BMC Medical Research Methodology

RESEARCH

Open Access

Artificial intelligence and machine learning in non-small cell lung cancer: the current state of the science on multi-omic applications

Savy Nistala[1], Julius Niyonzima[2], Raunak Chahal[1], Alina Hasan[3], Evelyn Ho[4], Arielle Janssens[5], Leroy W. Wheeler[6], Mark Nichols[6] and Saman Zeeshan[6*]

Abstract

Non-small cell lung cancer (NSCLC) accounts for the majority of lung cancer cases and remains a leading cause of cancer-related mortality globally. The biological heterogeneity of NSCLC challenges early detection, accurate staging, and optimal therapy. Recent advances in artificial intelligence (AI) and machine learning (ML) have enabled the integration of high-dimensional clinical, genomic, and imaging data, transforming cancer care. This narrative review critically examines recent AI/ML applications in NSCLC, emphasizing clinical validation, interpretability, and ethical deployment. We synthesize data from leading studies on imaging, genomics, and multimodal integration, highlighting how deep learning and ensemble methods have begun to outperform traditional diagnostic workflows. Persistent challenges, including dataset diversity, lack of external validation, interpretability, and ethical considerations must be addressed to realize translational impact. We advocate for large-scale, interdisciplinary collaboration to advance AI-powered personalized medicine and improve patient outcomes in NSCLC.

Keywords Artificial Intelligence, Machine Learning, NSCLC, Deep Learning, Ensemble Models, Genomics, Imaging, Prognosis, Diagnosis, Risk Factors

Introduction

Lung cancer remains the leading cause of cancer mortality globally, with NSCLC comprising the predominant histological subtype. Early detection is a critical determinant of survival, a fact underscored by the National Lung Screening Trial (NLST), which showed a 20% reduction in lung cancer mortality through low-dose computed tomography (CT) screening [1]. Nonetheless, most patients are diagnosed at advanced stages, and traditional risk models based on demographic and clinical attributes alone, such as age, gender, family history, and smoking status, offer only limited predictive precision [2].

The profound heterogeneity of NSCLC, characterized at the histopathologic and molecular levels, has prompted a shift in oncological research toward data-rich modeling. AI and ML have emerged as transformative tools in this

*Correspondence:
Saman Zeeshan
saman.zeeshan@umkc.edu
[1]School of Medicine, The University of Missouri-Kansas City, Kansas City, MO 64108, USA
[2]Department of Computer Sciences, School of Science and Engineering, The University of Missouri-Kansas City, Kansas City, MO 64110, USA
[3]Department of Applied Mathematics and Statistics, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA
[4]Department of Biological Sciences, School of Science and Engineering, The University of Missouri-Kansas City, Kansas City, MO 64110, USA
[5]Department of Chemical and Biomedical Engineering, College of Engineering, The University of Missouri-Columbia, Columbia, MO 65211, USA
[6]Department of Biomedical and Health Informatics, School of Medicine, The University of Missouri-Kansas City, Kansas City, MO 64108, USA

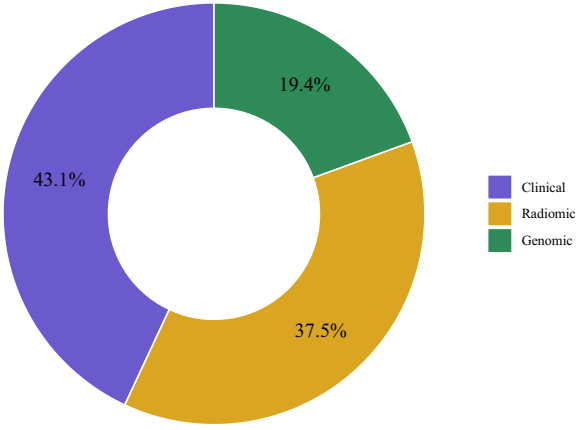

Fig. 1 Data modalities input into AI/ML models in NSCLC studies

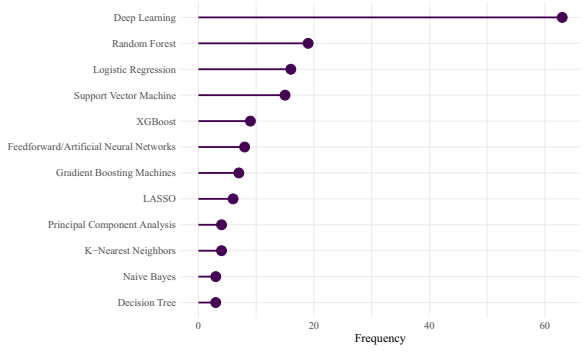

Fig. 2 Frequency of AI/ML algorithms mentioned in NSCLC studies

context, facilitating the integration and interpretation of multi-dimensional data sets, including genomics, proteomics, advanced imaging, and real-world clinical data [3, 4]. These technologies allow for detection of complex, non-linear patterns in high-dimensional data, leading to improved risk stratification, earlier diagnosis, and the promise of personalized medicine in NSCLC [5, 6]. This narrative review offers a critical synthesis of recent AI/ML advances in NSCLC and highlights future imperatives relating to rigorous validation, explainability, and ethical best practices. To orient the reader to the recent literature we reviewed, we first summarize the distribution of data modalities used across NSCLC AI/ML studies (Fig. 1) and the frequency of algorithms employed (Fig. 2).

Background

Non-small cell lung cancer (NSCLC) represents the majority of lung cancer cases and remains the leading cause of cancer-related mortality worldwide, with management transformed by advances in molecular profiling, targeted therapies, and immunotherapies that enable precision medicine approaches [7]. Current treatment strategies rely heavily on next-generation sequencing to identify actionable genomic alterations (AGAs) such as EGFR, ALK, ROS1, KRAS, and emerging targets like NRG1 fusions, guiding selection of targeted therapies that have demonstrated substantial survival benefits in both early-stage and advanced disease settings [7]. For patients without AGAs, immune checkpoint inhibitors have become the cornerstone of treatment, with recent expansion into perioperative settings showing improved pathologic response rates and event-free survival [7]. Novel therapeutic modalities including bispecific antibodies, antibody–drug conjugates, and combination approaches are addressing resistance mechanisms and expanding treatment options, with recent FDA approvals for agents targeting previously undruggable mutations [7].

Despite these advances, significant challenges remain including developing effective screening strategies for never-smokers, optimizing patient selection for adjuvant therapies, establishing correlations between surrogate endpoints and overall survival, and addressing resistance mechanisms that limit long-term treatment efficacy [7]. The evolving NSCLC landscape presents substantial opportunities for artificial intelligence and machine learning applications in molecular diagnostics, treatment selection, resistance prediction, and outcome optimization across the continuum of care from screening through advanced disease management.

The complexity of NSCLC is reflected in its clinicopathological and molecular diversity. Previous predictive models rested mainly on clinical and demographic data, but these variables alone lack the granularity and specificity required for precise prediction of tumor behavior, recurrence, or response to therapy. Integrating molecular data such as proteomic signatures, advanced imaging-based radiomics, and transcriptional profiles into risk models marks a necessary progression for accurate staging and prognostication [8].

Current advances leverage these multi-modal data streams with computational techniques capable of deep feature extraction and fusion. For example, the incorporation of electronic health records (EHRs), digital pathology, and next-generation sequencing datasets enables nuanced patient-specific modeling, as demonstrated by Levi et al., who used EHR-driven models to predict lung cancer risk [2]. The importance of including biologically relevant markers alongside clinical attributes is further underscored by Timilsina et al., who leveraged multi-omic pathway activity and chromosomal instability indices to stratify relapse risk in early-stage NSCLC [9]. The convergence of AI/ML with multimodal data offers the potential to dissect the biological underpinnings of NSCLC more effectively than one-dimensional approaches.

Methods

Literature search and study selection

A comprehensive PubMed search identified studies published within the five years preceding the review (approximately 2020–2025) using the keywords "machine learning," "artificial intelligence," "NSCLC," "diagnosis," "prognosis," "genomics," "imaging," and "risk factors." Studies describing the development or validation of AI/ML models in NSCLC were included and were subsequently organized by predominant data modality (clinical/EHR, genomic, or imaging) as well as by methodological framework and clinical endpoint. Within these categories, clinical variables such as age, sex, smoking history, comorbidities, and treatment were frequently incorporated, with integration of EHR data reported to improve prediction of survival, recurrence, and hospital length of stay; genomic profiling via RNA sequencing and mutation analysis was increasingly used for biomarker discovery and treatment-response prediction, with multi-omics models enhancing prognostic stratification; and radiomics and deep learning applied to CT, PET, and pathology images were used to advance NSCLC classification, staging, and response prediction, particularly through architectures such as convolutional neural networks, U-Net, and ResNet, which enable efficient segmentation and feature extraction. Across the corpus, models drew on three primary data modalities (clinical/EHR, imaging/radiomics, and genomics). Their relative use is depicted in Fig. 1.

