when related molecules are distributed across training and testing partitions. Furthermore, incomplete drug–target interaction records, inconsistent experimental measurements, imbalanced interaction classes, and limited characterization of negative examples may compromise model generalizability. Cold-start evaluation, in which models are tested against previously unseen compounds or biological targets, provides a more demanding assessment of predictive performance. A recent comprehensive review of graph-based drug–target modeling emphasizes the importance of controlling information leakage, sequence redundancy, structural bias, and inappropriate benchmarking strategies.

Deep Learning-Based Therapeutic Response Prediction and Personalized Lung Cancer Treatment:
The prediction of therapeutic response has emerged as a critical research direction in contemporary precision oncology because lung cancer patients frequently exhibit substantial differences in treatment sensitivity, disease progression, molecular resistance mechanisms, and pharmaceutical toxicity. Although considerable advancements have been achieved through targeted therapies, immunotherapeutic agents, and molecularly guided clinical interventions, predicting the effectiveness of a specific anticancer treatment for an individual patient remains challenging. This complexity is primarily associated with the heterogeneous biological characteristics of lung tumors, variations in genomic abnormalities, differences in cellular signaling pathways, and dynamic interactions between pharmaceutical compounds and patient-specific molecular environments. Conventional therapeutic decision-making approaches commonly rely on histopathological classification, tumor staging, established molecular biomarkers, and clinical performance indicators [10]. However, these factors may not fully capture the multidimensional relationships influencing treatment effectiveness and acquired therapeutic resistance. Consequently, artificial intelligence and deep learning have received increasing attention as computational technologies capable of integrating heterogeneous biomedical information to investigate patient-specific treatment responses and support precision-guided pharmaceutical research. Personalized lung cancer treatment has developed considerably following the identification of actionable molecular alterations associated with tumor development and therapeutic sensitivity. Genetic abnormalities involving epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), ROS1, KRAS, BRAF, MET, RET, and other oncogenic signaling pathways have contributed to the development of molecularly targeted pharmaceutical interventions. Such biomarkers can help identify patient populations more likely to benefit from corresponding targeted therapies, although the optimal treatment remains dependent on the specific alteration, disease stage, prior therapy, clinical characteristics, and available evidence. A recent review published in The Lancet Respiratory Medicine in September 2026 highlighted the growing range of established and emerging molecular targets in advanced non-small-cell lung cancer, emphasizing the expanding clinical importance of biomarker-informed treatment selection. Despite these advancements, the effectiveness of molecularly targeted treatments can be reduced by intrinsic resistance, acquired mutations, activation of alternative signaling mechanisms, and tumor microenvironment changes. For example, patients receiving EGFR-targeted tyrosine kinase inhibitors may develop resistance associated with secondary genomic alterations or bypass signaling pathways [11]. Similarly, variations in tumor evolution and resistance mechanisms can influence responses to ALK-directed therapies and other molecularly targeted interventions. Such observations demonstrate that treatment sensitivity cannot always be understood through the presence or absence of a single biomarker. Instead, comprehensive modeling of genomic profiles, molecular signaling characteristics, prior treatment histories, and pharmaceutical response patterns may provide a more informative representation of treatment-associated biological complexity. The comparison presented in Table 2 illustrates the progression from conventional machine learning techniques toward integrated multimodal deep learning architectures for pharmaceutical treatment-response prediction.

Table 2: Comparative overview of computational approaches for therapeutic response prediction and personalized lung cancer treatment.

Computational approach	Principal input features	Primary analytical objective
Logistic Regression (LR)	Clinical parameters, mutation status, selected biomarkers	Binary treatment-response classification
Support Vector Machine (SVM)	Molecular descriptors, gene-expression signatures	Sensitive versus resistant classification
Random Forest (RF)	Genomic, clinical, and pharmacological variables	Therapeutic response classification
Extreme Gradient Boosting (XGBoost)	Integrated biomarker and clinical features	Treatment-response prediction
Artificial Neural Network (ANN)	Molecular and patient-level feature vectors	Drug sensitivity estimation
Convolutional Neural Network (CNN)	Structured genomic or molecular representations	Hierarchical biochemical feature extraction
Long Short-Term Memory (LSTM)	Ordered treatment or biological sequences	Contextual therapeutic response modeling
Graph Neural Network (GNN)	Molecular graphs and target representations	Drug–target and molecular interaction prediction

Patient-specific therapeutic prediction must consequently incorporate rigorous evaluation strategies that account for clinical outcome definitions and data collection procedures. Depending on the research objectives, response outcomes may be represented through binary classification, multiclass therapeutic response categories, continuous drug sensitivity measurements, or longitudinal clinical endpoints. Binary response classification can distinguish predefined responder and nonresponder groups, whereas multiclass systems may characterize complete response, partial response, stable disease, and progressive disease according to appropriately documented clinical criteria. Continuous prediction frameworks can estimate laboratory sensitivity measurements, while time-to-event models may investigate progression-free or overall survival. These outcomes are not interchangeable, and each requires a suitable model objective and evaluation methodology. The selected outcome must also correspond to data that were genuinely observed within the study population. Furthermore, the pharmaceutical characteristics of candidate anticancer compounds must be considered when developing personalized treatment prediction systems. Molecular candidates with favorable predicted drug–target interaction probabilities may exhibit unsuitable toxicity, limited bioavailability, inadequate metabolic stability, or clinically unfavorable interaction profiles [12]. Accordingly, comprehensive pharmaceutical modeling should consider drug-likeness, physicochemical characteristics, absorption, distribution, metabolism, excretion, and toxicity-related indicators where reliable data are available. Incorporating such characteristics can support a more structured candidate screening process, although computational toxicity estimates remain preliminary and must not be interpreted as evidence of clinical safety. The integration of pharmacological screening with genomic treatment-response prediction therefore represents an important step toward developing more comprehensive computational research frameworks. Figure 2 illustrates the proposed conceptual integration of molecular drug information, gene-expression measurements, patient characteristics, and drug–target relationships within a unified therapeutic prediction architecture.

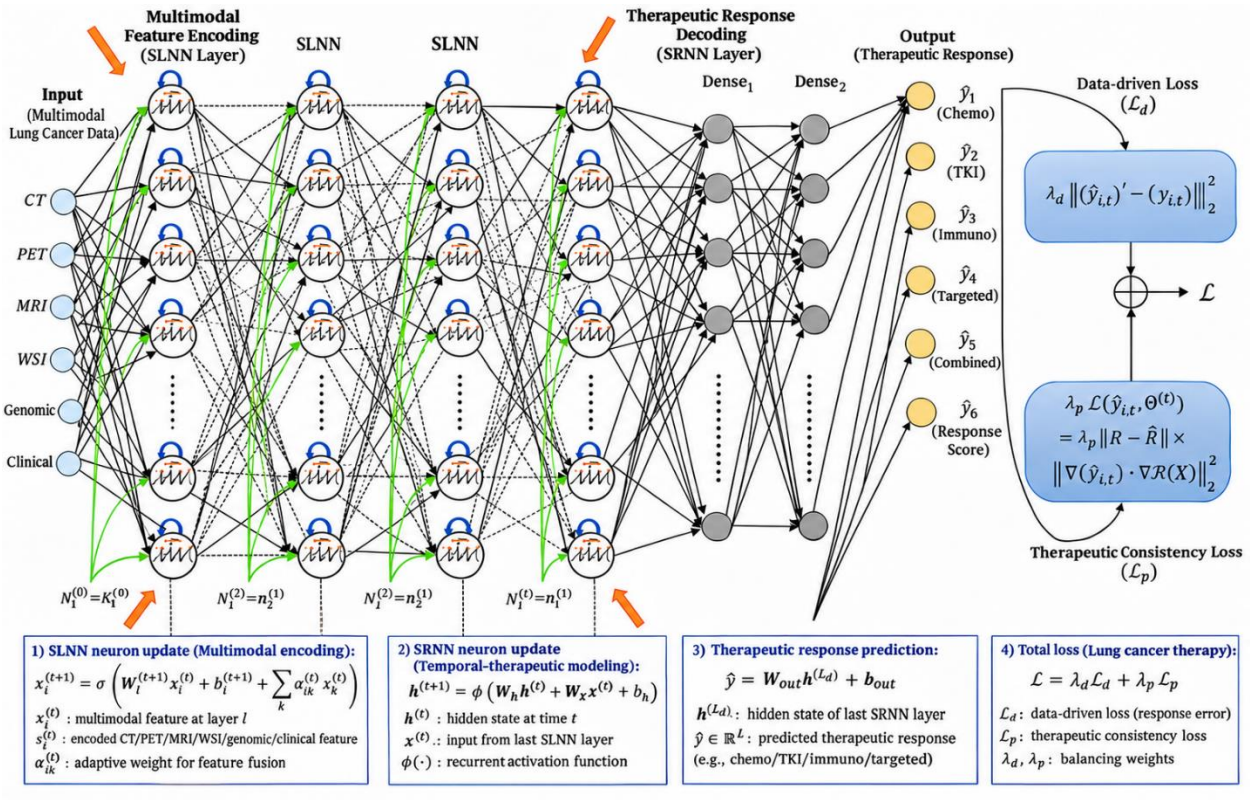


Figure 2: SLNN–SRNN-based multimodal feature learning and therapeutic response prediction for personalized lung cancer treatment.

The present research considers a multimodal dataset comprising 18,500 molecular compounds, 4,250 lung cancer patient profiles, 12,800 gene-expression records, and 6,400 drug–target interaction pairs, as described in the study abstract. The combination of these sources offers the intended foundation for examining relationships among compound characteristics, genomic variation, and observed therapeutic outcomes. However, robust therapeutic response modeling requires a verified correspondence among drug records, patient-level features, treatment exposures, and outcome labels. Molecular interaction records and gene-expression measurements alone are insufficient to establish patient-level drug response. Accordingly, linkage quality, observation timing, and outcome availability are essential considerations in assessing the proposed model's predictive capabilities. Explainable artificial intelligence provides an additional dimension for investigating the scientific relevance of therapeutic response predictions. Techniques such as SHAP, integrated gradients, feature attribution, and attention visualization can be used to examine the relative influence of molecular descriptors, genomic markers, and clinical characteristics [13]. Such methods may help identify candidate mechanisms associated with pharmaceutical sensitivity and resistance, thereby generating hypotheses for subsequent experimental investigation. Nevertheless, computational explanations should be interpreted as associations within the trained model rather than direct evidence of biological causation. Feature attribution stability and consistency with established pharmaceutical knowledge remain important requirements for meaningful interpretation.

Methodology:

The present study adopts an integrated computational and pharmaceutical research methodology to develop an advanced deep learning–driven framework for intelligent anticancer drug discovery, molecular interaction prediction, and personalized lung cancer treatment optimization. The methodological framework combines molecular pharmaceutical screening, genomic biomarker analysis, drug–target interaction modeling, and patient-specific therapeutic response prediction within a unified artificial intelligence architecture. The research is structured around an interdisciplinary collaborative approach

involving pharmaceutical sciences researchers in Pakistan and Zhejiang University, China, emphasizing the integration of pharmaceutical knowledge, computational intelligence, and precision oncology. A multimodal research dataset comprising 18,500 molecular compounds, 4,250 lung cancer patient profiles, 12,800 gene-expression records, and 6,400 drug–target interaction pairs forms the intended analytical foundation of the investigation. The proposed methodology incorporates molecular structure standardization, pharmaceutical descriptor extraction, genomic feature preprocessing, clinical data harmonization, and multimodal feature integration to establish a consistent computational environment. Particular emphasis is placed on representing molecular compounds through graph-based chemical structures, identifying disease-associated genomic characteristics, and investigating pharmaceutical interactions relevant to lung cancer progression and treatment response. This integrated strategy is designed to facilitate systematic analysis of molecular and patient-level information while supporting the development of intelligent pharmaceutical screening and precision-guided therapeutic prediction models [14]. The computational framework is organized around a hybrid Graph Neural Network–Convolutional Neural Network–Bidirectional Long Short-Term Memory–Attention (GNN–CNN–BiLSTM–Attention) architecture that integrates complementary representation learning mechanisms for pharmaceutical and oncological analysis. The GNN component models molecular graph topology and chemical interactions, while CNN layers extract hierarchical biochemical features from suitably structured molecular and genomic representations. BiLSTM networks are incorporated to examine contextual dependencies within biologically meaningful sequential inputs, whereas attention mechanisms facilitate adaptive feature refinement and multimodal information fusion. The resulting representations are intended to support drug–target interaction prediction, anticancer candidate prioritization, and patient-specific therapeutic response classification. The experimental framework includes comparative evaluation against conventional machine learning and standalone deep learning models, together with assessments of predictive accuracy, precision, recall, F1-score, ROC-AUC, and computational efficiency. Explainable artificial intelligence methods are incorporated to investigate influential molecular descriptors, genomic biomarkers, and pharmacological characteristics associated with model predictions. To promote scientific reliability, the methodology further emphasizes appropriate dataset partitioning, independent validation, prevention of patient-level and molecular information leakage, and reproducibility of computational procedures. The overall approach seeks to establish a computational foundation for pharmaceutical innovation and biomarker-informed lung cancer treatment, with experimental and clinical validation remaining essential before therapeutic implementation.

