International Transactions on Electrical Engineering and Computer Science
R. Kumari et.al., Vol: 4, No: 2, pp: 64-90, June 2025

Early Detection of Lung Cancer using Machine Learning Algorithms

Rimjhim Kumari[1], Shalu Kumari[1], Sharik Ahmad[1]

Abstract: Recent advances in imaging and sequencing technology have enabled a systematic advancement in the medical treatment of carcinoma of the lungs. Meanwhile, the human mind's ability to comprehend and make optimal use of the collection for these enormous amounts of knowledge is limited. Through the combination and investigation of this extensive and complex information, lung cancer has been extensively explained using a variety of perspectives from the gathered data, is made possible in great part by machine learning-based methodologies. We give a summary of machine learning based methods in this review that support the various facets of lung cancer treatment as well as diagnosis, involving prognostication, vaccinations, rapid identification, and auxiliary diagnostics. Furthermore, we highlight the challenges and opportunities regarding potential artificial intelligence uses for the illness.

Keywords: Omics datasets, Sequencing technologies, Intricate datasets, Immunotherapy, Machine learning

1. Introduction

Cancer of the lungs is one of the numerous malignancies that are commonly diagnosed and the primary cause of cancer-related mortality globally. Within five years of being diagnosed with lung cancer, 75% of the 2.20 per cent hundred million new people who are diagnosed yearly die [1 – 2]. Cancer treatment is more difficult because of the A great deal of cancer cell complexity and heterogeneity within the tumor (the Islamic faith), which result in resistance to therapies [3].

History
Received: 10-12-2024; Revised: 15-03-2025;
Accepted: 03-04-2025

Sharik Ahmad
shariqgee@gmail.com

[1]Department of Computer Science and Applications, Sharda School of Computing Science and Engineering, Sharda University, Greater Noida – 201306, India

Numerous big collaborative cancer initiatives have been facilitated by the ongoing technological advancements in cancer research over the past few decades, which have produced a huge number of databases for medical health care imaging and genomics [4 - 6]. With the aid of these databases, scientists can investigate thorough patterns of cancer progression from examination, therapy, and reactions to clinical effects [7]. However, using a range of high-dimensional data types for clinical activities requires a great deal of work and experience, even with the aid of category decrease methods like structure and magnitude elementalizations [8–11]. Another major problem for researchers is evaluating the continuously expanding databases linked to cancer. Consequently, using frameworks for data mining (also known as in order to comprehend the basic properties like many kinds of information that benefit physicians' decision-making has become more essential. A subfield of computational intelligence, the field of machine learning (ML) uses mathematical algorithms to identify patterns in data so that predictions can be made [12 - 19]. It has been widely utilized in state of the art techniques for prognosis prediction, TME deconstructing, trademark mining, prompt identification of cancer type classification, and medication sensitivity assessment [20 - 27].