Percentages reflect the relative frequency with which each data modality was used across the included primary NSCLC AI/ML studies. Each study could contribute to more than one category (e.g., a study using both radiomic and genomic data was counted once in each group), and review articles were excluded from these counts.

Data modalities & analytical frameworks

The algorithm families most commonly used in the reviewed studies are summarized in Fig. 2; below are key mechanisms.

Supervised learning algorithms

Support Vector Machine (SVM) constructs optimal hyperplanes in high-dimensional feature space to separate classes, using kernel functions to handle non-linear relationships, whereas Decision Trees (DT) create hierarchical rule-based models through recursive partitioning of data based on feature values that maximize class separation [10]. Random Forest (RF) builds an ensemble of decision trees using bootstrap samples and random feature subsets, aggregating their outputs to reduce overfitting and provide variable-importance rankings, while Logistic Regression (LR) models the probability of binary or categorical outcomes with a sigmoid function and can incorporate L1/L2 regularization to prevent overfitting [10]. Naive Bayes (NB) applies Bayes' theorem with strong feature-independence assumptions to compute posterior probabilities for fast probabilistic classification, and K-Nearest Neighbors (KNN) assigns labels based on the classes of the k most similar instances in feature space according to a chosen distance metric [10]. Adaptive Boosting (AdaBoost) sequentially trains weak learners while reweighting misclassified samples so that subsequent learners focus on difficult cases, and Extreme Gradient Boosting (XGBoost) implements gradient-boosted decision trees with regularization and efficient computation to achieve high predictive accuracy [10]. Convolutional Neural Networks (CNN) use convolution and pooling layers to learn hierarchical feature representations from grid-like input data such as images, whereas Long Short-Term Memory (LSTM) networks employ memory cells and gating mechanisms to capture long-range temporal dependencies in sequential data [10].

Unsupervised learning algorithms

Among the unsupervised learning algorithms, K-means clustering partitions data into k clusters by iteratively updating cluster centroids to minimize within-cluster variance, while Principal Component Analysis (PCA) performs linear dimensionality reduction by projecting data onto orthogonal components that capture maximum variance [10].

Diagnosis and histologic classification

Accurate diagnosis and subtype classification of non-small cell lung cancer (NSCLC) remains a historical challenge due to the disease's heterogeneity and the limitations of conventional methods. Recent years, however, have witnessed remarkable advances driven by machine learning (ML) and deep learning (DL) integrating clinical, imaging, and molecular data to elevate early detection, risk stratification, and histological classification to new levels of precision and scalability.

Early ML efforts focused on leveraging routinely available clinical and laboratory data for diagnostic differentiation. In a retrospective dual-center study including 2,312 lung cancer patients and 653 individuals with benign pulmonary nodules, Wei et al. compared XGBoost, LightGBM, Random Forest (RF), AdaBoost, and logistic regression (LR) using a multimodal feature set comprising demographic and behavioral variables, imaging-based nodule characteristics, tumor markers, and autoantibody profiles [1]. Cohort 1 was randomly split 80/20 into training and test sets, with fivefold cross-validation within the training subset to limit overfitting, and performance was then assessed in an independent external cohort (Cohort 2) [1]. For early diagnosis, LR provided the most favorable balance of sensitivity and specificity, achieving an

AUC of 0.716 (95% CI: 0.607–0.826), accuracy of 0.710, sensitivity of 70.3%, and specificity of 65.4% in the internal test set, with comparable external performance (AUC 0.712, accuracy 0.614, sensitivity 53.5%, specificity 77.1%), illustrating the potential of relatively simple models when trained on rich routine clinical data [1].

DL approaches have further advanced diagnostic performance, particularly for thoracic imaging. Shariff et al. developed a custom convolutional neural network with differential augmentation (CNN + DA) that perturbs hue, brightness, saturation, and contrast and trained it on the IQ-OTH/NCCD CT dataset comprising 1,097 images from 110 patients (40 malignant, 15 benign, 55 normal) using an 80/20 train–test split [11]. The model achieved 98.78% accuracy and significantly outperformed DenseNet, ResNet, and EfficientNetB0 on identical partitions ($p < 0.0001$); retraining and benchmarking on LC25000 histopathology, a TCIA Lung-PET-CT-Dx cohort, and NLST CT screening data confirmed strong cross-dataset performance, although generalizability still requires validation in a prospectively collected clinical cohort [11]. Complementing this, Primakov et al. introduced a fully automated three-stage pipeline that harmonizes CT data, isolates lung parenchyma, and applies an augmented 2-D U-Net for primary-tumor segmentation, trained on 999 standard-of-care CT scans from seven institutions with an internal test set of 93 patients and expert 3-D contours as ground truth [12]. Internally, lung-wise detection AUCs were approximately 0.96 with a median Dice similarity coefficient (DSC) around 0.85 and inference time under three seconds per patient; external validation on three additional datasets (238 scans from 236 patients) yielded AUC 0.98 and median DSC 0.82, and an in-silico reader study showed that automated contours were faster to obtain, more reproducible, and frequently preferred over manual segmentations, supporting readiness for workflow integration [12].

Subtype classification has likewise benefited from increasingly sophisticated DL architectures. Germain et al. used a lightweight TinyVGG-based CNN to classify NSCLC subtypes from brightfield images of cancer cell outgrowth in fibroblast co-culture across five lung cancer cell lines, including KRAS-mutant and EGFR-mutant patient-derived adenocarcinoma lines [13]. With daily imaging over nine days and data augmentation, the model achieved 100% accuracy for H520, 90% for H460, and 86% for A549 on Day 0; by Day 1, accuracy rose to 100% for A549 and 94% for H460, and the network reached 100% accuracy for HCC 4190 (EGFR oncogene) on Day 1 and HCC 4087 (KRAS oncogene) by Day 4, demonstrating that early morphological patterns can encode molecular subtype [13]. Yang et al. advanced adenocarcinoma characterization by developing a multiview convolutional spatial (MVCS) model inspired by ResNet and transformer-based learning, trained on LUNA16 and internally validated on 1,319 small cell pulmonary nodules (SCPNs) from 1,069 patients, with an external set of 160 SCPNs from 137 patients [14]. For triad classification of adenocarcinoma spectrum lesions (AAH–AIS, MIA, ADC), the MVCS fusion model achieved an accuracy of 0.69, weighted F1-score of 0.67, and Matthews correlation coefficient (MCC) of 0.43, outperforming expert radiologists and several state-of-the-art networks (e.g., Deep-RadNet accuracy 0.56, LCP-CNN 0.59, 3D DenseSharp 0.65), and highlighting the value of 3D spatial context for small nodule assessment [14].

Ensemble DL methodologies have delivered state-of-the-art performance across multiple diagnostic settings. Kumaran et al. combined VGG16, ResNet50, and InceptionV3 with Grad-CAM explainability to classify CT images into benign, malignant, and normal categories in a dataset of 1,097 images (561 malignant, 120 benign, 416 normal) [15]. The ensemble achieved 98.18% overall accuracy; for malignant cases, precision was 1.0000, recall 0.9929, and F1-score 0.9964, while for benign cases both precision and recall were 0.9333 (F1-score 0.9333), and for normal cases precision reached 0.9714, recall 0.9808, and F1-score 0.9761, indicating balanced high performance across all classes [15]. Vanitha et al. developed a MobileNet–Xception ensemble for histopathologic classification of lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue using 15,000 de-identified, pathologist-validated images (5,000 per class) resized to 224×224 pixels [16]. Their model achieved 99.44% overall accuracy, with per-class precision ranging from 0.9559 for squamous cell carcinoma to 1.0000 for benign tissue and colon adenocarcinoma, and recall values ≥ 0.9508 (lowest for lung adenocarcinoma), underscoring the potential of modular architectures to deliver robust, scalable diagnostic tools across imaging modalities [16]. Pushing the frontier further, Jian et al. integrated CNNs with gated recurrent units (CNN–GRU) to exploit both spatial and temporal CT features in the IQ-OTH/NCCD dataset [17]. By augmenting 1,097 original scans (561 malignant, 120 benign, 416 normal) to 1,683 images with class balancing (561 per class) and using an 80/20 train–test split, the model achieved 99.77% accuracy, 99.96% sensitivity, and 100% specificity on the internal test set; when applied without retraining to an independent CT-Scan dataset of 364 images (238 cancer, 126 non-cancer), accuracy remained 99.68%, although the external cohort was relatively small and non-prospective [17].