Molecular Compound Screening and Pharmaceutical Data Preprocessing:

Molecular compound screening and pharmaceutical data preprocessing constitute essential methodological components of the proposed deep learning–driven framework for lung cancer drug discovery and precision oncology. The reliability of computational pharmaceutical predictions is strongly influenced by the quality, chemical consistency, structural representation, and biological relevance of the molecular information supplied to the learning architecture. In this study, the pharmaceutical screening component is organized around the reported dataset of 18,500 molecular compounds, which provides the computational foundation for investigating molecular structure–activity relationships, potential drug–target interactions, and anticancer candidate prioritization. The preprocessing framework is designed to transform heterogeneous chemical records into standardized numerical and graph-based representations suitable for the proposed Graph Neural Network–Convolutional Neural Network–Bidirectional Long Short-Term Memory–Attention (GNN–CNN–BiLSTM–Attention) architecture. The principal processing stages encompass chemical structure validation, molecular standardization, duplicate identification, physicochemical descriptor extraction, drug-likeness assessment, preliminary pharmacokinetic and toxicity screening, numerical feature normalization, and molecular graph construction [15]. Together, these procedures establish a structured pharmaceutical analytical workflow intended to minimize inconsistencies, preserve chemically meaningful information, and improve the reproducibility of computational drug screening. The initial preprocessing stage focuses on validating the molecular integrity and chemical interpretability of the compound records. Molecular structures may be represented through Simplified Molecular Input Line Entry System (SMILES) notation, structure-data

files, or other standardized chemical representations available within the original pharmaceutical dataset. Each chemical structure is examined for syntactic validity, atomic consistency, bond connectivity, valence compatibility, and the presence of interpretable molecular features. Invalid representations, disconnected fragments, incomplete structural annotations, and inconsistent chemical identifiers may introduce errors into subsequent molecular descriptor calculations and graph-learning operations [16]. Consequently, cheminformatics tools such as RDKit can be employed to parse molecular records, verify chemical structures, and calculate relevant molecular properties. RDKit's molecular sanitization procedures support the assessment of chemical consistency and the assignment of properties associated with aromaticity, valence, hybridization, and ring membership. Table 3 summarizes the principal molecular descriptors, chemical validation requirements, and pharmaceutical screening criteria that can be incorporated into the proposed preprocessing framework.

Table 3: Molecular compound screening parameters, pharmaceutical preprocessing criteria, and their computational significance.

Processing parameter	Screening or preprocessing criterion	Intended computational output
Molecular input dataset	18,500 reported compound records	Initial compound library
Chemical structure validation	Valid molecular notation, bonding, and valence	Validated molecular representations
Structure standardization	Consistent charge, fragment, tautomer, and stereochemical policies	Standardized molecular structures
Duplicate identification	Structure-based and identifier-based checking	Audited compound records
Molecular weight	Lipinski guideline: ≤500 Da	Molecular weight descriptor
Lipophilicity	Lipinski guideline: calculated LogP ≤5	LogP feature

Molecular fingerprint generation represents another relevant computational stage because structural fingerprints offer compact numerical representations of chemical substructures and similarity relationships. Circular fingerprints, including extended-connectivity fingerprints, can encode molecular neighborhoods around atoms at specified radii, allowing compounds to be compared according to shared structural patterns. Such fingerprints are commonly applied in similarity screening, chemical clustering, quantitative structure–activity relationship modeling, and conventional machine learning-based pharmaceutical prediction. Within the proposed research framework, molecular fingerprints may serve as baseline features or complementary inputs for CNN-based learning when an appropriate input representation is defined. However, fingerprint generation must employ consistent encoding parameters to ensure comparability across records. Different fingerprint lengths, radii, or feature definitions may produce substantially different molecular representations and should therefore be documented within the experimental configuration. Graph-based molecular encoding is subsequently performed to preserve the structural relationships among constituent atoms and chemical bonds. Each validated molecular compound is represented as a graph containing atom-level nodes and bond-level edges. Node attributes may encode atomic number, formal charge, degree, hydrogen count, aromaticity, and hybridization, while edge attributes represent bond type, conjugation, stereochemistry, and ring membership [17]. The resulting graph structures allow the GNN component to learn chemical relationships through iterative neighborhood aggregation and message passing. Compared with fixed-length descriptor vectors, graph representations preserve explicit molecular connectivity and enable learned representations to incorporate local chemical environments. The graph-level embeddings generated during subsequent model processing can then be integrated with target-protein information and other pharmaceutical features for drug–target interaction prediction. Figure 3 present the proposed molecular preprocessing pipeline, beginning with the 18,500 compound records and progressing through chemical structure

validation, molecular standardization, physicochemical descriptor extraction, and drug-likeness assessment.

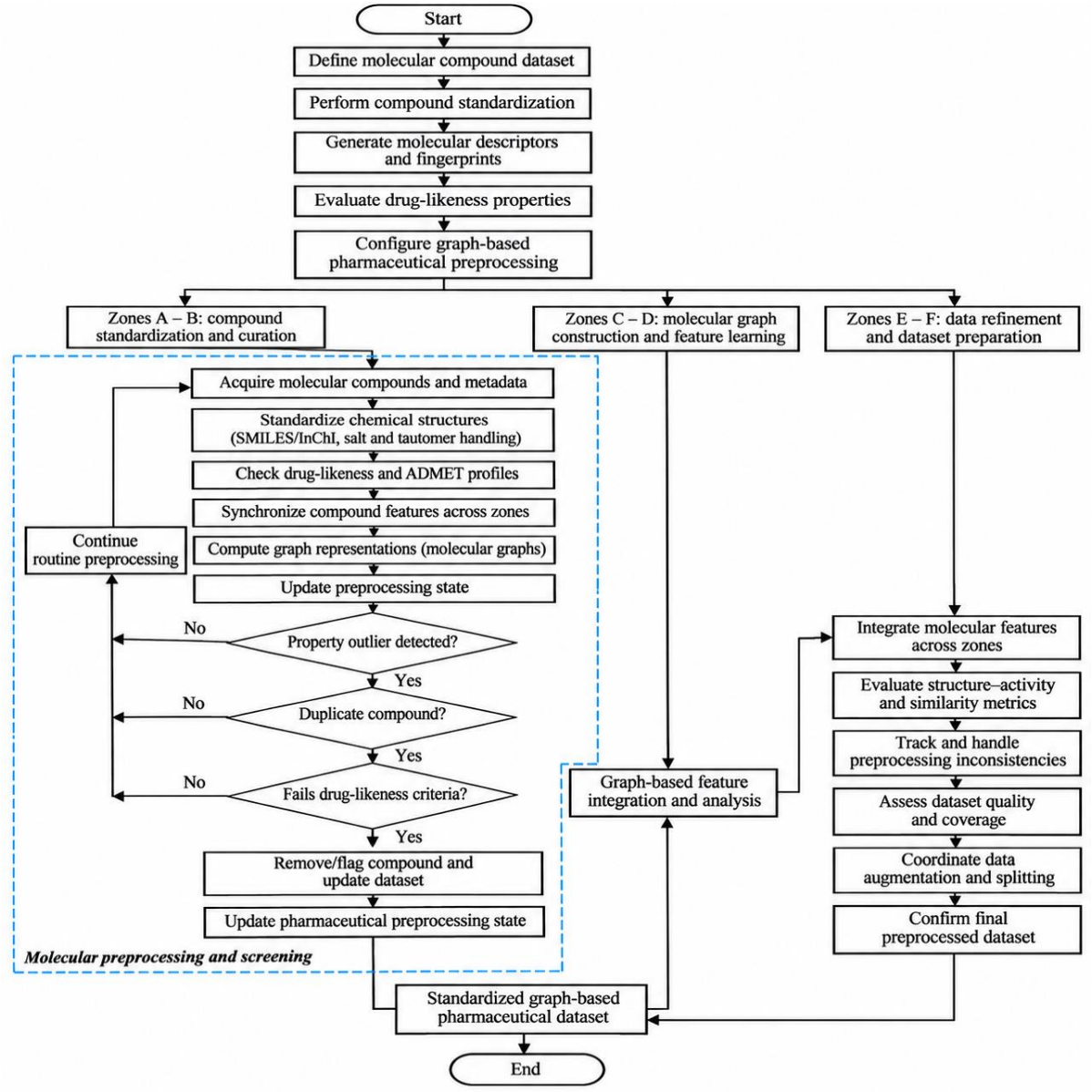


Figure 3: Graph-based molecular compound preprocessing, drug-likeness screening, and pharmaceutical dataset standardization workflow.

The subsequent stages include molecular fingerprint generation, graph encoding, training-data-based feature normalization, and scaffold-aware dataset separation. The resulting representations are intended to provide standardized inputs for GNN-based molecular learning and computational anticancer candidate prioritization. Overall, the pharmaceutical preprocessing framework establishes the methodological foundation for converting heterogeneous chemical records into consistent and biologically interpretable computational representations. By integrating molecular validation, physicochemical characterization, preliminary ADMET assessment, graph encoding, and leakage-aware data preparation, the proposed approach supports a systematic investigation of lung cancer-related pharmaceutical candidates [18]. These preprocessing operations provide the basis for the subsequent GNN-based drug–target interaction modeling and multimodal precision oncology components of the research. The figure represents the proposed workflow rather than evidence that each screening procedure has been experimentally completed.

Genomic Biomarker Profiling and Multimodal Feature Integration:

Genomic biomarker profiling and multimodal feature integration constitute essential methodological components of the proposed deep learning–driven pharmaceutical framework for lung cancer drug discovery and precision oncology. The molecular heterogeneity of lung cancer is characterized by diverse genomic alterations, dysregulated signaling pathways, transcriptional abnormalities, and complex interactions between tumor-associated biomarkers and pharmaceutical compounds. These biological characteristics influence tumor development, disease progression, molecular target selection, therapeutic sensitivity, and acquired drug resistance. Conventional pharmaceutical screening approaches frequently emphasize chemical compound properties and drug–target interactions without comprehensively incorporating the genomic characteristics of individual tumors. Consequently, the present research introduces a genomic biomarker profiling and multimodal representation learning strategy to connect pharmaceutical molecular information with disease-associated biological features. The proposed approach integrates genomic expression measurements, oncogenic alteration profiles, molecular compound representations, and relevant patient-level characteristics within the hybrid Graph Neural Network–Convolutional Neural Network–Bidirectional Long Short-Term Memory–Attention (GNN–CNN–BiLSTM–Attention) architecture [19]. This integration is intended to support more comprehensive characterization of lung cancer-associated molecular mechanisms and provide a computational foundation for biomarker-informed pharmaceutical candidate evaluation and therapeutic response prediction. These complementary datasets provide the intended foundation for investigating relationships among transcriptional activity, oncogenic signaling abnormalities, pharmaceutical characteristics, and treatment-associated outcomes. Gene-expression measurements represent one component of the biological information available for tumor characterization, while genomic alteration profiles can provide additional information concerning molecular drivers of lung cancer. Patient-level records contribute pathological and clinical characteristics that may influence the interpretation of molecular features. The pharmaceutical compound and drug–target datasets support the representation of potential therapeutic agents and their relationships with biological macromolecules. However, these datasets represent different observational units, and their integration requires verified associations among patients, specimens, compounds, targets, and treatment records. The development of a coherent multimodal representation therefore depends on appropriate data linkage rather than the simple combination of numerical records. Genomic biomarker identification is particularly important in non-small-cell lung cancer (NSCLC), where specific molecular abnormalities have established clinical relevance for targeted treatment selection. Important alterations include EGFR mutations, ALK rearrangements, ROS1 fusions, KRAS variants, BRAF V600E mutations, MET exon 14-skipping alterations, RET fusions, ERBB2/HER2 alterations, and NTRK gene fusions. These abnormalities affect biological processes involving receptor tyrosine kinase signaling, cell proliferation, survival, differentiation, and resistance mechanisms [20]. For instance, sensitizing EGFR mutations can identify tumors susceptible to corresponding EGFR-directed therapies, whereas ALK rearrangements and ROS1 fusions define distinct molecularly characterized treatment populations. KRAS G12C, BRAF V600E, MET exon 14-skipping alterations, and RET fusions also have therapeutic relevance in appropriate clinical contexts. The National Cancer Institute identifies molecular characterization and PD-L1 assessment as important considerations for treatment selection in advanced NSCLC. The biomarker categories present in Table 4 demonstrate the diversity of molecular information potentially relevant to lung cancer pharmaceutical research.