https://doi.org/10.62760/iteecs.4.2.2025.133

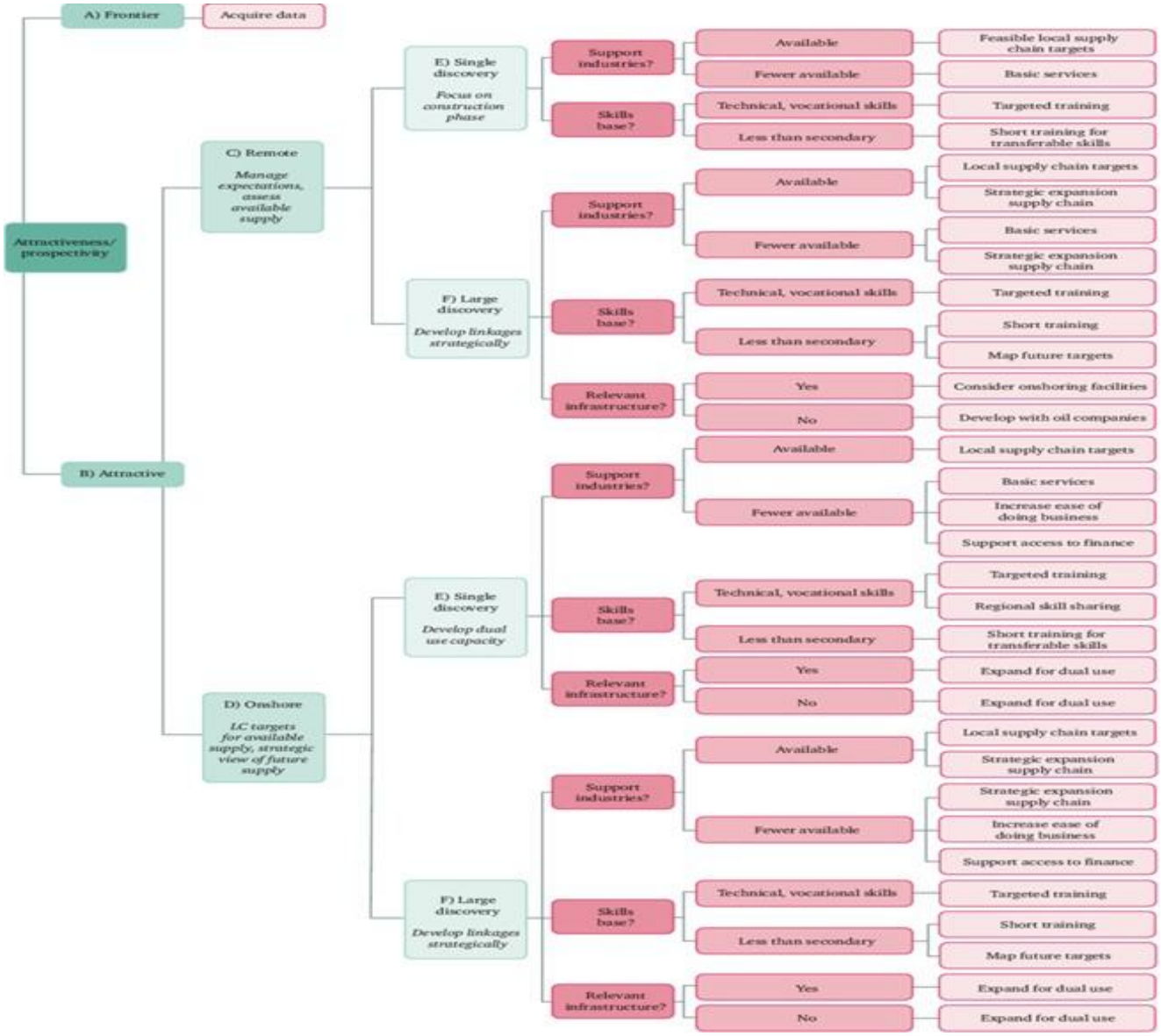

Fig. 1: Lung cancer applications utilizing machine learning models

For many years, it has also been employed in the supporting aid in the diagnosis and treatment of cancer [13 - 19]. The main machine learning techniques that utilized to incorporate challenging biological evidence (such imaging or sequencing data) for various lung cancer characteristics are compiled below (as seen in Fig. 1; S1 tables and the case of S2). We also highlight the main obstacles and opportunities for further machine learning applications in lung cancer clinical research practice. We anticipate that the assessment will improve knowledge of machine learning's prospects and skills in this area of research.

The remaining paper organized as, section 2 presents the machine learning methods review, section 3 bout the application of machine learning in cancer prediction, section 4 about the immunotherapy applications, section 5 about the challenges and conclusions presented in section 6.

2. Use machine learning to identify and diagnose lung cancer early

2.1 Using machine learning over early diagnosis and detection using diagnostic datasets

A key aspect in prompt diagnosis reduces deaths from carcinoma of the lung. The most popular method for keeping an eye on those who are more likely to get lung cancer is chest screening using low dose CT