Beyond imaging, integration of omics data is revolutionizing molecular-level diagnosis. Lin et al. analyzed transcriptomic profiles of mRNA, miRNA, and lncRNA from 535 LUAD samples and 59 normal controls in TCGA, applying six ML algorithms (KNN, Naive

Bayes, RF, Decision Tree, SVM, XGBoost) with tenfold cross-validation [18]. For mRNA, AUC values ranged from 0.9844 to 1.000 and precision–recall curve (PRC) AUC from 0.9943 to 1.000; for miRNA, AUC ranged from 0.9470 to 1.000 and PRC AUC from 0.9824 to 1.000; and for lncRNA, AUC ranged from 0.986 to 1.000 with PRC AUC from 0.9943 to 1.000, with average accuracies exceeding 90% and KNN, RF, and SVM achieving perfect AUC and PRC AUC (1.000) in both training and testing sets [18]. Diagnostic biomarkers including ADRB2, FAM189A2, CLEC3B, AGER, CAT, and RS1 all yielded ROC AUCs above 0.99 in TCGA, and external validation on GEO GSE7670 showed AGER with AUC 1.000 and other markers exceeding 0.96, suggesting that transcriptomics-informed models can approach near-perfect discrimination between malignant and non-malignant tissue, even at early stages [18]. Chen and Dhahbi applied overlapping feature-selection methods to RNA-seq data from 529–535 LUAD and 498 LUSC cases, using a 70/30 train–test split with fivefold cross-validation, and found that a PCA-based classifier achieved 94.2% accuracy for LUAD vs. LUSC discrimination [19]. By intersecting genes prioritized by differential expression, PCA loadings, LASSO, mRMR, and XGBoost, they derived a 131-gene classification panel and a 17-gene biomarker subset that was externally validated on GSE28582 (50 LUAD, 28 LUSC) via ROC/AUC analysis and linked to overall survival using the Kaplan–Meier Plotter resource, demonstrating both diagnostic and prognostic utility [19].

Risk stratification and multimodal decision support at the time of diagnosis have been enabled by models that integrate molecular, imaging, and clinical features. Timilsina et al. leveraged TCGA data from 1,348 stage I–II NSCLC patients to compute pathway activity scores for 14 oncogenic cascades using PROGENy and to derive aneuploidy scores from somatic copy-number alterations [9]. Eight regressors including support vector regression (SVR), random forest regression (RFR), and k-nearest neighbors regression (KNNR) were evaluated over 10 trials, and five pathways (aneuploidy, TRAIL, WNT, VEGF, EGFR) emerged as core predictive features: SVR yielded the lowest RMSE for aneuploidy, RFR for WNT, TRAIL, and EGFR, and KNNR for VEGF, illustrating how pathway-level transcriptomics and genomic instability metrics can forecast relapse and identify high-risk patients who may benefit from intensified surveillance or adjuvant therapy [9]. Meng et al. developed a gradient boosting machine (GBM) model for small (≤ 3 cm) pulmonary nodules in two Chinese hospitals, incorporating imaging characteristics, seven tumor markers and autoantibodies, and clinical data from patients with confirmed benign or malignant lesions [20]. After LASSO feature selection identified eight key predictors (VEGF, TAABs, malignancy probability, average CT value, nodule diameter, solid proportion, gender, pleural retraction), GBM was benchmarked against generalized linear models, RF, DL, and Naive Bayes; GBM achieved the highest AUCs (0.931 = validation and 0.990 = test), with validation and test accuracies of 0.857 and 0.955, respectively, whereas RF achieved the highest sensitivities (0.914 and 0.991) but less balanced performance [20]. Aslani et al. addressed the challenge of sub-10-mm nodules using DeepCAD-NLM-L, a time-series DL model trained on NLST and evaluated on withheld NLST data and the SUMMIT study [21]. Clinical variables ($n = 273$) were normalized by min–max scaling, and nine key features (e.g., age, pack-years, smoking status, education) were selected via random forest [21]. In 100 difficult cases (25 malignant, 75 benign) evaluated at T0, T1, and T2, the retrained DeepCAD-NLM-L model achieved high sensitivity (0.92) with moderate specificity, performing comparably to two thoracic radiologists who showed lower sensitivity (0.60–0.72) but higher specificity (0.72–0.77) and highlighting the potential of time-aware DL to flag high-risk nodules earlier in screening workflows [21].

Taken together, these studies demonstrate a rapid convergence of clinical, imaging, and molecular data with advanced ML and DL methods. Modern architectures be they logistic regression ensembles, CNN variants, or integrated omics pipelines are demonstrably on the cusp of clinical maturity for NSCLC diagnosis, facilitating faster and more personalized cancer management. This evolution toward explainable, generalizable, and highly performant diagnostic models marks a pivotal advancement in the era of precision oncology.

Staging and metastatic assessment

Accurate staging in non-small cell lung cancer (NSCLC) is critical for prognostication and therapeutic planning, profoundly influencing survival outcomes. Recent innovations in machine learning (ML), deep learning (DL), and radiomics have significantly enhanced the granularity, accuracy, and clinical utility of staging methodologies heralding a new era of multimodal intelligent systems that transcend the confines of traditional imaging and clinical assessment.

In a seminal dual-center study, Wei et al. applied Extreme Gradient Boosting (XGBoost) to pathological staging and demonstrated clear gains over traditional models: the XGBoost classifier achieved an internal validation AUC of 0.913 (95% CI: 0.862–0.963), with accuracy 87.5%, sensitivity 90.9%, and specificity 81.4%, and maintained strong external performance (AUC 0.882, accuracy 0.854), underscoring the capacity of advanced ML to more reliably inform treatment selection across diverse clinical settings [1]. Radiomics-based models have further refined risk stratification in early-stage disease. Zhu

et al. retrospectively analyzed 168 resected T1N0M0 lung adenocarcinomas (93 low-risk, 75 intermediate/high-risk by IASLC grade), randomly splitting the cohort 7:3 into a 117-patient training set and 51-patient internal test set [22]. A seven-feature radiomics signature was combined with tumor diameter and consolidation-to-tumor ratio and evaluated across seven ML classifiers; on the held-out test set, the XGBoost-based combined model delivered the best performance, with an AUC of 0.975 (95% CI: 0.943–1.000), perfect sensitivity (1.00), and an F1-score of 0.89, enabling differentiation between candidates for wedge resection and those requiring more extensive surgery, although generalizability beyond this single-center cohort remains to be established [22].

Multimodal models have been particularly impactful for metastasis detection. Wang et al. used support vector machine (SVM)–based radiomics models on [99mTc]-MDP SPECT/CT to distinguish bone metastases from benign bone lesions in lung cancer patients, analyzing 141 cases (69 metastatic, 72 benign) randomly split 7:3 into training and testing sets [4]. Lesions were manually segmented in ITK-SNAP, 944 radiomics features were extracted from SPECT and CT, and LASSO was applied to build single-modality (SPECT-only or CT-only) and bimodal SVM radiomics models [4]. The optimal radiomics model combined structural (CT) and metabolic (SPECT) information, achieving AUCs of 0.919 and 0.907 in the training and testing sets, respectively, while an integrated radiomics–clinical model that added two clinical features to the SPECT + CT radiomics features further improved performance, with AUCs of 0.939 and 0.925 and representing the best-performing approach [4].

In nodal staging, Ren et al. used data from 260 patients to build clinical-biological (CBI), radiomics (Rad), and clinico-biological-radiomics (CBR) models for FDG-PET/CT assessment of mediastinal–hilar lymph node status; the CBR model achieved AUCs of 0.90 (training) and 0.89 (testing), with sensitivity 81.73% and 74.47% and specificity 87.18% and 93.55%, respectively, compared with markedly lower AUCs (0.74/0.73) and specificities (< 60%) for the CBI model and intermediate performance for the Rad model (AUC 0.76/0.83, specificity 60.26%/70.97%), highlighting the added value of integrating radiomics with clinical and biological data for more accurate and less invasive lymph-node staging [5]. DL has also begun to uncover occult nodal disease often missed by conventional criteria. Shimada et al. analyzed preoperative CT scans from 720 patients with clinical stage 0–IA NSCLC undergoing radical resection and systematic lymph node dissection, training a 3D CNN with a modified U-Net/VGG-16 backbone to extract tumor features [23]. Among the derived variables, the average CT value of the solid tumor component emerged as the most powerful predictor of pathological nodal (pN) status, achieving an AUC of 0.761 with a cut-off of −103 Hounsfield units and outperforming traditional size-based criteria, thereby offering a practical tool to flag early-stage patients at elevated risk of nodal metastasis [23].