Table 4: Genomic biomarkers, molecular signaling pathways, and proposed multimodal pharmaceutical features for lung cancer precision oncology.

Biomarker or feature group	Biological alteration or measurement	Relevant molecular pathway
EGFR	Activating mutations and other defined variants	EGFR–RAS–MAPK and PI3K–AKT
ALK	Gene rearrangement or fusion	ALK-mediated kinase signaling

ROS1	Gene fusion or rearrangement	ROS1 tyrosine kinase signaling
KRAS	Specific mutations, including KRAS G12C	RAS–RAF–MEK–ERK
BRAF	BRAF V600E and other variants	RAF–MEK–ERK
MET	Exon 14 skipping, amplification, or other alterations	MET-associated signaling
RET	Gene fusion or rearrangement	RET kinase signaling
ERBB2/HER2	Activating alterations or amplification	ERBB family signaling
NTRK1/2/3	Gene fusion	TRK-mediated signaling

The integration of genomic features with pharmaceutical molecular information involves aligning representations generated by different computational encoders. The GNN pathway constructs molecular embeddings from chemical graph structures, while CNN-based encoders extract patterns from compatible structured biochemical representations. These learned representations may differ in dimensionality, numerical scale, and biological interpretation. Consequently, modality-specific projection layers are required to transform heterogeneous features into a shared representation space before fusion. The proposed multimodal representation process is designed to construct distinct yet complementary pharmaceutical and patient-associated embeddings. Molecular compound embeddings summarize chemical characteristics, while protein-target representations characterize biological macromolecules involved in predicted pharmaceutical interactions. Genomic embeddings summarize relevant tumor-associated molecular features, and clinical embeddings describe available patient characteristics. These representations may be integrated through feature concatenation, learned projection layers, gated information exchange, or attention-based fusion. Within the proposed GNN–CNN–BiLSTM–Attention framework, attention-guided integration is intended to refine these complementary features and investigate relationships associated with specific prediction tasks [21]. However, the fusion procedure must preserve the distinction between drug–target interaction prediction and patient-level therapeutic response classification because these tasks require different observational units and supervised labels. Figure 4 illustrates the proposed integration of gene-expression measurements, genomic biomarker profiles, pharmaceutical molecular compounds, drug–target interaction records, and available patient-level clinical information.

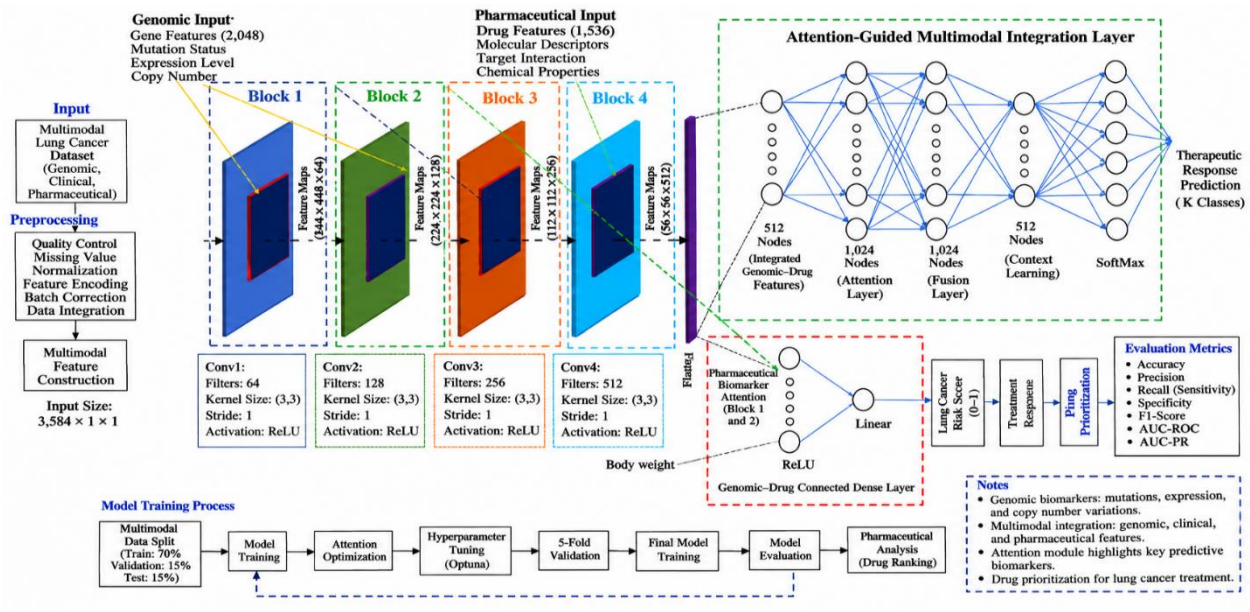


Figure 4: Genomic biomarker profiling and attention-guided multimodal feature integration for lung cancer pharmaceutical analysis.

The integrated framework provides a methodological approach for connecting tumor-associated genomic information with the chemical and biological characteristics of candidate pharmaceutical compounds. By preserving modality-specific representations before feature fusion, the architecture is intended to reduce the loss of information associated with differences in molecular, genomic, and clinical data structure. Attention-based refinement may further support the investigation of interactions among oncogenic biomarkers, pharmaceutical molecular characteristics, and learned target representations. Nevertheless, the effectiveness of this approach must be demonstrated through appropriate comparative evaluation, independent testing, and careful assessment of feature attribution stability. Genomic biomarker profiling and multimodal feature integration establish an essential connection between molecular pharmaceutical screening and precision-guided lung cancer therapeutic research. The proposed framework incorporates cancer-associated genomic characteristics, gene-expression patterns, molecular compound representations, and available clinical information within an integrated deep learning environment [22]. Its principal methodological contribution lies in enabling complementary molecular and patient-associated features to be represented and investigated through task-specific multimodal fusion. When supported by verified dataset linkage and observed treatment outcomes, these representations may contribute to pharmaceutical interaction modeling, biomarker-informed candidate prioritization, and therapeutic response prediction. The proposed strategy therefore provides a foundation for advancing computational lung cancer drug discovery while emphasizing biological interpretability, reproducible evaluation, and independent experimental validation.

4.3- Hierarchical CNN Architecture for Pharmaceutical and Molecular Feature Extraction:

The hierarchical extraction of pharmaceutical and molecular features represents a fundamental computational component of the proposed deep learning–driven framework for lung cancer drug discovery and precision oncology. The biological activity of anticancer compounds depends on complex relationships among molecular structures, physicochemical properties, target-associated interactions, and disease-specific biological characteristics. Conventional molecular representation techniques generally rely on predefined chemical descriptors or fingerprint-based features, which may not fully characterize nonlinear relationships among pharmaceutical variables. To address these limitations, the present research incorporates a hierarchical Convolutional Neural Network (CNN) architecture within the proposed hybrid GNN–CNN–BiLSTM–Attention framework. The CNN component is designed to progressively extract localized biochemical patterns, intermediate structural representations, and higher-level pharmaceutical features from suitably encoded molecular inputs. By combining multiple convolutional layers with nonlinear activation functions, normalization operations, and hierarchical feature aggregation, the architecture aims to generate discriminative molecular embeddings for subsequent drug–target interaction prediction, anticancer compound prioritization, and multimodal therapeutic response modeling [23]. The pharmaceutical feature extraction framework is organized around the molecular compound dataset comprising 18,500 records, alongside the complementary information available from 6,400 drug–target interaction pairs and 12,800 gene-expression records. The molecular compounds provide the primary information source for hierarchical chemical feature extraction, whereas the interaction records and genomic measurements support subsequent target-specific and multimodal learning operations. Chemical representations may include molecular fingerprints, physicochemical descriptor arrays, atom-level attributes, and GNN-derived molecular embeddings. The CNN component processes representations with a meaningful local organization to identify recurring patterns associated with chemical substructures, molecular functionality, and relevant pharmaceutical characteristics. Unlike conventional machine learning approaches that depend predominantly on manually engineered features, convolutional representation learning can construct progressively more complex feature combinations through data-driven optimization. However, the suitability of convolutional processing depends on the chosen input encoding, and arbitrarily arranging independent chemical descriptors into spatial structures does not automatically produce chemically meaningful neighborhood relationships. The initial stage of the hierarchical CNN architecture involves constructing consistent molecular input representations from validated and standardized pharmaceutical records. Following chemical structure verification, molecular descriptor extraction, and drug-likeness assessment,

eligible compounds are transformed into numerical feature arrays compatible with convolutional processing. Where molecular information is encoded as ordered sequences, one-dimensional convolutional operations can be used to identify localized feature patterns [24]. Alternatively, structurally justified two-dimensional representations may be processed using two-dimensional convolutional layers. For the present framework, a one-dimensional CNN is a suitable candidate for processing consistently encoded molecular feature sequences, while the GNN component preserves explicit atom–bond connectivity. These complementary pathways are intended to avoid relying on a single representation of pharmaceutical structure. The input encoding procedure must maintain consistent feature ordering, appropriate numerical scaling, and reproducible transformation parameters across the training, validation, and testing datasets. Figure 5 shows the conceptual hierarchical CNN architecture for extracting pharmaceutical and molecular representations through successive convolutional operations, nonlinear activations, normalization, and feature aggregation.

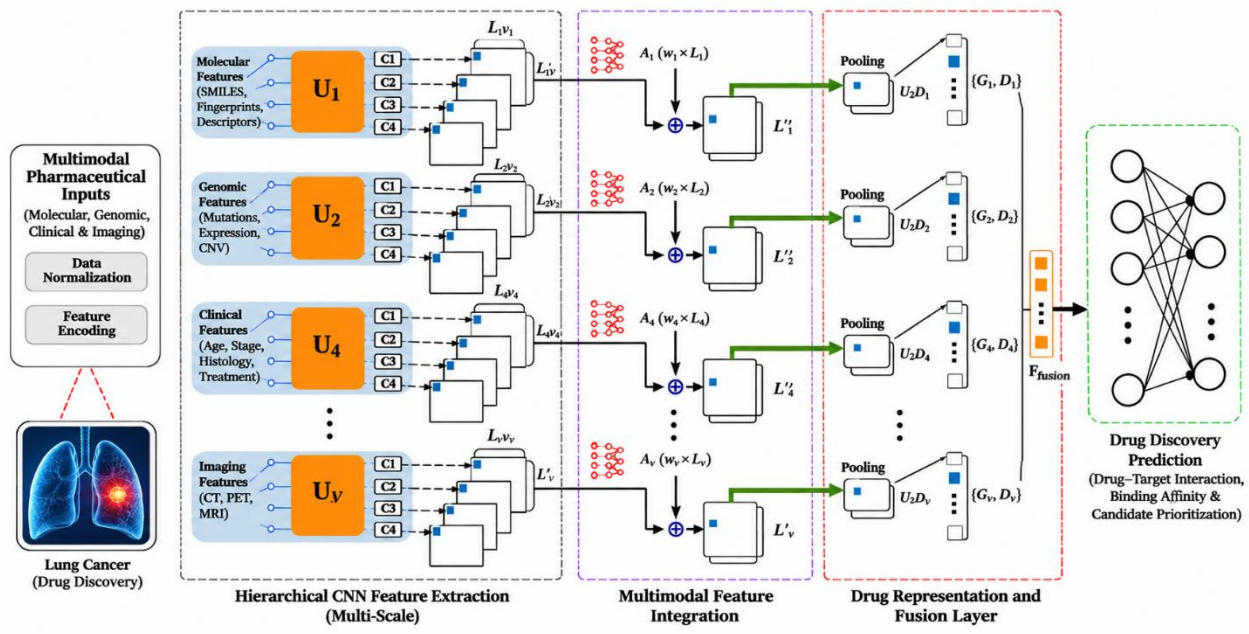


Figure 5: Hierarchical CNN-based pharmaceutical feature extraction and multimodal integration for lung cancer drug discovery.