scans. The computer-aided diagnostics (coronary artery system was created. to assist medical professionals in interpreting medical imaging data to increase diagnostic efficiency [28 - 29], and it is a helpful second opinion for doctors [30]. Nodule the process of segmentation and feature extraction. The traditional feature-based CAD problem has three stages: decision-making, professional reasoning (grouping), and (Fig. 2). Several methods, such as logarithmic modeling (also known as [31–33], straight classifier research, or the LDA technique, [34], use the measured textural features of specific nodules in CT images along with patient's clinical factors as input data to train an ML classifier for assessing the risk of malignancy. A lump size, helpful, place, gather, frontier and pulmonary findings from scans using CT are typically included in these metrics. Age, gender, smoking exposure, parents' history of cancer of the lung, and specimen collection time are among the diagnostic factors. Nevertheless, these characteristics are primarily arbitrary and subjective, and they typically fall short of providing a thorough and quantitative description of the appearance of malignant nodules. More research has been the emergence of the use of deep learning (DL) methods, especially neural networks), has led to efforts to integrate Db-based theories into the computerized imaging (CAD) system in order to improve its reliability and reduce its percentage of positives as well as length of operation for the identification of lung tumors (see Table 1) [35 - 36]. Similar to information-based CAD systems, the workflow for these representations typically consists of three steps: nodule recognition and segmentation, clinical judgment deductive reasoning, and nodule feature extraction [37]. The DL-based CAD system can automatically collect and extract intrinsic attributes of a harmful tumor [38 - 39] and model, in comparison to typical feature-based CAD applications. The nodule's three-dimensional structure (Fig. 2). For instance, Krizhevsky et al. [40] extracted the axial, coronal, and sagittal 2D-view feature vectors of the nodule from CT images in order to create OverFeatis the basis for this model [41 - 42]. The newly integrated CNN models provide a comprehensive and worldwide analysis of tumours enabling characteristic personality development utilizing images obtained by CT. Buty et al. [37] developed a supplemental CNN model for nodule segmentation by applying a circular periodic structure [43]. A deeply convolutional neural network (DCNN)-based network was used to extract the segmented nodule's texture and intensity attributes (also known as "appearance" features) and shape descriptors (also known as "shape" features) [41]. For the downstream categorization, "shape" and "appearance" characteristics were combined. Venkadesh et al. [44] also employed the collaborative strategy. Using both distinct models, 2D-ResNet50-based [45], for instance and 3D-Inception-V1 [46], to identify both characteristics of a pulmonary nodule.

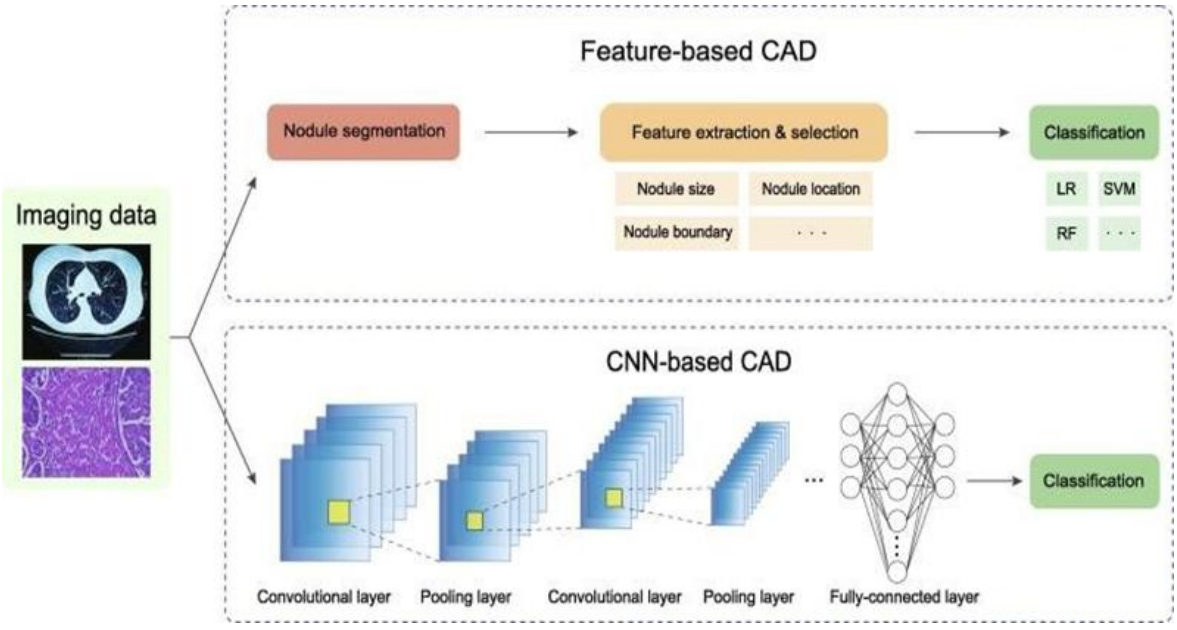

Fig. 2: Systems that use DL and feature-based CAD

Table. 1: Research on machine learning for early detection and therapy utilizing data from images