Expanding multimodal intelligence, Zhong et al. introduced the Deep Learning Nodal Metastasis Signature (DLNMS), a ResNet18-based model that fuses PET and CT imaging at the feature level; across an internal cohort (n = 1911), external cohort (n = 355), and prospective cohort (n = 999), DLNMS achieved AUROCs between 0.875 and 0.958 for prediction of occult N1/N2 metastases and consistently outperformed PET-only, CT-only, and clinical models as well as human readers, while radiogenomic analyses linked high-risk signatures to adverse mutation and microenvironmental patterns [24]. Zhong et al. developed a CT-based deep learning model to predict pathologic N2 metastasis and survival in clinical stage I NSCLC, achieving AUCs of 0.82, 0.81, and 0.81 in an internal test set (n = 266), external cohort (n = 133), and prospective cohort (n = 300), respectively, significantly outperforming clinical models (AUCs 0.61–0.70, all $p < 0.001$) [25]. The deep learning-derived N2 risk score stratified patients by overall survival (adjusted HR: 2.9, 95% CI: 1.2–6.9; $p = 0.02$) and recurrence-free survival (adjusted HR: 3.2, 95% CI: 1.4–7.4; $p = 0.007$), with log-rank $p < 0.001$ for both outcomes [25].

Collectively, these studies exemplify the transformative potential of advanced ML and DL methodologies, especially when integrated across clinical, radiologic, and molecular domains. XGBoost and other cutting-edge algorithms not only improve accuracy and generalizability in staging but also enable individualized surgical planning and more refined risk-adapted management. Deep learning models, particularly those leveraging multi-modal imaging and radiogenomics, now offer tools capable of detecting occult metastatic disease and providing actionable prognostic information beyond human capability. The emergence of these technologies is driving a paradigm shift: from traditional one-size-fits-all protocols toward precision staging strategies tailored to the unique clinical, radiological, and molecular contexts of each patient. As these models undergo wider prospective validation and are embedded into routine workflows, they promise to close persistent diagnostic gaps, optimize treatment selection, and ultimately improve patient outcomes in NSCLC.

Prognosis and treatment response modeling

The application of predictive modeling in non-small cell lung cancer (NSCLC) has transformed the landscape of prognosis and treatment response. Machine learning models, in particular, now serve as critical tools for enhancing risk stratification and guiding therapy selection. Yang et al. retrospectively analyzed early-stage

(Stage I & II) NSCLC patients from The Cancer Genome Atlas (TCGA) cohort using clinical and genomic features to train classification and regression trees (CART), feedforward neural networks (FFNN), and least-squares support vector machines (LS-SVM) [6]. For recurrence prediction, CART achieved the best performance with an AUC of 0.82 for both lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), whereas LS-SVM and FFNN achieved more modest AUCs of 0.72–0.75 [6]. For survivability prediction, CART achieved an AUC of 0.767 for LUAD, while FFNN slightly outperformed CART in LUSC with AUCs of 0.837 and 0.815, respectively [6]. Decision tree models revealed key factors including TNM stage, gender, and specific gene mutations (NF1, ERBB2, TP53, EGFR, KRAS) that drive recurrence and survival outcomes, highlighting the interpretability advantage of tree-based models for guiding risk-adapted surveillance and treatment decisions [6].

Multiomic integration has enabled more granular molecular prognostication. Hua and Li curated a TCGA LUAD cohort with complete mRNA, lncRNA, DNA methylation, somatic mutation, and clinical data, applied ten multiomics clustering algorithms to 336 lactylation-related genes to define two cancer subtypes (CS1, CS2), and identified 28 signature lactylation-related genes differentially expressed between subtypes [26]. Through univariate Cox regression across TCGA, GSE31210, and GSE13213, they selected nine hub lactylation-related genes (HNRNPC, PPIA, BZW1, GAPDH, H2AFZ, RAN, KIF2C, RACGAP1, WBP11) and combined them with ten survival machine-learning methods, finding that a random survival forest model achieved AUCs of 0.956, 0.975, and 0.954 for 1-, 3-, and 5-year overall survival in the TCGA training set [26]. External validation in GSE13213 (n = 117) and GSE31210 (n = 226) yielded AUCs of 0.730, 0.683, 0.668 and 0.833, 0.793, 0.783 for 1-, 3-, 5-year OS, respectively, with CS1 consistently showing superior survival to CS2 [26]. High-risk patients exhibited more advanced stage (T3-4, N1-2–3, M1, III-IV), higher recurrence rates, shorter survival, and differential predicted sensitivity to immunotherapy and chemotherapy [26].

Postoperative and long-term outcome prediction has similarly benefitted from ML. Kojima et al. retrospectively studied 650 NSCLC patients who underwent R0 resection with systematic lymph-node dissection at a single center (April 2017–April 2022), training multiple algorithms including random forest, gradient boosting, XGBoost, AdaBoost, and logistic regression on tumor characteristics and lymph-node dissection variables to predict recurrence [27]. In the validation set, the random forest model achieved the highest accuracy (0.88), F1 score (0.68), and lowest Brier score (0.11), with ROC AUC of 0.89 and PR AUC of 0.73, while AdaBoost achieved slightly higher ROC AUC (0.93) and PR AUC (0.75) but lower accuracy (0.84) and F1 score (0.58) [27]. Based on overall balanced performance across all metrics, random forest was selected as the final model and achieved ROC AUC 0.92, PR AUC 0.79, accuracy 0.83, F1 score 0.64, and Brier score 0.10 on the test set [27]. In a larger single-center cohort of 1,049 completely resected stage I–IIIA NSCLC patients, Kinoshita et al. used an XGBoost-Cox framework incorporating 17 clinicopathologic variables and 52 perioperative laboratory measures (e.g., CRP, albumin, blood counts, coagulation, liver and renal function, CEA, CYFRA) to predict survival, achieving strong 5-year time-dependent AUCs of 0.890 for disease-free survival, 0.926 for overall survival, and 0.960 for cancer-specific survival and outperforming TNM stage alone, thereby underscoring the prognostic value of systemic inflammation, nutritional status, and coagulation markers despite the absence of external validation [28].

Deep learning has been deployed to estimate metastatic risk and radiotherapy outcomes. Dudas et al. trained a deep learning model on 478 NSCLC patients treated with SBRT (50 Gy in five fractions using IMRT or VMAT) after PET/CT staging, using clinical variables, dosimetric metrics, and tumor morphology to predict distant metastasis [29]. The algorithm achieved a concordance index of 0.60 ± 0.02 during cross-validation and 0.61 on the test set [29]. In the full cohort, higher mean dose within 3 cm of the PTV (Dmean,PTV3cm) was associated with increased DM risk (HR 1.94, $p < 0.001$), but this relationship was confounded by treatment technique, tumor size, and sphericity [29]. In a stratified analysis of patients with more spherical tumors (sphericity > 0.5) treated with VMAT (n = 204), higher Dmean,PTV3cm became a significant protective factor (HR 0.36, 95% CI: 0.17–0.79, $p = 0.007$), demonstrating how DL and explainable AI can uncover complex dose-geometry-outcome relationships that are obscured by confounding variables [29].

Predicting therapeutic response is central to personalized oncology, and recent work has made significant strides in both immunotherapy and chemotherapy contexts. For immunotherapy, Trebeschi et al. analyzed contrast-enhanced CT scans from 152 stage-IV NSCLC patients treated with anti-PD-1, training a 3D CNN to perform image-to-image registration between baseline and 8–12-week follow-up scans and extracting 96 latent feature maps summarizing morphological change [30]. A random forest using these features achieved AUCs of 0.69 for 1-year overall survival and 0.67 for durable clinical benefit, with peak performance (AUC 0.75) when applied to scans obtained 3–5 months after treatment initiation, suggesting a clinically actionable window for adaptive management [30]. Peng et al. pursued a genomics-based strategy, using pre-treatment somatic mutation profiles from 915 advanced NSCLC patients receiving PD-1/

PD-L1 blockade to train a 55-gene 1D CNN, with training in 429 patients from POPLAR/OAK ctDNA studies and validation in tissue-NGS cohorts from UCMC (n = 137) and MSKCC (n = 349) [31]. Across all cohorts, the CNN achieved AUCs around 0.96 for distinguishing durable clinical benefit from non-benefit, outperformed PD-L1 expression and tumor mutational burden, and stratified patients into CNN-high and CNN-low groups with markedly different progression-free and overall survival; an ensemble nomogram combining CNN, SVM, and random forest scores further delineated low-, intermediate-, and high-risk groups with consistent prognostic separation across centers and sequencing platforms [31].

Rakaee et al. added a pathology-based dimension with Deep-IO, a CNN applied to pre-treatment H&E slides from 614 advanced NSCLC patients in a US cohort, where tile-level outputs were aggregated to predict RECIST response, yielding an AUC of 0.75 (95% CI 0.64–0.85) in the internal test set and independently predicting progression-free and overall survival in multivariable models [32]. External validation in 344 European patients produced an AUC of 0.66 (95% CI 0.60–0.72) and confirmed Deep-IO as an independent predictor of PFS (HR 0.56) and OS (HR 0.53); combining Deep-IO with PD-L1 expression improved AUC to 0.70 and raised the response rate to 51% in Deep-IO-high/PD-L1-high tumors compared with 41% for PD-L1 ≥ 50% alone, highlighting the complementary value of morphology-aware DL [32].