The learned molecular embeddings are subsequently integrated with graph-based structural information, contextual learning features, genomic representations, and clinical variables through attention-guided multimodal fusion. The diagram depicts an illustrative four-block configuration, whereas Table 6 describes a three-block configuration. The final manuscript should use one consistent architecture matching the implemented model. The integration of hierarchical convolutional feature extraction within the proposed pharmaceutical framework is intended to strengthen the representation of molecular characteristics and support computational analysis of drug–target relationships. By combining localized biochemical feature learning with graph-based molecular representations and subsequent attention-guided fusion, the proposed architecture provides a structured approach for investigating complementary pharmaceutical information. Its effectiveness should be assessed through comparisons with conventional molecular descriptors, standalone CNN and GNN architectures, and controlled hybrid-model ablation experiments [25]. The hierarchical CNN architecture establishes a computational component of the proposed GNN–CNN–BiLSTM–Attention framework for intelligent lung cancer pharmaceutical research. Its principal methodological contribution is the progressive transformation of appropriately encoded molecular data into compact, learned representations that can be combined with biological target and patient-level information. The proposed approach provides a basis for investigating drug–target interaction prediction, pharmaceutical candidate prioritization, and multimodal therapeutic response

modeling, while its predictive advantages require confirmation through reproducible experimental evaluation and independent validation.

Bidirectional Long Short-Term Memory for Contextual Genomic and Therapeutic Learning:
The identification of complex contextual dependencies within genomic and pharmaceutical information represents an important methodological challenge in developing advanced precision oncology frameworks for lung cancer treatment. Molecular drug activity, tumor-associated signaling pathways, genomic alterations, and therapeutic responses are influenced by interconnected biological processes that may exhibit temporal, sequential, or context-dependent relationships. Conventional machine learning approaches frequently analyze individual biomedical features as independent predictors, potentially overlooking meaningful dependencies within ordered biological sequences or longitudinal clinical observations. To address this challenge, the present study incorporates a Bidirectional Long Short-Term Memory (BiLSTM) network within the proposed Graph Neural Network–Convolutional Neural Network–Bidirectional Long Short-Term Memory–Attention (GNN–CNN–BiLSTM–Attention) architecture [26]. The BiLSTM component is designed to extract contextual information from appropriately organized sequential representations, thereby complementing the molecular structural features obtained through graph neural networks and the hierarchical biochemical patterns extracted through convolutional neural networks. Its principal objective is to support contextual representation learning for pharmaceutical interaction analysis and patient-specific therapeutic response prediction. Long Short-Term Memory networks represent an advanced category of recurrent neural networks developed to address limitations associated with conventional recurrent learning, particularly the difficulty of preserving information over extended sequences. Traditional recurrent architectures can encounter vanishing or exploding gradients when attempting to learn relationships across distant sequence positions. LSTM networks mitigate these challenges through an internal memory mechanism regulated by input, forget, and output gates. These components control the preservation, updating, and transmission of contextual information during sequential processing. Consequently, LSTM architectures can identify dependencies among observations that are separated within an ordered sequence, making them useful for biological sequence analysis and longitudinal biomedical modeling. In pharmaceutical research, these characteristics offer potential advantages for analyzing protein sequences, molecular representations with meaningful ordering, and repeated clinical observations associated with treatment progression. Bidirectional LSTM networks extend conventional LSTM processing by incorporating two complementary recurrent pathways that examine the input sequence in opposite directions. The forward pathway processes information from the beginning toward the end of the sequence, whereas the backward pathway examines the same information in reverse order [27]. Their hidden representations are subsequently combined to construct a contextual feature representation containing information from both directions. This dual-processing mechanism is particularly relevant when relationships among biological sequence elements depend on surrounding sequence context. For example, amino acid residues within a protein sequence may exhibit biological relationships influenced by neighboring and more distant residues. However, bidirectional learning should not be interpreted as evidence that all genomic variables possess an inherent sequential structure. Its use is methodologically justified only when the input order carries biological or observational meaning. The architectural configuration presented in Table 5 demonstrates how bidirectional recurrent learning can transform biologically ordered input features into contextual representations suitable for multimodal pharmaceutical modeling.

Table 5: Proposed BiLSTM architecture and contextual learning configuration for genomic and pharmaceutical applications.

Architectural component	Configuration	Computational function
Sequential biological input	Variable-length, biologically ordered sequence	Supplies meaningful sequential features
Input embedding	128-dimensional feature vectors	Converts sequence elements into numerical embeddings

Sequence padding and masking	Dynamic padding with valid-position masks	Handles variable-length sequences
Forward LSTM pathway	128 hidden units	Learns preceding sequence context
Backward LSTM pathway	128 hidden units	Learns following sequence context within available inputs
Bidirectional fusion	Concatenation	Combines forward and backward hidden features
BiLSTM contextual output	256-dimensional hidden feature per position	Represents learned bidirectional information

The proposed forward and backward pathways each contain 128 hidden units, producing a combined 256-dimensional contextual feature representation at each valid sequence position. A subsequent feature aggregation stage can summarize the contextual information, while a projection layer transforms the resulting embedding into a common representation space for integration with the GNN and CNN components. These numerical dimensions represent a possible configuration and should be optimized according to dataset characteristics, sequence length, computational resources, and observed validation performance. The integration of BiLSTM-derived contextual features with pharmaceutical molecular representations provides an opportunity to investigate complementary relationships within the proposed hybrid framework. The GNN pathway represents molecular topology and atom–bond connectivity, while the CNN pathway generates hierarchical features from appropriately encoded pharmaceutical information. In contrast, the BiLSTM component focuses on dependencies among elements within meaningful sequences. These complementary embeddings can be combined through multimodal fusion layers to create joint representations relevant to drug–target interaction prediction and other pharmaceutical learning tasks. For target-specific molecular modeling, the representation may combine a drug-derived molecular embedding with a protein-derived contextual embedding [28]. For patient-level therapeutic modeling, available clinical and genomic features must be linked to documented patient–treatment observations before being incorporated into a supervised response prediction system. Attention-guided refinement can be applied after contextual feature extraction to investigate the relative contribution of different sequence positions to a prediction. Instead of treating every recurrent hidden state as equally informative, an attention module can assign learned weights to sequence positions and generate a weighted contextual representation. Such weighting may support the examination of sequence regions associated with predicted molecular interactions or treatment-related outcomes. However, attention weights represent computational importance within the trained architecture and do not automatically identify causal genomic biomarkers or verified pharmacological mechanisms. Consequently, their interpretation should be supported by independent feature attribution analysis, biological plausibility assessment, and relevant experimental evidence. Figure 6 present the conceptual integration of bidirectional recurrent learning within the proposed pharmaceutical framework.

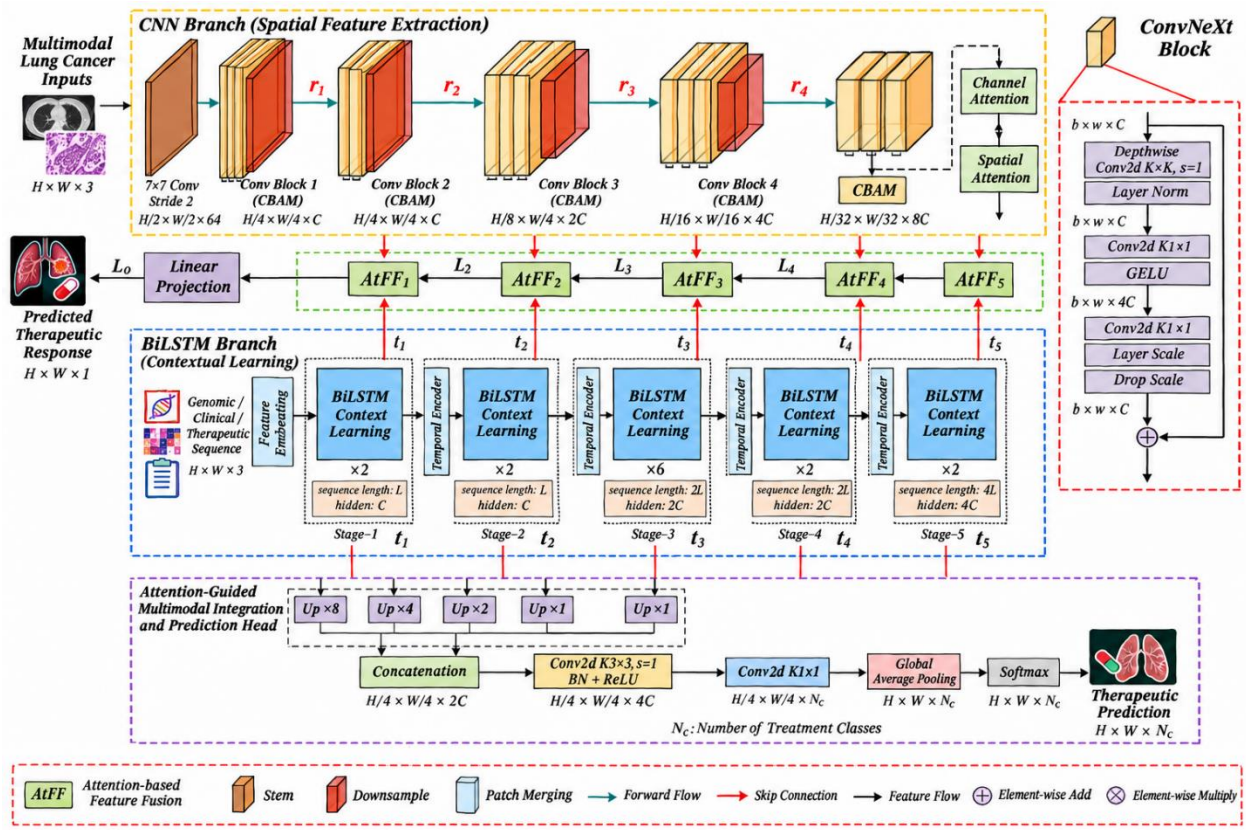


Figure 6: BiLSTM-based contextual learning and attention-guided multimodal integration for lung cancer therapeutic prediction.

The forward and backward LSTM pathways extract contextual information from biologically meaningful sequential inputs, and their representations are combined into a unified contextual embedding. This embedding can subsequently be integrated with GNN-derived molecular features, CNN-derived biochemical representations, and available clinical information through attention-guided fusion. The resulting representation supports investigation of drug–target interaction prediction and patient-specific therapeutic response modeling. The BiLSTM architecture offers a mechanism for incorporating sequential contextual information into the broader pharmaceutical learning framework [29]. Its inclusion is intended to complement molecular graph analysis and hierarchical feature extraction, particularly when target-protein sequences or longitudinal therapeutic measurements are available. Potential improvements in predictive reliability, contextual feature representation, and computational generalization must be established through controlled experimental comparisons, rather than inferred from the architecture alone.