Publication	Feature Extraction	Classification Model	Sample Size	Imaging Data Type	Performance	Validation Method	Feature Selection/ Input	Highlight/ Advantage	Shortcoming
McWilliams et al. [31]	NA	LR	2961	CT images	AUC (0.907–0.960)	Hold-out	Clinical risk factors + nodule characteristics on CT images	High AUC in small nodules using extracted features	Affected by nodule characteristics selection
Riel et al. [32]	NA	LR	300	CT images	AUC (0.706–0.932)	Hold-out	Clinical factors + nodule characteristics	Comparable performance to human observers	Performance reliant on nodule size
Kriegsman et al. [34]	NA	LDA	326	MALDI	Accuracy (0.991)	Hold-out	Mass spectra from ROIs of MALDI images	Maintains high accuracy on FFPE biopsies	Relies on MALDI stratification
Mohammed et al. [37]	Spherical harmonics; DCNN	RF	1018	CT images	Accuracy (0.793–0.824)	10-fold cross-validation	CT patches + radiologist binary segmentations	Shape and appearance integration improves accuracy	No benchmarking comparisons used
Zhang et al. [38]	3D CNN-based multi-task model	3D CNN-based multi-task model	1018	CT images	Accuracy (0.9126)	10-fold cross-validation	3D CT volume feature	Higher accuracy than other methods	Ground truth defined by radiologists may be arbitrary
Ciompi et al. [39]	3D CNN-based multi-task model	3D CNN-based multi-task model	6960	CT images	DSC (0.91); accuracy (0.97)	10-fold cross-validation	3D CT volume feature	Combines clustering with attention features	Segmentation might fail outside lung regions
Krizhevsky et al. [40]	OverFeat	SVM; RF	1729	CT images	AUC (0.868)	10-fold cross-validation	3D CT volume, nodule position, and diameter	Classifies malignancy based on location	CT may not always localize the nodule
Venkadesh et al. [44]	2D-ResNet50; 3D-Inception-V4	Ensemble of 2 CNN models	16,429	CT images	AUC (0.86–0.96)	10-fold cross-validation	3D CT volume + nodule coordinates	Higher AUC than other benchmark models	Difficult to define precise nodule position
Ardila et al. [47]	Mask-RCNN; RetinaNet; Inception-V1	Mask-RCNN; RetinaNet; Inception-V1	14,851	CT images	AUC (0.944)	Hold-out	Patient history + current/prior CT volumes	High AUC; benefits from multiple scans	Demands large datasets
AbdulJabbar et al. [52]	Micro-Net; SC-CNN	Ensemble CNN	100	Histological images	Accuracy (0.913)	Hold-out	H&E-stained tumor slides	Can annotate rare datasets	Annotation accuracy varies
Coudray et al. [55]	Multi-task CNN	Multi-task CNN	1634	Histological images	AUC (0.733–0.856)	Hold-out	512×512 tiles from whole-slide images	Can predict mutation	Mutation prediction accuracy

								types like EGFR, STK11, etc.	not very high
Lin et al. [59]	DCGAN + AlexNet	DCGAN + AlexNet	22,489	CT images	Accuracy (0.9986)	Hold-out	Synthetic CT images	GAN generates synthetic lung cancer images	No benchmarking used
Ren et al. [60]	DCGAN + VGG-DF	DCGAN + VGG-DF	15,000	Histopathological images	Accuracy (0.9962); F1-score: 0.9984	Hold-out	Synthetic histopathological images	GAN-enhanced model improves accuracy	64×64 image patch too small for medical tasks