Chemotherapy response has likewise been modeled using imaging-based DL and radiomics. In a two-center cohort of 343 patients receiving first-line chemotherapy, Chang et al. trained a deep multiple-instance learning model on pretreatment CT, using 301 patients from Shengjing Hospital (163 responders, 138 non-responders) split 70/10/20 into training, validation, and internal test sets [33]. A VGG16 backbone with attention-based MIL pooling yielded an internal test accuracy of 0.883 and AUC of 0.982 for predicting RECIST response, and maintained good performance in an external cohort (22 responders, 20 non-responders) from a second hospital with different scanners (accuracy 0.833, AUC 0.940), demonstrating cross-center generalizability despite small external numbers [33]. In a complementary radiomics study of 315 patients treated with first-line chemotherapy, the same group extracted 1,688 features from intratumoral volumes and concentric 0–3, 3–6, 6–9, and 9–12-mm peritumoral shells on baseline CT, reduced them via LASSO to 20 features, and tested logistic regression, SVM, and random forest classifiers [34]. A logistic-regression model combining features from the 0–3-mm and 3–6-mm peritumoral bands achieved the best internal performance (AUC 0.97, accuracy 92.7%) and preserved good discrimination in the 43-patient external cohort (AUC 0.85), underscoring the predictive value of fine-grained peritumoral heterogeneity [34].

Progress has also been notable in predicting response to combination and targeted therapies. Zhou et al. retrospectively evaluated 90 advanced NSCLC patients receiving first-line chemo-immunotherapy, using 70 cases (38 with durable clinical benefit, 32 without) for feature selection and training a radiomics–clinical SVM (RC-SVM) based on 851 CT radiomic features and multiple clinical and blood markers [35]. After reducing to 26 imaging and 6 clinical variables and then to six principal components, the RC-SVM achieved an AUC of 0.91 in the derivation set and 0.73 in cross-validation; when applied unchanged to a 20-patient external test set, it yielded an AUC of 0.84 (95% CI 0.80–0.89), suggesting that fusing CT radiomics with routine clinical data can enhance prediction of chemo-immunotherapy benefit, though evidence remains limited to single-center cohorts [35]. For EGFR-targeted therapy, Qureshi et al. combined demographic and clinical data with molecular-dynamics-derived descriptors of EGFR–gefitinib interaction (4 dynamic cross-correlation features, 4 MM/PBSA-like energy terms, 5 geometric pocket descriptors) in a pooled dataset of 201 EGFR-mutant patients spanning four RECIST response categories (19 CR, 118 PR, 30 SD, 34 PD) [36]. An XGBoost classifier trained with nested cross-validation achieved 97.5% accuracy for four-class response prediction, markedly outperforming energy-only or geometry-only models, though all evaluation remained internal to this dataset without independent external validation [36].

Song et al. developed a multimodal DL model combining 3D-ResNet CT features with basic clinicopathologic variables (age, sex, smoking history, histology) to predict ALK fusion status in a primary cohort of 651 surgically resected NSCLC patients; the CT-only network achieved an AUC of 0.80, which improved to 0.85 with clinical features, and external validation in 286 patients from two additional hospitals confirmed similar performance (AUC 0.85, 95% CI 0.80–0.89) [37]. In a separate cohort of 91 ALK-positive advanced NSCLC patients treated with crizotinib, those predicted as ALK-positive by the model had significantly longer progression-free survival than model-negative cases (16.8 vs 7.5 months), supporting the use of integrated CT-clinical DL as a noninvasive biomarker screen [37]. Cross-cutting work has addressed privacy and workflow efficiency. Liu et al. used a multicenter cohort of 245 NSCLC patients from four institutions treated with definitive chemoradiotherapy and planning CT to train a custom 3D CNN (conv3DNet) for treatment-response prediction, comparing centralized training with simulated two-client and real-world three-client federated learning configurations

[38]. In the simulated environment, centralized models achieved AUCs around 0.72 (DL1: AUC 0.718), which was matched or slightly exceeded by the two-client federated model (FL1: AUC 0.725) [38]. In real-world settings using three-client federated learning, the model maintained reasonable performance (FL2 with hospitals A-B-D: AUC 0.698, tested on hospital C; FL4 with hospitals A-C-D: AUC 0.672, tested on hospital B), demonstrating that federated learning can preserve patient privacy while maintaining comparable multicenter accuracy, though performance variations reflect differences in data distribution across institutions [38].

Chen et al. focused on workflow automation in surgical planning, developing a DL-based segmentation tool to identify and classify pulmonary vessels and bronchi on preoperative CT in 20 patients undergoing segmentectomy [39]. The model achieved 85% accuracy for vessel/bronchus detection versus 80% with manual annotation ($p = 1.00$), 80% vs 95% for vessel classification ($p = 0.34$), and 70% vs 80% for overall segmentation accuracy ($p = 0.72$); although small sample size limited statistical power, the tool reliably reconstructed 3D vascular structures of high surgical relevance, suggesting potential to reduce manual segmentation burden and flag anatomic variants during planning [39]. Finally, spatially resolved radiomics is redefining how intratumoral heterogeneity is leveraged for outcome prediction. Ye et al. retrospectively analyzed 178 patients with stage IB–III NSCLC from four centers who received neoadjuvant immunochemotherapy and had pre-treatment contrast-enhanced CT plus standardized IASLC pathologic assessment of pathologic complete response (pCR) [40]. In 108 patients from the derivation center, a conventional radiomics logistic-regression model built from eight LASSO-selected features achieved a training AUC of 0.778 and accuracy of 0.685, whereas an intratumoral "habitat" model that clustered voxel-wise radiomic features into four subregions and fused subregional descriptors improved training AUC to 0.861 and accuracy to 0.815 [40]. External validation in 70 patients from three additional centers confirmed that the habitat model maintained superior discrimination and accuracy (AUC 0.781, accuracy 0.743) compared with the traditional radiomics model (AUC 0.723, accuracy 0.686), supporting spatially resolved heterogeneity modeling as a more informative, noninvasive strategy for predicting pCR after neoadjuvant immunochemotherapy in resectable NSCLC [40].

Together, these advances illustrate how the fusion of AI with clinical, imaging, molecular, dosimetric, and pathology data is rapidly moving NSCLC care from generalized protocols toward truly individualized management, as models that predict recurrence, survival, and treatment benefit become more robust, interpretable, and ready for integration into real-world workflows.

Atypical driver mutations and residual gaps

Despite significant advances in the application of artificial intelligence (AI) and machine learning (ML) for non-small cell lung cancer (NSCLC) diagnostics, a critical limitation remains: the ability of current models to accurately identify and classify cases involving atypical or rare driver mutations. While conventional models have demonstrated excellent performance for common mutations and typical imaging features, the heterogeneity of NSCLC including rare alterations such as those involving NTRK, RET, MET exon 14 skipping, ROS1, BRAF, HER2, and NRG1 poses substantial challenges for robust prediction and classification [41]. These rare mutations are difficult to detect due to their low prevalence, diverse clinical manifestations, and the scarcity of large, annotated datasets encompassing such variants [41].

Existing AI/ML models often rely on training data dominated by more prevalent mutations (e.g., EGFR, ALK), potentially leading to reduced sensitivity and specificity when tasked with predicting less common molecular subtypes [41]. As a result, patients exhibiting uncommon genomic profiles may be misclassified or overlooked, with subsequent implications for targeted therapy selection and clinical outcomes. It is essential that future AI/ML approaches are designed to incorporate diverse, multi-center data, including underrepresented mutation classes, and leverage multimodal information such as radiomics, clinical features, and molecular data to enhance model generalizability.

Moreover, innovative strategies such as transfer learning, data augmentation, and semi-supervised learning can help mitigate the challenges posed by limited sample sizes of rare mutation carriers. Ensemble and deep learning architectures that accommodate multi-class outputs and can adaptively recalibrate for imbalanced datasets are particularly promising. Ultimately, the refinement of AI/ML tools to robustly predict and classify atypical driver mutations in NSCLC will not only improve diagnostic precision but also expand individualized therapeutic decision-making, ensuring equitable progress in precision oncology for all patient subgroups.