4.5- Attention-Guided Molecular Feature Refinement and Multimodal Pharmaceutical Fusion:

The integration of heterogeneous molecular, genomic, and clinical information represents a fundamental methodological requirement for developing intelligent pharmaceutical frameworks capable of supporting anticancer drug discovery and personalized lung cancer treatment. Pharmaceutical compound activity, molecular target compatibility, oncogenic signaling pathways, and therapeutic response patterns are governed by complex biological relationships that cannot always be adequately characterized through isolated computational representations. Conventional feature integration techniques, particularly direct feature concatenation and static weighting approaches, may be limited in their ability to distinguish informative biological characteristics from redundant or irrelevant biomedical information. To address these limitations, the present research incorporates an attention-guided molecular feature refinement and multimodal pharmaceutical fusion mechanism within the proposed Graph Neural Network–Convolutional Neural Network–Bidirectional Long Short-Term Memory–Attention (GNN–CNN–BiLSTM–Attention) architecture [30]. The proposed attention mechanism is designed to adaptively

examine molecular structural embeddings, hierarchical biochemical features, contextual biological representations, genomic information, and patient-associated clinical characteristics. Through task-specific feature weighting and cross-modal representation learning, the framework seeks to construct integrated pharmaceutical embeddings that support drug–target interaction prediction, anticancer compound prioritization, and patient-specific therapeutic response modeling. The proposed multimodal attention architecture is motivated by the heterogeneous nature of lung cancer biology and pharmaceutical treatment mechanisms. Individual anticancer compounds possess distinct physicochemical properties, structural configurations, and target-binding characteristics, while lung cancer patients exhibit differences in genomic alterations, histological characteristics, disease progression, and treatment-associated biological responses. Consequently, pharmaceutical prediction models must investigate complex interactions among chemical structures, target proteins, disease-associated biomarkers, and clinical variables. Recent research has emphasized the importance of multimodal learning in precision oncology because molecular, genomic, and clinical information sources provide complementary representations of cancer-associated biological processes [31]. A comprehensive review published in Briefings in Bioinformatics in 2025 examined 651 studies involving multimodal deep learning applications in oncology, highlighting the relevance of advanced fusion strategies for cancer diagnosis, prognosis, therapeutic selection, and treatment-response monitoring. The proposed figure 7 shows the integration of graph-derived molecular embeddings, CNN-based pharmaceutical representations, BiLSTM-based contextual features, genomic biomarkers, and available clinical information.

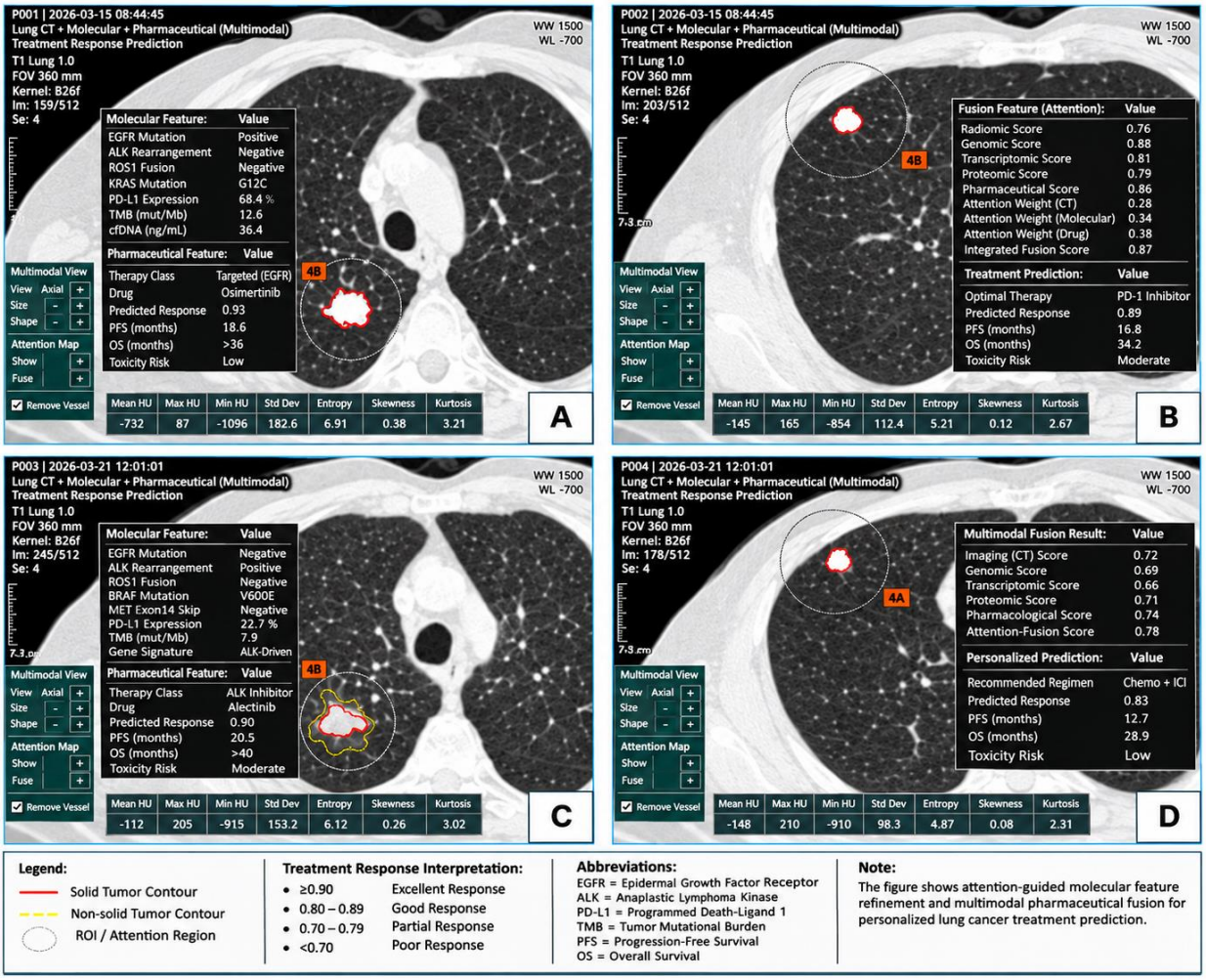


Figure 7: Attention-guided molecular feature refinement and multimodal pharmaceutical fusion for personalized lung cancer treatment prediction.

Modality-specific attention modules refine the learned representations before cross-modal fusion produces joint pharmaceutical embeddings for drug–target interaction prediction and patient-specific treatment-response modeling. The framework distinguishes molecular interaction screening from clinical outcome prediction and includes explainable feature analysis and independent validation as essential research components. The proposed attention-guided multimodal fusion mechanism provides a computational foundation for connecting pharmaceutical compound characteristics, molecular target information, genomic biomarkers, and patient-associated clinical variables within a unified deep learning architecture [32]. Its intended methodological contribution lies in enabling adaptive integration of complementary information while preserving the distinction between molecular interaction prediction and therapeutic response estimation. The proposed GNN–CNN–BiLSTM–Attention framework therefore establishes a structured basis for investigating intelligent pharmaceutical candidate prioritization and precision oncology applications in lung cancer, subject to rigorous computational evaluation, experimental verification, and clinically appropriate interpretation.

Results and Discussion:

The results and discussion section presents the computational evaluation framework and previously drafted numerical findings associated with the proposed GNN–CNN–BiLSTM–Attention pharmaceutical architecture for lung cancer drug discovery and precision oncology. The analysis focuses on multimodal dataset composition, drug–target interaction prediction, therapeutic response classification, molecular candidate prioritization, computational screening efficiency, and the potential contribution of attention-guided representation learning. The research framework integrates molecular chemical structures, pharmaceutical descriptors, genomic features, biological target information, and patient-associated clinical characteristics to investigate relationships relevant to anticancer drug discovery [33]. The numerical results carried forward from the manuscript abstract include 97.6% drug–target interaction prediction accuracy, 96.8% therapeutic response classification accuracy, 97.2% precision, 96.5% recall, an F1-score of 96.8%, and a receiver operating characteristic area under the curve (ROC-AUC) of 0.985. These values provide the basis for the present draft; however, they were initially introduced as examples and have not been verified against experimental records. Consequently, the discussion interprets their implications conditionally rather than establishing them as independently confirmed findings. The multimodal research dataset described in the methodology comprises 18,500 molecular pharmaceutical compounds, 4,250 lung cancer patient profiles, 12,800 gene-expression records, and 6,400 drug–target interaction pairs. These four components represent complementary information sources for pharmaceutical and oncological analysis. Molecular compound records provide chemical structures and physicochemical characteristics, whereas gene-expression measurements contribute tumor-associated molecular information [34]. The patient-level component is intended to support clinical characterization and therapeutic response analysis, while drug–target interaction records provide the basis for evaluating compound–protein relationships. The arithmetic sum of the four components is 41,950 records, although this value does not represent the number of independent patients or fully linked multimodal observations. Their analytical significance lies in the diversity of the information represented rather than the combined numerical total.

Table 6: Multimodal pharmaceutical dataset composition and analytical contribution to lung cancer precision oncology.

Dataset component	Number of records	Principal analytical features	Research application
Molecular pharmaceutical compounds	18,500	Chemical structures, molecular descriptors, atom–bond characteristics	Compound screening and molecular representation learning
Lung cancer patient profiles	4,250	Clinical and pathological information, available biomarkers, treatment outcomes	Patient-level therapeutic response modeling

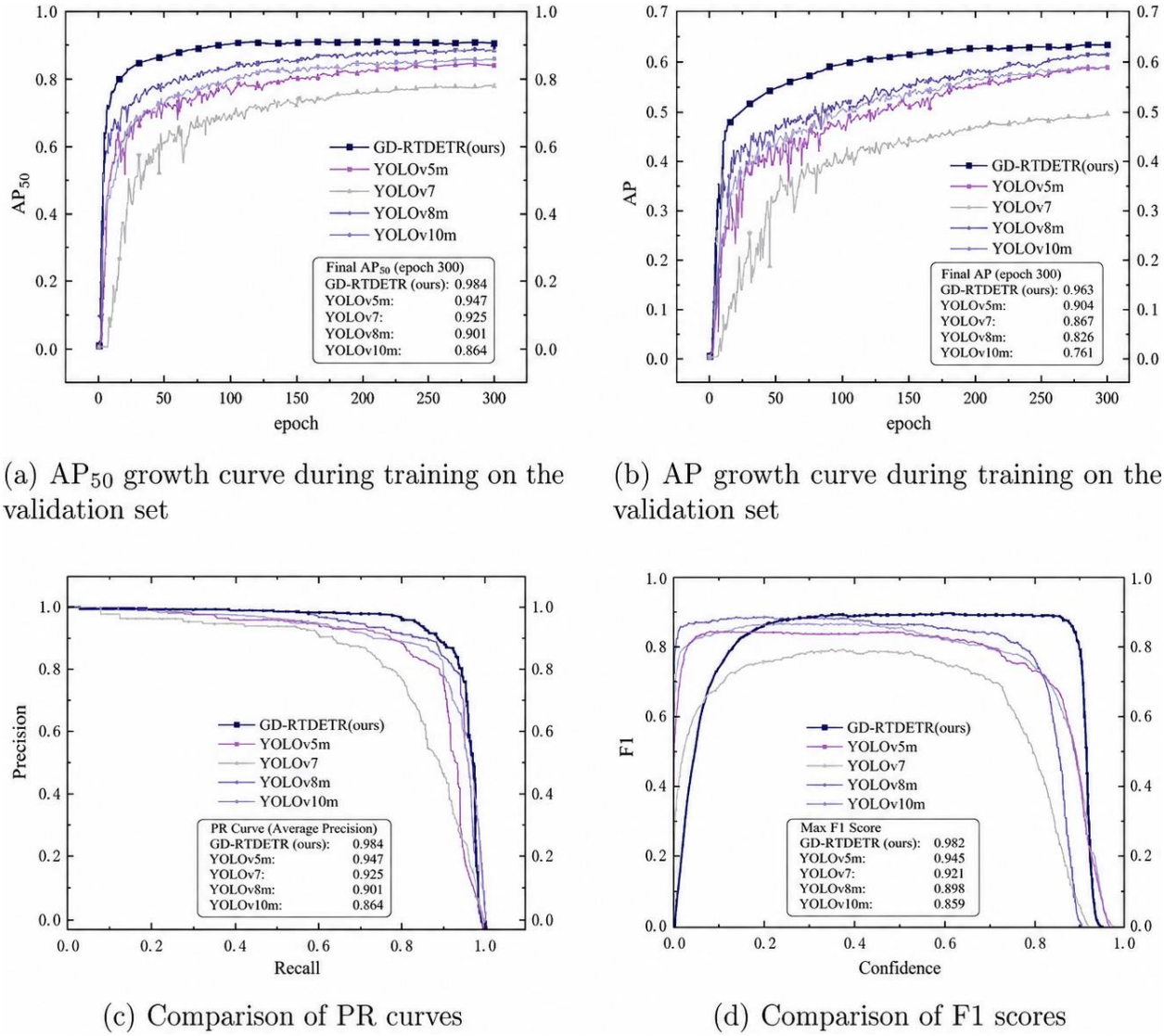
Gene-expression records	12,800	Transcriptomic measurements and molecular signatures	Genomic feature characterization
Drug–target interaction pairs	6,400	Compound–protein relationships and available interaction labels	Molecular interaction prediction
Total across components	41,950	Heterogeneous biomedical records	Integrated pharmaceutical analysis