Note: NA signifies not applicable

Utilizing the unprocessed CT scans, the combination of CNN model has the superior capacity to differentiate between nodules of different sizes and malignant nodules. The support vector machines (SVM), random forest (RF), neural networks (NNs), and LR are examples of machine learning (ML) approaches that can be frequently used to conduct medical decision inference with the aid of features extracted from the most recent generation of CNN models. Interestingly, CNN models were also used in several studies to draw final clinical judgments. Using only the input CT data, Ardila et al. [47] presented a comprehensive method for methodically modeling the Objectives related to carcinoma of the lung risk category and identification. Three CNN models provided the foundation for their strategy: three models: a Mask-RCNN [48] model for lung tissue segmentation, an updated system [49] model for cancer area of interest (ROI) identification, and full-volume model that utilizes 3Dinflated Inception-V1 for predicting disease [50 - 51]. In addition to CT images, CNN-generated models are often employed in visualizing immunohistochemistry to help with lung cancer detection. Additional biological information regarding tumors along level of cells could be obtained using immune histochemical photography. than CT imaging. In order to do this, Abdul Jabbar and colleagues [52] segmented separate cells from IHC (immunohistochemistry) images and the associated hematoxylin and eosin (H&E) stained images using a Micro-Net [53] theory for anatomical border detection, followed by a SC-CNN [54] strategy. The segmented cells were then used for cell type categorization in order to evaluate arrangement of various cell kinds in the visual representations. The immune evasion and distinct development of lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD) can be identified with great resolution using this model. For example, based on H&E-stained histological whole-slide images, the tissue was classified as LUAD, LUSC, or normal in another study [55] using the latest version of in method [51]. The model's ability to detect structural changes many lung cancer driver genes is one of the work's highlights. Namely, STK and arrive11, the EGFR gene the format known as F1, SETBP1, the KRAS gene, and the protein the TP53 gene in a given tissue. It should be mentioned that when training fresh examples, some studies employed transfer learning to boost the model's efficiency and endurance, because of the dataset's high complexity and plenty of resources [38, 55]. Even while CAD already makes extensive use of these ML methods, the problem is only a small percentage of pictures. CNN-drafting systems which are feature-based alongside those which have roots are developed differently. The fundamental traits that define a suspicious nodule can be automatically extracted and extracted using DL-based CAD systems, in contrast to feature-based CAD systems. SVM means Labeling is present in support vector machines, logistic reconstruction (LR), convolutional neuronal networks (CNN), and random forests (RF). When a complex CNN model is trained using few preliminary data points, overfitting may happen. By creating bogus images, Recently, models based on generative adversarial networks, also called GANs, have been used in order to boost the accuracy of discriminative classifiers [56]. Chuquicusma et al. [57] generated simulated lung nodule CT scans using an extensive segmentation GAN (DCGAN methods) [58] model. problem of overestimation in the classification of carcinoma of the lung is currently solved by more recent research by combining GAN

algorithms with other CNN models. Lung disease categorization, Lin et al. [59] used a technique with two stages that included original and manufactured datasets. They created synthetic lung cancer images using a DCGAN models [41]. Ren as well as colleagues [60], also did a similar study. To enhance the data, they also employed DCGAN (Digital [58]. Later, for better performance by avoiding overfitting issues with pre-trained model auto-selection, they produced VGG-DF, a regularization-enhanced model learning model to combat data bias. Rapid diagnosis and identification are made possible by machine learning on omics sequencing information. An essential false discovery rate has made it difficult to implement the recommended periodic medical imaging testing for high-risk patients [61 - 62]. New methods for the quick Therefore, there is an urgent need for lung cancer detection. Numerous techniques for the early identification of lung cancer are made achievable by recent sequencing technologies [63]. Accurately identifying the subtypes of lung cancer is essential in the interim for directing the best course of treatment. With the exception of targeted therapy, the two most commonly seen subtypes of lung cancer, LUAD (~45%) and LUSC (~25%), are frequently treated similarly [64]. However, research has shown that the biochemical signatures of LUAD and LUSC differ significantly, and it has been proposed that they be treated and classified as distinct malignancies [65 - 66]. From a computational standpoint, the classification problem includes both subtype identification and early detection. In large pan-cancer sequencing datasets, prior machine learning research has demonstrated the effectiveness and progress on form of cancer categorizing and early recognition [67–75], which could support the diagnosis of lung cancer. Numerous genetic variations are known to be present in cancer cells, and the aggregation of these differences could act as indicators that document the mutational variations in different kinds of disease [3, 76, 77].