Discussion

As reflected in our corpus (Fig. 2), deep learning models have emerged as the dominant algorithmic class in NSCLC workflows, consistently outperforming traditional statistical approaches when applied to multimodal inputs (Fig. 1). The application of AI/ML to NSCLC offers clear advantages but also exposes important gaps that shape priorities for future work. AI/ML models increasingly exceed traditional statistical methods in risk prediction, classification, and outcome estimation, particularly when they integrate imaging, genomic, and clinical data to enable more personalized risk stratification

Table 1 FDA-cleared AI tools and digital biomarkers for NSCLC diagnosis, risk stratification, and treatment selection. FDA-cleared and FDA breakthrough device designation tools are listed. Regulatory status reflects authorizations as of January 2026

Company/Developer	Product Name	Clinical Application	Modality	FDA Status	NSCLC-Specific Use
Qure.ai	qXR-LN	Identifies and localizes lung nodules on chest X-rays; assists clinicians as a second-reader aid	Chest X-ray	FDA-cleared	Earlier detection of suspicious nodules potentially representing NSCLC
Qure.ai	qCT LN Quant	Quantifies solid lung nodules; tracks volumetric growth over time; provides malignancy risk scores	Non-contrast chest CT	FDA-cleared	Longitudinal monitoring and malignancy risk stratification of indeterminate nodules
Optellum	Virtual Nodule Clinic	Radiomics-based imaging AI providing Lung Cancer Prediction score for risk-stratified nodule management	Chest CT	First FDA-cleared imaging AI digital biomarker for lung cancer	Risk-stratified clinical management decisions for suspicious pulmonary nodules
Fovia Ai	F.A.S.T. aiCockpit CT Lung Nodule	Streamlines description, measurement, and false-positive rejection of lung nodules on CT	Chest CT	FDA-cleared	Improved efficiency and consistency in NSCLC nodule workup and triage
Intuitive	ion Endoluminal System software	AI-enhanced robotic bronchoscopy navigation to reach small, peripheral lung nodules for precise biopsy	Robotic bronchoscopy with AI navigation	FDA-cleared	More accurate tissue sampling of suspected NSCLC lesions
Serial CTRS collaborators	Serial CTRS (AI Prognostic Tool)	Stratifies NSCLC patients into high- versus low-risk mortality categories based on serial CT imaging	Serial CT-based analysis	FDA Breakthrough Device Designation	Prognostic risk stratification and survival prediction in diagnosed NSCLC
HeartLung.AI	AutoChamber	Automatically measures cardiac chamber volumes and left ventricular mass from non-contrast chest CT	Non-contrast chest CT (screening)	FDA-cleared	Opportunistic cardiovascular risk assessment in lung cancer screening populations
Roche	VENTANA TROP2 (EPR20043) RxDx Device	Computational pathology system combining assay and digital pathology algorithm for TROP2 expression assessment	Digital pathology (tissue biopsy)	FDA Breakthrough Device Designation	Identified advanced/metastatic non-squamous NSCLC patients likely to benefit from datopotamab deruxtecan therapy
Invenio Imaging	NIO Lung Cancer Reveal	Real-time AI evaluation of bronchoscopic lung biopsy samples during procedure to confirm tissue adequacy	Optical imaging of biopsy specimens (intra-procedural)	FDA Breakthrough Status	Ensures adequate tissue sampling for diagnosis and biomarker testing in suspected NSCLC

Abbreviations: *FDA* U.S. Food and Drug Administration, *NSCLC* Non-small cell lung cancer, *CT* Computed tomography, *RxDx* Theranostic (treatment selection)

and treatment planning [42, 43]. Automated image and data analysis can also reduce manual workload and inter-observer variability, supporting more efficient and standardized clinical workflows [43].

Although most AI/ML models for NSCLC remain at the investigational stage, a small but growing set of AI-enabled products have progressed to regulatory clearance, underscoring tangible clinical translation. These tools span the full care continuum, from nodule detection and risk stratification (e.g., Qure.ai's qXR-LN [47] and qCT LN Quant [48], Optellum's Virtual Nodule Clinic [49], and Fovia Ai's F.A.S.T. aiCockpit CT Lung Nodule [50]) to AI-assisted bronchoscopic navigation (Intuitive's Ion platform) [51], prognostic modeling with FDA Breakthrough Device Designation (Serial CT Response Score) [52], opportunistic cardiovascular risk profiling during lung cancer screening (HeartLung.AI AutoChamber) [53], and AI-driven companion diagnostics and intra-procedural pathology support for advanced disease (Roche's VENTANA TROP2 RxDx Device [54] and Invenio's NIO Lung Cancer Reveal [55]) (Table 1). Collectively, these examples illustrate that despite pervasive challenges related to generalizability, interpretability, and workflow integration discussed above, industry-led solutions are beginning to embed AIML into routine NSCLC pathways for early detection, risk-adapted surveillance, and biomarker-guided therapy.

At the same time, several limitations constrain translation into routine practice. Many published models are developed from relatively small, single-center cohorts, which restricts generalizability and heightens the risk of bias, and external validation remains absent for a substantial proportion of studies [44]. Interpretability is a further challenge because complex deep learning architectures often function as "black boxes," complicating clinician trust and regulatory assessment, while ethical and privacy considerations including data

security, algorithmic bias, and informed consent must be systematically addressed for responsible deployment [44, 45].

Looking ahead, research should focus on more sophisticated multimodal integration strategies, approaches to improve model transparency, and rigorously designed prospective, multicenter validation studies [46]. The incorporation of emerging technologies such as advanced multi-omics platforms, real-time monitoring systems, and adaptive treatment strategies holds particular promise for further refining prognostication and tailoring therapy, ultimately moving NSCLC management closer to fully individualized, data-driven care.

Conclusion

The application of AI/ML technologies to NSCLC management has demonstrated remarkable potential across the entire spectrum of clinical care, from early diagnosis and staging to prognosis prediction and treatment optimization. The evidence consistently shows that multimodal integration approaches outperform single-modality methods, with performance metrics often exceeding those of expert clinicians. Key achievements include near-perfect diagnostic accuracy (> 95%) in transcriptomic-based classification, robust staging performance (AUC: 0.913), and effective prognosis prediction (AUC: 0.92) across diverse patient populations.

The integration of imaging, genomic, transcriptomic, and clinical data through sophisticated ML algorithms has enabled the development of comprehensive diagnostic and prognostic frameworks that support personalized treatment strategies. These advances offer significant potential for improving patient outcomes through earlier detection, more accurate staging, and optimized treatment selection.

However, successful clinical implementation requires addressing challenges related to data standardization, model validation, and workflow integration. Future research should focus on developing more sophisticated multimodal integration strategies, improving model interpretability, and conducting prospective validation studies to translate these promising technologies into routine clinical practice. The continued advancement of AI/ML applications in NSCLC management holds great promise for transforming cancer care and improving patient outcomes in the era of precision medicine.

Abbreviations

AdaBoost	Adaptive Boosting
AEs	Autoencoders
AI/ML	Artificial Intelligence/Machine Learning
ALP	Alkaline Phosphatase
CAGE	Tumor-associated Gene
CB	Chronic Bronchitis
CC	Squamous Cell Carcinoma Antigen
CEA	Carcinoembryonic Antigen
CART	Classification and Regression Trees
CNN	Convolutional Neural Networks
CNV	Copy Number Variation
COPD	Chronic Obstructive Pulmonary Disease
DB-ResNet	Dual-Branch Residual Network
DLR	Deep Learning Reconstruction
HER	Electronic Health Record
EGFR	Estimated Glomerular Filtration Rate
ELISA	Enzyme-Linked Immunosorbent Assay
FDG	Fludeoxyglucose
FFNNs	Feedforward Neural Networks
FPN	Feature Pyramid Network
FPR	False Positive Rate
GAGE7	G Antigen 7
GANs	Generative Adversary Networks
GBM	Gradient Boosting Machine
GBU4-5	RNA Helicase Autoantibody
GEO	Gene Expression Omnibus
GLM	Generalized Linear Models
Grad-CAM	Gradient-weighted Class Activation Mapping
HIR	Hybrid Iterative Reconstruction
ICC	Interclass Correlation Coefficient
IGTV	Internal Gross Tumor Volume
IMRT	Intensity-Modulated Radiation Therapy
KNN	K-Nearest Neighbor
KRAS	Kristen Rat Sarcoma Viral Oncogene Homolog
LASSO	Least Absolute Shrinkage and Selection Operator Regression
LightGBM	Gradient Boosting Machine
LIDC-IDRI	Lung Image Database Consortium and Image Database Resource Initiative
LNs	Lymph Nodes
LRA	Learning Rate Adjustment
LR	Linear Regression
LS-SVM	Least Squares Support Vector Machine
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
LYM%	Lymphocyte Percentage
MAGEA1	Melanoma Antigen A1
MBIR	Model-Based Iterative Reconstruction
mRMR	Minimum Redundancy Maximum Relevance
N and M	Nodes and Metastasis
NB	Naive Bayes
NGS	Next-Generation Sequencing
NLST	National Lung Screening Trial
NRL	Neutrophil to Lymphocyte Ratio
NSCLC	Non-Small Cell Lung Cancer
NSLT	National Lung Screening Trials
pN	Pathological Nodal Status
PET	Positron Emission Tomography
PGP 9.5	Protein Gene Product
PRC	Precision-Recall Curve
PTV	Planned Target Volume
Rad	Radiomics
RBF	Radial Basis Function
RB1	Retinoblastoma Protein
RFE	Recursive Feature Elimination
RF	Random Forest
RFR	Random Forest Regression
RNN	Recurrent Neural Networks
ROC	Receiver Operating Characteristics
RNA-Seq	RNA Sequencing
SCPN	Small Cell Pulmonary Nodule
SBRT	Stereotactic Body Radiation Therapy
SND	Systematic Nodal Dissection
SMOTE	Synthetic Minority Over-Sampling Technique
SOX2	SRY-Box Transcription Factor 2
SPECT/CT	Single Photon Emission Computed Tomography
SVR	Support Vector Regression
SVM	Support Vector Machine
TCGA	The Cancer Genome Atlas
UV	Ultraviolet
VEGF	Vascular Endothelial Growth Factor