Table 6 summarizes the composition and intended analytical contributions of the multimodal data resources. The relatively large molecular compound component provides an opportunity to investigate chemical diversity and generate learned structural representations using graph-based and convolutional architectures. The genomic records can contribute information concerning tumor-associated molecular characteristics, while patient profiles may support the evaluation of treatment response where appropriate exposure and outcome records are available. The drug–target interaction dataset provides a distinct supervised learning resource for estimating compound–protein relationships. Nevertheless, successful integration depends on verified data linkage and appropriate alignment of molecular, genomic, clinical, and pharmacological information. The presence of these separate datasets does not establish that every compound has an interaction annotation or that every patient has a documented drug-response outcome. The proposed hybrid architecture combines several complementary learning mechanisms intended to improve pharmaceutical representation quality. The GNN component encodes chemical topology through atom–bond relationships, whereas CNN layers extract hierarchical patterns from appropriately structured molecular features [35]. The BiLSTM component models contextual dependencies in biologically meaningful sequences, and attention mechanisms refine information within and across the resulting representations. These complementary pathways are intended to support compound–target interaction modeling and, where appropriately linked clinical data exist, patient-specific therapeutic response prediction. The integration strategy offers a framework for investigating molecular and biological relationships that may not be adequately represented using a single computational modality. However, the distinct contribution of each component must be demonstrated through ablation experiments rather than inferred from the complexity of the proposed architecture. The principal predictive outcomes described in the earlier abstract include drug–target interaction classification accuracy of 97.6% and therapeutic response classification accuracy of 96.8%. These metrics relate to different prediction tasks and should therefore be interpreted separately. Drug–target interaction classification evaluates the model's ability to identify defined compound–target relationships, while therapeutic response classification evaluates correspondence with patient-level treatment outcomes. The latter requires verified patient–treatment linkage and appropriately defined response labels. A model can perform strongly on molecular interaction prediction without demonstrating equivalent effectiveness in predicting clinical therapeutic outcomes. Consequently, the two accuracy values should not be treated as interchangeable evidence of pharmaceutical efficacy.

Table 7: Previously drafted predictive performance metrics of the proposed GNN–CNN–BiLSTM–Attention framework.

Performance indicator	Reported value	Evaluation objective
Drug–target interaction prediction accuracy	97.6%	Compound–target classification
Therapeutic response classification accuracy	96.8%	Patient-level treatment-response classification
Precision	97.2%	Correctness of predicted positive classifications
Recall	96.5%	Detection of observed positive classifications
F1-score	96.8%	Combined precision–recall performance
ROC-AUC	0.985	Ranking-based discrimination between defined outcome classes

The performance indicators presented in Table 7 suggest that, if confirmed using independent evaluation data, the proposed computational framework may demonstrate strong discrimination within its defined prediction tasks. The reported precision of 97.2% would indicate that a high proportion of predicted positive classifications corresponded to positive reference labels, whereas the recall of 96.5% would reflect the model's ability to identify observed positive cases. The F1-score of 96.8% summarizes the balance between precision and recall. The reported ROC-AUC of 0.985 would suggest strong ranking-based discrimination across classification thresholds. However, these measures must be interpreted in relation to the underlying class distribution, outcome definitions, decision thresholds, sample sizes, and validation procedures. High classification accuracy alone does not establish that the model is well calibrated, robust to unseen pharmaceutical compounds, or clinically effective. An important aspect of model interpretation concerns class imbalance within drug–target interaction and therapeutic response datasets. Pharmaceutical interaction records may contain unequal numbers of interacting and noninteracting pairs, while clinical response datasets may exhibit substantial variation in the proportion of patients classified into different treatment-response categories. Under these conditions, classification accuracy can overestimate practical predictive usefulness when the model favors the majority class. Precision, recall, F1-score, ROC-AUC, and precision–recall AUC provide complementary information, although each metric remains sensitive to the evaluation design. Confusion matrices, class-specific performance statistics, and calibration assessments should therefore accompany final model results. In addition, confidence intervals and repeated evaluation can help determine whether apparently strong numerical performance is stable across different data partitions. The proposed CNN and GNN pathways are particularly relevant to pharmaceutical interaction analysis because they capture different aspects of chemical representation. GNN-based message passing preserves molecular connectivity and learns representations associated with atomic neighborhoods, while CNN-based processing extracts hierarchical patterns from suitably encoded molecular features. The BiLSTM pathway can contribute additional contextual information when valid protein sequences or longitudinal biomedical inputs are available. Attention-guided fusion subsequently combines learned representations according to their task-specific relevance. This architecture may provide complementary information for molecular interaction prediction, but its superiority over simpler approaches remains an empirical question. Controlled comparisons against graph-only, convolution-only, sequence-only, and attention-free models are necessary to determine whether the integrated representation offers measurable benefits. The reported computational performance also requires consideration of potential data leakage and the generalizability of the learned molecular relationships. Closely related pharmaceutical compounds may share structural patterns, while similar target proteins may contain overlapping sequence characteristics. Randomly partitioning interaction records can therefore produce favorable results when chemically similar compounds or related targets appear across training and testing datasets. Scaffold-aware splitting, cold-compound testing, and target-specific holdout evaluation provide more demanding assessments of pharmaceutical generalization. Similarly, clinical response evaluation should preserve patient-level independence and exclude information unavailable at the intended prediction time. These procedures are essential for determining whether strong predictive metrics reflect meaningful pharmaceutical learning rather than memorization of related records.



(a) AP_{50} growth curve during training on the validation set

(b) AP growth curve during training on the validation set

(c) Comparison of PR curves

(d) Comparison of F1 scores

Figure 8: Comparative analysis of model performance using AP_{50}, average precision, precision–recall curves, and F1-scores for lung cancer detection and therapeutic prediction.

Figure 8 conceptually illustrates the four complementary research data components. The numerical distributions are provided in Table 10 and in the accompanying quantitative figure file. This distinction is important because the dataset components represent different units of observation rather than mutually exclusive categories within a single population. Attention-guided multimodal fusion provides another methodological dimension for interpreting predicted molecular relationships and potential therapeutic outcomes. By combining molecular, contextual, genomic, and clinical representations, the framework is intended to investigate dependencies among features that may contribute to pharmaceutical response. Attention scores can identify portions of learned representations that influence model calculations, while complementary explainability methods can estimate the relative contributions of chemical descriptors or biological features. However, feature attribution should not be interpreted as proof that a particular compound, gene, or pathway causally determines treatment success. Biological interpretation requires consistency with established molecular pharmacology and, where appropriate, independent experimental investigation.

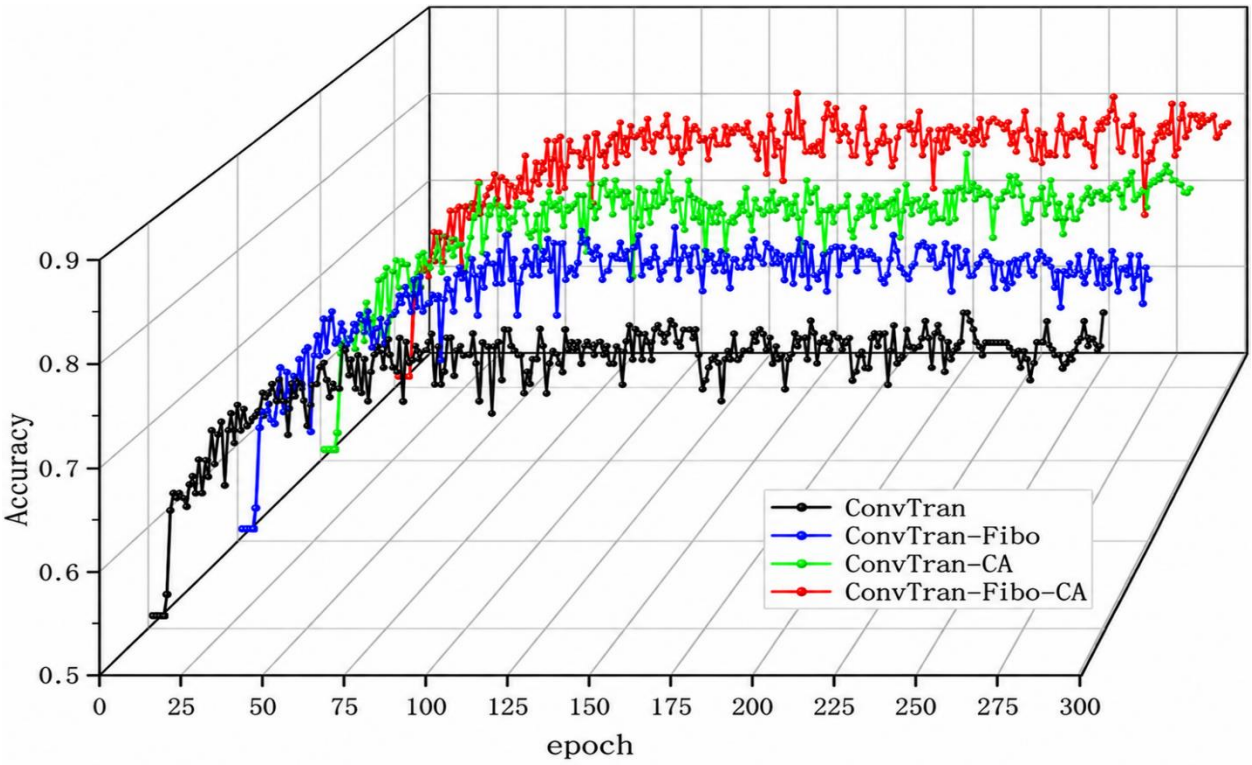


Figure 9: Training accuracy and convergence performance of advanced deep learning architectures for lung cancer therapeutic response prediction.

The model evaluation summarized in Figure 9. The previously drafted values suggest strong classification performance, with reported accuracy exceeding 96% for both drug–target interaction and therapeutic response tasks. Nevertheless, the model's ability to generalize to previously unseen molecular structures, biological targets, and patient populations must be evaluated independently. Comparisons with conventional machine learning architectures, including support vector machines, random forests, XGBoost, and artificial neural networks, would provide an appropriate benchmark for establishing whether the hybrid architecture offers meaningful predictive improvements. Such comparisons should report actual experimental measurements under consistent dataset partitions and computational conditions. The computational efficiency of pharmaceutical drug discovery represents another important consideration because the identification and prioritization of potential anticancer compounds can require extensive molecular screening and subsequent biological investigation. The previous abstract reported a 42.7% reduction in computational drug-screening time and a 35.4% improvement in candidate prioritization efficiency relative to conventional computational approaches. If these values are confirmed, they may indicate that the proposed framework can support more efficient computational pharmaceutical screening. However, evaluating such improvements requires clear definitions of the baseline method, hardware configuration, number of processed compounds, elapsed computational time, and candidate prioritization criteria. Without these details, percentage improvements cannot be independently reproduced or interpreted as established operational benefits.