It offered an overview of predictive CAD from imaging datasets, immunotherapy research, prognosis and treatment response prediction, data integration and biomarker extraction from multi omics datasets, and sequencing based lung cancer early detection are some of the machine learning techniques applied in many lung cancer treatment fields. Machine learning, or ML, half-maximal inhibitory concentration, or IC50]; Deep learning uses DL. Computer-aided diagnostic (CAD), copy number variation (CNV), antigen of human leukocytes (HLA), computed tomography (CT), and matrix-assisted laser desorption/ionization (MALDI (Multi) are all acronyms. RECIST, Solid Tumor Response Evaluation Criteria; Tumor-infiltrating lymphocytes, or TILs. The science of machine learning is also referred to as ML, computed tomography as CT, and the LDA technique, or linear discriminant analysis. AUC stands for area under the curve, ROI for area of interest, and LR for logistic regression. The acronyms for matrix-assisted laser desorption/ionization are MALDI analyzer, Convolutional neural network (CNN), dice similarity coefficient (DSC), and formalin-fixed paraffin) are acronyms. SVMs, or support vector machines; Hematoxylin and eosin, or H&E; DCNN (deep convolutional generative neural network), SC-CNN (spatially constrained convolution neural network), DCGAN (deep convolutional generative adversarial network), RCNN (region-cNN), RF (randomized forest), and both two and three-dimensional (2D and 3D). Usually, cross-validation accounts for more variation between potential differentiates training, validation, and test data and is stronger than hold out. Cross-validation takes longer than the straightforward holdout strategy, though. Certain structural characteristics of a sample of science can be measured simultaneously using various sequencing techniques. Statistical and machine learning methods use feature selection or differential analysis to increase efficiency and decrease overfitting. The acquired omics and clinical variables are then concatenated by ML models as input for the detection along with evaluation of lung cancer. It is thought to be a dependable technique to investigate possible circulating tumor markers using fragments of DNA without cells (cf DNA), circulatory tumor genetic material (ctDNA), exosomes, its methylation, nucleotide (miRNA, or messenger), and circulate carcinoma cells (CTCs) [63], as in Fig. 3. Numerous By combining these liquid biopsy data, classification algorithms such as SVM, among others radio frequency (RF), and the LR may have used to identify tumors with high detection rates [78–81].

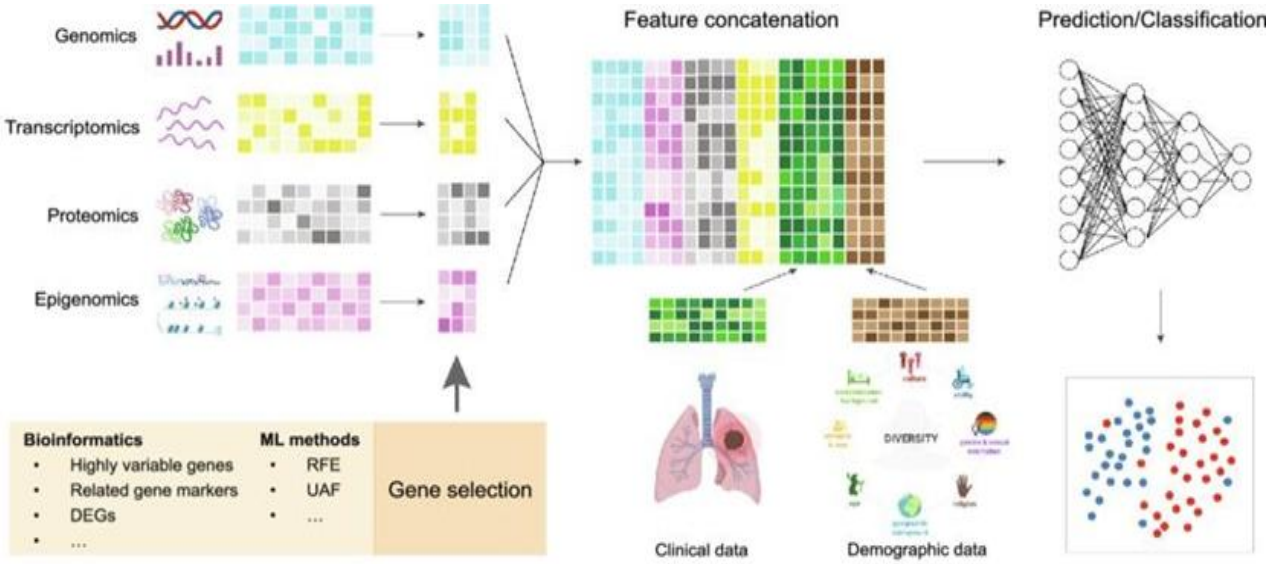

Fig. 3: Omics dataset, also investigation for studies on cancer of the lung.