VMAT	Volumetric Arc Therapy
WBC	White Blood Cells
WNT	Wingless-Related Integration Site

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-026-02821-4.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions
The idea for the review was conceived by SZ, who led the conceptual development, and supervised the project. AH and JN conducted the initial literature search. SN and RC performed advanced research, data synthesis and interpretation. EH, AJ, and LWW contributed to critically revising the manuscript for important intellectual content. MN finalized the paper for submission. All authors contributed to the writing, reviewed the final manuscript, and approved it for submission.

Funding
The authors declare no funding was received for this work.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate.
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 15 October 2025 / Accepted: 4 March 2026
Published online: 09 March 2026

References

1. Wei W, Wang Y, Ouyang R, Wang T, Chen R, Yuan X, et al. Machine learning for early discrimination between lung cancer and benign nodules using routine clinical and laboratory data. Ann Surg Oncol. 2024;31(12):7738–49. https://doi.org/10.1245/s10434-024-15762-3.
2. Levi M, Lazebnik T, Kushnir S, Yosef N, Shlomi D. Machine learning computational model to predict lung cancer using electronic medical records. Cancer Epidemiol. 2024;92:102631. https://doi.org/10.1016/j.canep.2024.102631.
3. Gao Q, Yang L, Lu M, Jin R, Ye H, Ma T. The artificial intelligence and machine learning in lung cancer immunotherapy. J Hematol Oncol. 2023;16(1):55. https://doi.org/10.1186/s13045-023-01456-y.
4. Wang H, Chen Y, Qiu J, Xie J, Lu W, Ma J, et al. Machine learning based on SPECT/CT to differentiate bone metastasis and benign bone lesions in lung malignancy patients. Med Phys. 2024;51(4):2578–88. https://doi.org/10.1002/mp.16839.
5. Ren C, Zhang F, Zhang J, Song S, Sun Y, Cheng J. Clinico-biological-radiomics (CBR) based machine learning for improving the diagnostic accuracy of FDG-PET false-positive lymph nodes in lung cancer. Eur J Med Res. 2023;28(1):554. https://doi.org/10.1186/s40001-023-01497-6.
6. Yang Y, Xu L, Sun L, Zhang P, Farid SS. Machine learning application in personalised lung cancer recurrence and survivability prediction. Comput Struct Biotechnol J. 2022;20:1811–20. https://doi.org/10.1016/j.csbj.2022.03.035.
7. Jeon H, Wang S, Song J, Gill H, Cheng H. Update 2025: management of non-small-cell lung cancer. Lung. 2025;203(1):53. https://doi.org/10.1007/s00408-025-00801-x.
8. Alsinglawi B, Alshari O, Alorjani M, Mubin O, Alnajjar F, Novoa M, et al. An explainable machine learning framework for lung cancer hospital length of stay prediction. Sci Rep. 2022;12(1):607. https://doi.org/10.1038/s41598-021-04608-7.
9. Timilsina M, Fey D, Buosi S, Janik A, Costabello L, Carcereny E, et al. Synergy between imputed genetic pathway and clinical information for predicting recurrence in early stage non-small cell lung cancer. J Biomed Inform. 2023;144:104424. https://doi.org/10.1016/j.jbi.2023.104424.
10. Sarker IH. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci. 2021;2(3):160. https://doi.org/10.1007/s42979-021-00592-x.
11. Shariff V, Paritala C, Ankala KM. Optimizing non small cell lung cancer detection with convolutional neural networks and differential augmentation. Sci Rep. 2025;15(1):15640. https://doi.org/10.1038/s41598-025-98731-4.
12. Primakov SP, Ibrahim A, van Timmeren JE, Wu G, Keek SA, Beuque M, et al. Automated detection and segmentation of non-small cell lung cancer computed tomography images. Nat Commun. 2022;13(1):3423. https://doi.org/10.1038/s41467-022-30841-3.
13. Germain A, Sabol A, Chavali A, Fitzwilliams G, Cooper A, Khuon S, et al. Machine learning enabled classification of lung cancer cell lines co-cultured with fibroblasts with lightweight convolutional neural network for initial diagnosis. J Biomed Sci. 2024;31(1):84. https://doi.org/10.1186/s12929-024-01071-0.
14. Yang X, Chu XP, Huang S, Xiao Y, Li D, Su X, et al. A novel image deep learning-based sub-centimeter pulmonary nodule management algorithm to expedite resection of the malignant and avoid over-diagnosis of the benign. Eur Radiol. 2024;34(3):2048–61. https://doi.org/10.1007/s00330-023-10026-2.
15. Kumaran SY, Jeya JJ, R MT, Khan SB, Alzahrani S, Alojail M. Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. BMC Med Imaging. 2024;24(1):176. https://doi.org/10.1186/s12880-024-01345-x.
16. Vanitha K, R MT, Sree SS, Guluwadi S. Deep learning ensemble approach with explainable AI for lung and colon cancer classification using advanced hyperparameter tuning. BMC Med Inform Decis Mak. 2024;24(1):222. https://doi.org/10.1186/s12911-024-02628-7.
17. Jian W, Haq AU, Afzal N, Khan S, Alsolai H, Alanazi SM, et al. Developing an innovative lung cancer detection model for accurate diagnosis in AI healthcare systems. Sci Rep. 2025;15(1):22945. https://doi.org/10.1038/s41598-025-03960-2.
18. Lin L, Bao Y. Development and validation of machine learning models for diagnosis and prognosis of lung adenocarcinoma, and immune infiltration analysis. Sci Rep. 2024;14(1):22081. https://doi.org/10.1038/s41598-024-73498-2.
19. Chen JW, Dhahbi J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. Sci Rep. 2021;11(1):13323. https://doi.org/10.1038/s41598-021-92725-8.
20. Meng L, Zhu P, Xia K. Application value of the automated machine learning model based on modified CT index combined with serological indices in the early prediction of lung cancer. Front Public Health. 2024;12:1368217. https://doi.org/10.3389/fpubh.2024.1368217.
21. Aslani S, Alluri P, Gudmundsson E, Chandy E, McCabe J, Devaraj A, et al. Enhancing cancer prediction in challenging screen-detected incident lung nodules using time-series deep learning. Comput Med Imaging Graph. 2024;116:102399. https://doi.org/10.1016/j.compmedimag.2024.102399.
22. Zhu J, Tao J, Zhang F, Yao J, Chen K, Wang Y, et al. Machine learning algorithms for predicting malignancy grades of lung adenocarcinoma and guiding treatments: CT radiomics-based comparisons. J Thorac Dis. 2025;17(4):2423–40. https://doi.org/10.21037/jtd-2025-310.
23. Shimada Y, Kudo Y, Maehara S, Fukuta K, Masuno R, Park J, et al. Artificial intelligence-based radiomics for the prediction of nodal metastasis in early-stage lung cancer. Sci Rep. 2023;13(1):1028. https://doi.org/10.1038/s41598-023-28242-7.
24. Zhong Y, Cai C, Chen T, Gui H, Deng J, Yang M, et al. PET/CT based cross-modal deep learning signature to predict occult nodal metastasis in lung cancer. Nat Commun. 2023;14(1):7513. https://doi.org/10.1038/s41467-023-42811-4.
25. Zhong Y, She Y, Deng J, Chen S, Wang T, Yang M, et al. Deep learning for prediction of N2 metastasis and survival for clinical stage I non-small cell lung cancer. Radiology. 2022;302(1):200–11. https://doi.org/10.1148/radiol.2021210902.