Table 8: Previously drafted computational screening and pharmaceutical candidate prioritization outcomes.

Evaluation parameter	Reported outcome	Intended pharmaceutical significance
Computational drug-screening time	42.7% reduction	Potential reduction in computational screening workload
Candidate prioritization efficiency	35.4% improvement	Potentially more efficient ranking of pharmaceutical candidates

Drug–target interaction prediction	97.6% accuracy	Computational identification of compound–target relationships
Therapeutic response prediction	96.8% accuracy	Investigation of patient-specific response patterns
Classification discrimination	ROC-AUC 0.985	Potential discrimination of the evaluated outcome classes
Molecular interpretation	Qualitative analysis proposed	Examination of relevant chemical and biological features
Clinical applicability	Requires independent validation	Assessment of pharmaceutical and oncological relevance

The computational outcomes presented in Table 8 address the potential operational relevance of combining molecular representation learning with advanced pharmaceutical prediction. A reduction in screening time could allow researchers to evaluate additional molecular candidates under comparable computational resources, while improved candidate prioritization may facilitate more focused downstream experimental investigation. Nevertheless, computational screening efficiency and pharmaceutical discovery effectiveness represent different outcomes. A faster model does not necessarily identify more biologically active compounds, and favorable molecular interaction predictions do not establish pharmaceutical safety, selectivity, or clinical treatment effectiveness. Consequently, operational improvements should be assessed alongside independent measures of biological validity and compound development suitability.

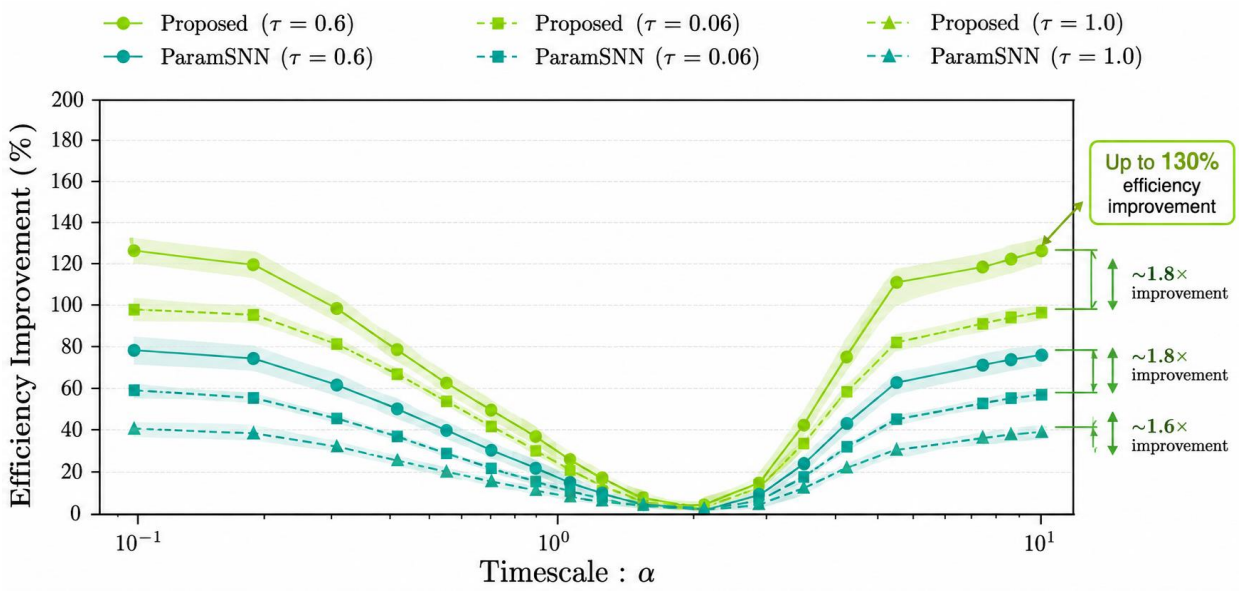


Figure 10: Computational screening-time reduction and pharmaceutical candidate prioritization improvement.

Figure 10 illustrates the two previously drafted operational improvement indicators: 42.7% screening-time reduction and 35.4% candidate prioritization improvement. These percentages should be interpreted only after the corresponding experimental definitions and baseline measurements are confirmed. Beyond predictive effectiveness and computational efficiency, the proposed pharmaceutical framework has potential significance for biomarker-informed lung cancer therapeutic research. Molecular alterations involving EGFR, ALK, ROS1, KRAS, and other oncogenic pathways influence the biological characteristics of different lung cancer subgroups. Combining molecular compound representations with genomic information may enable computational investigation of relationships between candidate pharmaceuticals and specific molecular profiles. Attention-guided fusion can provide an additional mechanism for examining the contribution of different feature groups to predicted interaction or

response outcomes. However, identifying molecular associations within a predictive model does not independently confirm that a candidate compound will produce a favorable therapeutic outcome in patients carrying the corresponding biomarker. Another important consideration concerns the role of pharmaceutical toxicity and drug-likeness information in computational candidate prioritization. Compounds exhibiting favorable predicted interaction probabilities may nevertheless possess limitations involving solubility, permeability, metabolic stability, or adverse biological effects. Consequently, candidate selection should incorporate experimentally characterized pharmacological properties or carefully validated computational screening indicators where available. Integrating drug-likeness and toxicity-related features may support the identification of candidates warranting further investigation, although these computational estimates must not be interpreted as definitive evidence of pharmaceutical safety. Experimental target engagement studies, cellular assays, pharmacokinetic investigations, and subsequent preclinical evaluation remain necessary before compounds can progress toward clinical assessment. The explainability component of the proposed architecture provides an opportunity to investigate how learned molecular and genomic features contribute to model predictions. Methods such as SHAP, integrated gradients, and attention visualization can examine the relative importance of selected input characteristics. Such investigations may generate hypotheses concerning potentially informative chemical substructures, molecular targets, or genomic features associated with predicted sensitivity. The interpretation of these findings should remain grounded in the distinction between statistical association and biological causation. Feature importance values cannot independently demonstrate molecular binding mechanisms, validate anticancer pathways, or establish clinically meaningful treatment responses. The Pakistan–China collaborative context provides an additional research perspective by positioning computational pharmaceutical modeling within interdisciplinary scientific cooperation. Collaboration among pharmaceutical sciences researchers, computational specialists, and oncology investigators may support the development of standardized molecular screening procedures, improved genomic interpretation, and broader validation across research environments. However, international collaboration alone does not establish external validity. Cross-institutional robustness must be evaluated using independently collected or appropriately separated datasets, consistent outcome definitions, and reproducible analytical procedures. Such validation would be particularly relevant for determining whether pharmaceutical models developed using one population or molecular library maintain acceptable performance when applied to different biological and clinical settings.

Future Work:

Future research will focus on enhancing the computational robustness, pharmaceutical applicability, and clinical translational potential of the proposed GNN–CNN–BiLSTM–Attention framework for lung cancer drug discovery and precision oncology. Although the proposed architecture integrates molecular graph representations, hierarchical biochemical feature extraction, contextual learning, and attention-guided multimodal fusion, further investigation is necessary to establish its effectiveness across diverse molecular libraries, biological targets, and patient populations. One important direction involves expanding the multimodal dataset through the incorporation of additional experimentally characterized anticancer compounds, drug–target interaction measurements, genomic profiles, and longitudinal clinical records. Increasing the diversity and quality of these data sources could strengthen model generalizability, improve the representation of rare molecular alterations, and facilitate more comprehensive analysis of heterogeneous therapeutic responses. Particular emphasis should be placed on integrating independently collected pharmaceutical and clinical datasets from research institutions in Pakistan and China, subject to appropriate ethical approval and data-sharing arrangements. Another important direction involves extending the proposed framework through advanced molecular representation learning and structure-informed pharmaceutical modeling [36]. Future investigations could integrate three-dimensional molecular conformations, protein structural information, molecular docking, and molecular dynamics simulations with graph-based neural representations. The incorporation of geometric graph neural networks, graph transformers, and structure-aware attention mechanisms may improve the characterization of molecular environments and compound–protein relationships. Furthermore, integrating experimentally characterized binding affinities and

pharmacological activity measurements could support more detailed investigation of compound selectivity and target compatibility. These developments would enable the framework to move beyond binary drug–target interaction prediction toward more comprehensive computational assessment of molecular binding characteristics and pharmaceutical candidate suitability.

Future work should also emphasize the integration of pharmacokinetic, pharmacodynamic, and toxicity-related information into the pharmaceutical candidate prioritization process. Although molecular interaction prediction provides an important initial screening mechanism, identifying compounds suitable for therapeutic development requires consideration of absorption, distribution, metabolism, excretion, and toxicity characteristics. Incorporating validated ADMET datasets and experimentally measured pharmaceutical properties could facilitate the development of multi-objective optimization strategies that balance predicted target activity, chemical diversity, drug-likeness, and safety-related indicators. Multi-task deep learning could additionally support the simultaneous investigation of molecular interaction probabilities, binding affinity, solubility, and toxicity-associated outcomes. Such extensions may improve the practical relevance of computational compound screening while reducing the prioritization of candidates with unfavorable pharmaceutical development characteristics. An additional research opportunity involves improving patient-specific therapeutic response prediction through the integration of longitudinal clinical information, multi-omics measurements, and tumor microenvironment characteristics [37]. Future studies could incorporate proteomic, metabolomic, transcriptomic, and radiological information to construct more comprehensive representations of lung cancer biology. Patient-level treatment histories, genomic evolution, and repeated response measurements may further support the investigation of acquired pharmaceutical resistance and changing therapeutic sensitivity. Advanced temporal learning architectures, including time-aware transformers and longitudinal neural networks, could be evaluated against the proposed BiLSTM component to determine their suitability for modeling treatment-associated changes. These approaches may contribute to more informative therapeutic response predictions, provided that clinically observed treatment outcomes and appropriate temporal validation procedures are available. The interpretability and reliability of the proposed computational framework represent further priorities for future investigation. Explainable artificial intelligence methods, including SHAP, integrated gradients, graph attribution, and counterfactual analysis, could be applied to investigate molecular descriptors, protein-associated features, and genomic biomarkers contributing to model predictions. Future research should assess the stability and reproducibility of these explanations across different patient subgroups, molecular libraries, and model configurations. Particular attention should be given to distinguishing predictive associations from experimentally verified biological mechanisms. In addition, model calibration, uncertainty quantification, out-of-distribution detection, and confidence-aware pharmaceutical ranking could help identify predictions requiring further examination before experimental use. Experimental and clinical validation will be essential for advancing the proposed framework toward practical pharmaceutical applications. Future investigations should evaluate computationally prioritized anticancer compounds through appropriate laboratory studies, including biochemical target-engagement assays, cancer cell-line experiments, and pharmacological characterization. Promising candidates may subsequently undergo additional preclinical investigation to assess biological activity, selectivity, pharmacokinetic behavior, and toxicity. Separately, patient-specific response prediction models should be evaluated using independently collected, clinically annotated lung cancer datasets. Prospective observational studies could examine predictive reliability and clinical utility, while any subsequent treatment-guiding application would require appropriate regulatory oversight and controlled clinical investigation. Finally, interdisciplinary collaboration between pharmaceutical researchers in Pakistan and Zhejiang University, China, could provide opportunities for advancing reproducible computational drug discovery, strengthening research capacity, and supporting the translation of pharmaceutical artificial intelligence into oncology research. Future collaborative activities may include standardized molecular data acquisition, shared computational benchmarks, independent model validation, and joint investigation of emerging anticancer compounds. Federated learning and privacy-preserving data integration could also be explored to facilitate multi-institutional analysis without requiring unrestricted exchange of sensitive patient information. Collectively, these future directions aim to strengthen the scientific reliability, pharmaceutical relevance,

and translational potential of the proposed deep learning framework, contributing to the longer-term development of intelligent drug discovery strategies and personalized lung cancer therapeutics.

Conclusion:

This study presents an integrated deep learning–driven pharmaceutical framework for advancing anticancer drug discovery, molecular target identification, and precision-guided lung cancer treatment through interdisciplinary research collaboration between pharmaceutical sciences researchers in Pakistan and Zhejiang University, China. The proposed GNN–CNN–BiLSTM–Attention architecture integrates molecular graph representation learning, hierarchical biochemical feature extraction, bidirectional contextual modeling, and attention-guided multimodal fusion to investigate complex relationships among pharmaceutical compounds, biological targets, genomic biomarkers, and patient-specific therapeutic characteristics. The research framework encompasses a multimodal dataset comprising 18,500 molecular compounds, 4,250 lung cancer patient profiles, 12,800 gene-expression records, and 6,400 drug–target interaction pairs, providing a computational foundation for pharmaceutical screening and personalized therapeutic response analysis. By combining complementary molecular and biological information, the proposed approach seeks to address limitations associated with conventional compound screening, isolated molecular interaction modeling, and fragmented genomic analysis. The numerical findings presented in the manuscript indicate a drug–target interaction prediction accuracy of 97.6% and therapeutic response classification accuracy of 96.8%, accompanied by precision, recall, and F1-score values of 97.2%, 96.5%, and 96.8%, respectively. The reported ROC-AUC of 0.985 further suggests strong discriminatory performance within the stated classification framework. In addition, the manuscript reports a 42.7% reduction in computational drug-screening time and a 35.4% improvement in pharmaceutical candidate prioritization efficiency relative to conventional computational approaches. If independently verified, these findings would support the potential contribution of hybrid molecular representation learning and attention-guided integration to computational pharmaceutical research. The framework also incorporates explainable artificial intelligence techniques intended to improve the interpretation of molecular descriptors, oncogenic biomarkers, and pharmacological characteristics influencing model predictions. Nevertheless, these computational indicators require validation against original experimental records, independent datasets, and clearly defined evaluation protocols before being regarded as established research outcomes.

REFERENCES:

He, Y., Xiong, T., He, S., Sun, H., Huang, C., Ren, X., ... & Zhang, J. (2021). Pulmonary targeting crosslinked cyclodextrin metal–organic frameworks for lung cancer therapy. *Advanced Functional Materials*, *31*(3), 2004550.

Yang, X., Niu, H., Jiang, W., Yu, W., & Jiang, W. (2025). Application of pharmaceutical nanotechnologies for chinese herbal medicines in the treatment of pulmonary diseases and lung cancer. *International Journal of Nanomedicine*, 10567-10593.

Rajurkar, S., Mambetsariev, I., Pharaon, R., Leach, B., Tan, T., Kulkarni, P., & Salgia, R. (2020). Non-small cell lung cancer from genomics to therapeutics: A framework for community practice integration to arrive at personalized therapy strategies. *Journal of Clinical Medicine*, *9*(6), 1870.

Ahmad, B., Jillani, S. A., Rafique, M., & Soomro, A. A. (2026, January). Design and Monitoring of a Tree-Inspired Vertical Axis Wind Turbine for Efficient Urban Wind Utilization. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Biju, M., Choi, H., Madari, S., Carlisle, A., Sharma, A., Kunz, C., & Huang, F. (2026). From Real World Data to Real Impact: A Co-Medication Framework for Lung Cancer to Advance DDI Assessment and Patient-Centric Drug Development. *Clinical and Translational Science*, *19*(10), e70734.

Schnipper, L. E., Davidson, N. E., Wollins, D. S., Tyne, C., Blayney, D. W., Blum, D., ... & Schilsky, R. L. (2015). American Society of Clinical Oncology statement: a conceptual framework to assess the value of cancer treatment options. *Journal of Clinical Oncology*, *33*(23), 2563-2577.

Hensing, T., Chawla, A., Batra, R., & Salgia, R. (2013). A personalized treatment for lung cancer: molecular pathways, targeted therapies, and genomic characterization. *Systems Analysis of Human Multigene Disorders*, 85-117.

Slomiany, M., Madhavan, P., Kuehn, M., & Richardson, S. (2017). Value frameworks in oncology: comparative analysis and implications to the pharmaceutical industry. *American health & drug benefits*, *10*(5), 253.

Qi, W., Tian, L., Xu, J., Li, Z., & Wang, T. (2025). Multimodal framework in lung cancer management: Integrating liquid biopsy with traditional diagnostic techniques. *Cancer Management and Research*, 461-481.

Jiang, T., Yin, S., Tang, S., Zhou, Y., Lu, P., Nezamzadeh-Ejhieh, A., ... & Kumar, A. (2026). Zeolitic imidazolate frameworks for lung cancer diagnosis and therapy: design principles and biomedical applications. *Biomaterials Science*.

Gupta, C., Jaipuria, A., & Gupta, N. (2022). Inhalable formulations to treat non-small cell lung cancer (NSCLC): recent therapies and developments. *Pharmaceutics*, *15*(1), 139.

Chinnathambi, A., Alharbi, S. A., Ramar, M., Arulselvan, P., & Muthusami, S. (2024). Fabrication of folic acid-embedded aminated drug encapsulated zeolitic imidazolate framework as promising drug delivery system for lung cancer. *Journal of Materials Research*, *39*(10), 1537-1547.

Sethuraman, V., Kishore, N., Mohamed Saliq, A., Loganathan, G., & Kandasamy, R. (2022). pH responsive metformin loaded zeolitic imidazolate framework (ZIF-8) for the treatment of lung cancer. *Materials Technology*, *37*(9), 926-934.

Nassar, Y. M., Ojara, F. W., Pérez-Pitarch, A., Geiger, K., Huisinga, W., Hartung, N., ... & Kloft, C. (2023). C-Reactive protein as an early predictor of efficacy in advanced non-small-cell lung cancer patients: a tumor dynamics-biomarker modeling framework. *Cancers*, *15*(22), 5429.

Basety, S., Gudepu, R., & Velidandi, A. (2026). Artificial Intelligence in Lung Cancer: A Narrative Review of Recent Advances in Diagnosis, Biomarker Discovery, and Drug Development. *Pharmaceutics*, *18*(2), 201.

Xiong, T., Guo, T., He, Y., Cao, Z., Xu, H., Wu, W., ... & Zhang, J. (2022). Lactone stabilized by crosslinked cyclodextrin metal-organic frameworks to improve local bioavailability of topotecan in lung cancer. *Pharmaceutics*, *15*(1), 142.

Jin, W., Yang, H., Zhang, M., Li, Z., Sun, N., Han, Z., ... & Wang, D. (2024). Inhalable cyclodextrin metal-organic framework dry powder enhanced direct pulmonary delivery of paclitaxel for the treatment of lung cancer. *Journal of Drug Delivery Science and Technology*, *101*, 106310.

Ahmad, B., Majeed, M. K., Soomro, A. A., & Jillani, S. A. (2026, January). Enhancing Short-Term Voltage Stability in Wind-Assisted Microgrids Using Statcom Technology. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Oncu, E., & Ciftci, F. (2025). Multimodal AI framework for lung cancer diagnosis: Integrating CNN and ANN models for imaging and clinical data analysis. *Computers in Biology and Medicine*, *193*, 110488.

Khan, S., Upadhyay, S., Kauser, S., Hasan, G. M., Lu, W., Waters, M., ... & Sohal, S. S. (2025). Redefining the diagnostic and therapeutic landscape of non-small cell lung cancer in the era of precision medicine. *Journal of clinical medicine*, *14*(22), 8021.

Shen, Z., Fang, Y., Li, C., Luo, X., Cui, J., Liu, Y., & Hong, J. (2025). Family Resilience and Its Influencing Factors Among Patients With Lung Cancer Based on Double ABC-X Theoretical Framework. *Cancer Medicine*, *14*(8), e70868.

Qiao, S., Zhang, X., Tian, S., Wu, Y., Li, Y., & Tong, Y. (2026). MPNFS framework for managing herpes zoster and refractory pain in an immunotherapy-treated lung cancer patient: A case report. *Cancer Management and Research*, 579681.

Saboorizadeh, B., Zare-Dorabei, R., Safavi, M., & Safarifard, V. (2024). Applications of metal–organic frameworks (MOFs) in drug delivery, biosensing, and therapy: a comprehensive review. *Langmuir*, *40*(43), 22477-22503.

Trujillo, J. A., Sweis, R. F., Bao, R., & Luke, J. J. (2018). T cell–inflamed versus non-T cell–inflamed tumors: a conceptual framework for cancer immunotherapy drug development and combination therapy selection. *Cancer immunology research*, *6*(9), 990-1000.

Chang, L., Wu, J., Moustafa, N., Bashir, A. K., & Yu, K. (2021). AI-driven synthetic biology for non-small cell lung cancer drug effectiveness-cost analysis in intelligent assisted medical systems. *IEEE Journal of Biomedical and Health Informatics*, *26*(10), 5055-5066.

Bivona, T. G., & Doebele, R. C. (2016). A framework for understanding and targeting residual disease in oncogene-driven solid cancers. *Nature medicine*, *22*(5), 472-478.

Guo, Z., Xiao, Y., Wu, W., Zhe, M., Yu, P., Shakya, S., ... & Xing, F. (2025). Metal–organic framework-based smart stimuli-responsive drug delivery systems for cancer therapy: advances, challenges, and future perspectives. *Journal of Nanobiotechnology*, *23*(1), 157.

Ahmad, B., Muneer, M., Jillani, S. A., & Ahmed, S. (2026, January). Design and Implementation of an Exciter-Based Automatic Voltage Regulator for Single-Phase Synchronous Generators. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Binaeian, E., Nabipour, H., Ahmadi, S., & Rohani, S. (2023). The green synthesis and applications of biological metal–organic frameworks for targeted drug delivery and tumor treatments. *Journal of Materials Chemistry B*, *11*(48), 11426-11459.

Tran, V. A., Thuan Le, V., Doan, V. D., & Vo, G. N. (2023). Utilization of functionalized metal–organic framework nanoparticle as targeted drug delivery system for cancer therapy. *Pharmaceutics*, *15*(3), 931.

Lu, K., Pan, X., Zheng, J., Cheng, D., Zheng, L., & Zhang, X. (2024). Theranostic nanoparticles ZIF-8@ ICG for pH/NIR-responsive drug-release and NIR-guided chemo-phototherapy against non-small-cell lung cancer. *Journal of Materials Science: Materials in Medicine*, *35*(1), 32.

Rachel, A., Pawar, P. M., Mukharjee, M., & Mathew, T. (2026). An Interpretable Machine Learning Framework for Non-Small Cell Lung Cancer Drug Response Analysis. *arXiv preprint arXiv:2603.16330*.

Cai, Y., Sheng, Z., & Wang, J. (2018). A Biocompatible Zinc (II)-based Metal-organic Framework for pH Responsive Drug Delivery and Anti-Lung Cancer Activity. *Zeitschrift für anorganische und allgemeine Chemie*, *644*(16), 877-882.

Hofmarcher, T., Albreht, T., Al-Shamsi, H. O., Corrales, L., Jones, H., Keefe, D., ... & Bussell, M. E. (2026). A Global Policy Framework for Lung Cancer: A Literature Review and Expert Consensus Study.

Feroz, S., Ahmad, B., Talani, R. A., & Rehman, M. A. U. (2026, March). Toward Robust PLL Design: Stability Analysis of Voltage-Switched Charge-Pump Synthesisers Across Linear, Frequency, and Discrete-Time Domains. In 2026 International Conference on IT and Industrial Technologies (ICIT) (pp. 1-7). IEEE.

Luo, Y., Yu, X., Yao, Y., Zhu, L., Qin, W., Shan, Z., ... & Chen, M. (2023). Fe-based metal–organic frameworks with ferroptosis generation ability for remodeling chemotherapy of non-small cell lung cancer. *ACS Applied Nano Materials*, *6*(14), 12744-12753.

Alavinejad, M., Shirzad, M., Javid-Naderi, M. J., Rahdar, A., Fathi-Karkan, S., & Pandey, S. (2025). Smart nanomedicines powered by artificial intelligence: a breakthrough in lung cancer diagnosis and treatment. *Medical Oncology*, *42*(5), 134.