26. Hua M, Li T. Multiomic machine learning on lactylation for molecular typing and prognosis of lung adenocarcinoma. Sci Rep. 2025;15(1):3075. https://doi.org/10.1038/s41598-025-87419-4.
27. Kojima K, Samejima H, Okishio K, Tokunaga T, Yoon H, Atagi S. Impact of the number of dissected lymph nodes on machine learning-based prediction of postoperative lung cancer recurrence: a single-hospital retrospective cohort study. BMJ Open Respir Res. 2024;11(1):e001926. https://doi.org/10.1136/bmjresp-2023-001926.
28. Kinoshita F, Takenaka T, Yamashita T, Matsumoto K, Oku Y, Ono Y, et al. Development of artificial intelligence prognostic model for surgically resected non-small cell lung cancer. Sci Rep. 2023;13(1):15683. https://doi.org/10.1038/s41598-023-42964-8.
29. Dudas D, Dilling TJ, Naqa IE. A deep learning-informed interpretation of why and when dose metrics outside the PTV can affect the risk of distant metastasis in SBRT NSCLC patients. Radiat Oncol. 2024;19(1):127. https://doi.org/10.1186/s13014-024-02519-1.
30. Trebeschi S, Bodalal Z, Boellaard TN, Tareco Bucho TM, Drago SG, Kurilova I, et al. Prognostic value of deep learning-mediated treatment monitoring in lung cancer patients receiving immunotherapy. Front Oncol. 2021;11:609054. https://doi.org/10.3389/fonc.2021.609054.
31. Peng J, Zhang J, Zou D, Xiao L, Ma H, Zhang X, et al. Deep learning to estimate durable clinical benefit and prognosis from patients with non-small cell lung cancer treated with PD-1/PD-L1 blockade. Front Immunol. 2022;13:960459. https://doi.org/10.3389/fimmu.2022.960459.
32. Rakaee M, Tafavvoghi M, Ricciuti B, Alessi JV, Cortellini A, Citarella F, et al. Deep learning model for predicting immunotherapy response in advanced non-small cell lung cancer. JAMA Oncol. 2025;11(2):109–18. https://doi.org/10.1001/jamaoncol.2024.5356.
33. Chang R, Qi S, Wu Y, Song Q, Yue Y, Zhang X, et al. Deep multiple instance learning for predicting chemotherapy response in non-small cell lung cancer using pretreatment CT images. Sci Rep. 2022;12(1):19829. https://doi.org/10.1038/s41598-022-24278-3.
34. Chang R, Qi S, Zuo Y, Yue Y, Zhang X, Guan Y, et al. Predicting chemotherapy response in non-small-cell lung cancer via computed tomography radiomic features: peritumoral, intratumoral, or combined? Front Oncol. 2022;12:915835. https://doi.org/10.3389/fonc.2022.915835.
35. Zhou Z, Guo W, Liu D, Micha JRN, Song Y, Han S. Multiparameter prediction model of immune checkpoint inhibitors combined with chemotherapy for non-small cell lung cancer based on support vector machine learning. Sci Rep. 2023;13(1):4469. https://doi.org/10.1038/s41598-023-31189-4.
36. Qureshi R, Basit SA, Shamsi JA, Fan X, Nawaz M, Yan H, et al. Machine learning based personalized drug response prediction for lung cancer patients. Sci Rep. 2022;12(1):18935. https://doi.org/10.1038/s41598-022-23649-0.
37. Song Z, Liu T, Shi L, Yu Z, Shen Q, Xu M, et al. The deep learning model combining CT image and clinicopathological information for predicting ALK fusion status and response to ALK-TKI therapy in non-small cell lung cancer patients. Eur J Nucl Med Mol Imaging. 2021;48(2):361–71. https://doi.org/10.1007/s00259-020-04986-6.
38. Liu Y, Huang J, Chen JC, Chen W, Pan Y, Qiu J. Predicting treatment response in multicenter non-small cell lung cancer patients based on federated learning. BMC Cancer. 2024;24(1):688. https://doi.org/10.1186/s12885-024-12456-7.
39. Chen X, Wang Z, Qi Q, Zhang K, Sui X, Wang X, et al. A fully automated noncontrast CT 3-D reconstruction algorithm enabled accurate anatomical demonstration for lung segmentectomy. Thorac Cancer. 2022;13(6):795–803. https://doi.org/10.1111/1759-7714.14322.
40. Ye G, Wu G, Zhang C, Wang M, Liu H, Song E, et al. CT-based quantification of intratumoral heterogeneity for predicting pathologic complete response to neoadjuvant immunochemotherapy in non-small cell lung cancer. Front Immunol. 2024;15:1414954. https://doi.org/10.3389/fimmu.2024.1414954.
41. Gou Q, Gou Q, Gan X, Xie Y. Novel therapeutic strategies for rare mutations in non-small cell lung cancer. Sci Rep. 2024;14(1):10317. https://doi.org/10.1038/s41598-024-61087-2.
42. Pokhriyal SC, Shukla A, Gupta U, Al-Ghuraibawi MMH, Yadav R, Panigrahi K. Application of artificial intelligence in neuroendocrine lung cancer diagnosis and treatment: a systematic review. Cureus. 2024;16(5):e61012. https://doi.org/10.7759/cureus.61012.
43. Yang D, Miao Y, Liu C, Zhang N, Zhang D, Guo Q, et al. Advances in artificial intelligence applications in the field of lung cancer. Front Oncol. 2024;14:1449068. https://doi.org/10.3389/fonc.2024.1449068.
44. Sinha T, Khan A, Awan M, Bokhari SFH, Ali K, Amir M, et al. Artificial intelligence and machine learning in predicting the response to immunotherapy in non-small cell lung carcinoma: a systematic review. Cureus. 2024;16(5):e61220. https://doi.org/10.7759/cureus.61220.
45. Sebastian AM, Peter D. Artificial intelligence in cancer research: trends, challenges and future directions. Life (Basel). 2022;12(12):1991. https://doi.org/10.3390/life12121991.
46. Abbaker N, Minervini F, Guttadauro A, Solli P, Cioffi U, Scarci M. The future of artificial intelligence in thoracic surgery for non-small cell lung cancer treatment: a narrative review. Front Oncol. 2024;14:1347464. https://doi.org/10.3389/fonc.2024.1347464.
47. Qure.ai Technologies. qXR-LN: 510(k) premarket notification K231805. Silver Spring (MD): U.S. Food and Drug Administration; 2023 Dec 22 [cited 2026 Jan 24]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf23/K231805.pdf
48. Qure.ai Technologies. qCT LN Quant: 510(k) premarket notification K240740. Silver Spring (MD): U.S. Food and Drug Administration; 2024 Aug 16 [cited 2026 Jan 24]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf24/K240740.pdf.
49. Optellum Ltd. Virtual Nodule Clinic: 510(k) premarket notification K202300. Silver Spring (MD): U.S. Food and Drug Administration; 2021 Mar 4 [cited 2026 Jan 24]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf20/K202300.pdf.
50. Fovia, Inc. aiCockpit AI Viewer (F.A.S.T. aiCockpit CT Lung Nodule): 510(k) premarket notification K241546. Silver Spring (MD): U.S. Food and Drug Administration; 2025 Feb 25 [cited 2026 Jan 24]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf24/K241546.pdf.
51. Intuitive Surgical, Inc. Ion Endoluminal System (Model IF1000): 510(k) premarket notification K182188. Silver Spring (MD): U.S. Food and Drug Administration; 2019 Feb 13 [cited 2026 Jan 24]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf18/K182188.pdf.
52. Onc.AI. Serial CT Response Score (Serial CTRS) AI model awarded FDA Breakthrough Device Designation for metastatic NSCLC prognosis. New York (NY): Onc.AI; 2025 Feb 5 [cited 2026 Jan 24]. Available from: https://resources.flatiron.com/press/onc.ai-awarded-fda-breakthrough-device-designation-for-serial-ct-response-score-deep-learning-model.
53. HeartLung Technologies, Inc. AutoChamber: 510(k) premarket notification K240786 (automated CT-based chamber size analysis for opportunistic cardiovascular risk profiling). Silver Spring (MD): U.S. Food and Drug Administration; 2024 Oct 10 [cited 2026 Jan 24]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf24/K240786.pdf.
54. Roche Diagnostics. VENTANA TROP2 (EPR20043) RxDx Device: AI-driven computational pathology companion diagnostic with FDA Breakthrough Device Designation for NSCLC. Basel (Switzerland): F. Hoffmann-La Roche Ltd; 2025 Apr 29 [cited 2026 Jan 24]. Available from: https://www.roche.com/media/releases/med-cor-2025-04-29.
55. Invenio Imaging, Inc. NIO Lung Cancer Reveal image analysis module: AI-based tool with FDA Breakthrough Device Designation for bronchoscopic lung forceps biopsies. Santa Clara (CA): Invenio Imaging; 2024 Oct 30 [cited 2026 Jan 24]. Available from: https://www.prnewswire.com/news-releases/invenio-imaging-receives-fda-breakthrough-device-designation-for-ai-based-image-analysis-module-for-lung-cancer-biopsies-302086501.html.

Publisher's Note