For the classification of lung cancer subtypes, somatic mutations, including updates, alterations, and single-nucleotide mutations (sequence variants), usually exhibit unique disease signatures [82]. Classifiers are trained for LUAD–LUSC categorization, research has thus used somatic mutations as input characteristics [83]. Numerous of these mutations, particularly driver mutations, have the ability to alter expression levels, which affects gene function and disrupts signaling pathways within cells. Consequently, distinct cancer types have varying degrees of protein expression [84 - 85]. These distinct expression profiles of cancer types impose constraints on machine learning models, which can use. To categorize patients' malignancy (benign or malignant) and subtypes.), genome sequencing is used as the data source [86–89]. Similarly, it has been demonstrated that there is a substantial association between differential gene expression and copy number variation (CNV, which stands), which is extensively present in cancer cells [90]. Thus, CNVs can also be used for developing neural network models for specific cancer classification in lung cancer research [81,91,92]. Notably, a recurrent hidden Markov model (HMM) was proposed by Daemen et al. [92] and offers good classification accuracy for locating larger chromosomal sites with changing copy counts. Jurmeister et al. [93] sought to determine if the found malignant nodule was a primary lung cancer or the spread of another cancer. used genetic methylation patterns as input features relatively recently. Whenever every generated gene is used as an attribute of input directly, overfitting could happen [94]. Many research used a variety of computational strategies to choose a few genes associated with cancer in to improve their models for machine learning (see Fig. 3). Algorithms that use machine learning to learn were used to select characteristics in a number of studies. To choose the optimal markers for model training, Liang et al. the article [80] and Whitney et al. [86], for instance used the least absolute shrinkage and selection operator (LASSO) technique, whereas Aliferis et al. [89] was determined selected extensively cancer-associated genes using unilateral association filtering (UAF) models and recursive feature elimination (a request for information) [95]. Further, a number of studies use unsupervised models to cluster sample population classifications, after which marker genes are identified for the various groups [96 - 97]. Some studies combined machine learning-based models with statistical methods to pick characteristics. Shankar along with colleagues [81] strengthened the CNV feature by creating an identical pattern characteristic deviation (also known as the score rather than using copy number variations as the input feature directly. This method is more stable and less vulnerable to variations in sample quality. Several statistical tests, including ordinary fold changes, ordinary t-tests, SAM-statistics, and moderated t-statistics, were combined by Daemen et al. [92] in order to choose a gene with substantial differential expression collection.

Table. 2: ML-related publications on early diagnosis and detection with sequencing data

Publication	ML Method	Sample Size	Sequencing Data Type	Performance	Validation Method	Feature Selection	Highlight / Advantage	Shortcoming
Mathios et al. [78]	LR model with a LASSO penalty	799	cfDNA fragment	AUC (0.98)	10-fold cross-validation	cfDNA fragment features, clinical risk factors, and CT imaging features	Framework for combining cfDNA fragmentation profiles with lung cancer markers	DNA variations in late-stage disease may affect cfDNA detection
Chabon et.al., [79]	5-NN; 3-NN; NB; LR; DT	160	ctDNA	AUC (0.69–0.98)	Leave-one-out cross-validation	SNV + CNV features	Establishes ML framework for early detection of lung cancers using cfDNA	Sampling bias exists (most are smokers)
Liang et al. [80]	LR	296	ctDNA	AUC (0.816)	10-fold cross-validation	Nine DNA methylation markers	ML framework for lung cancer detection using DNA methylation markers	Limited by small number of methylation markers
Raman et al. [81]	RF; SVM; LR with ridge, elastic net; LASSO regularization	843	cfDNA	mAUC (0.896–0.936)	Leave-one-out cross-validation	Copy number profiling of cfDNA	Framework for using copy number profiling of cfDNA as lung cancer biomarker	Feature selection methods may reduce overfitting but can limit AUC
Kobayashi et al. [83]	Diet Networks with EIS	954	Somatic mutation	Accuracy (0.8)	5-fold cross-validation	SNVs, insertions, and deletions across 1796 genes	EIS stabilizes Diet Networks training process	Interpretability varies between datasets
Whitney et al. [86]	LR	299	RNA-seq of BECs	AUC (0.81)	10-fold cross-validation	Lung cancer-associated and clinical covariate RNA markers	Model keeps sensitivity for small and peripheral lesions	Selected genes vary across datasets
Podolskiy et al. [87]	KNN; NB; SVM; C4.5	529	RNA-seq	AUC (0.91)	Hold-out	RNA-seq	Systematically compares models for lung cancer subtype classification	Feature selection needed to reduce overfitting
Choi et al. [88]	Ensemble model with elastic net LR; SVM; hierarchical LR	2285	RNA-seq of bronchial brushing samples	AUC (0.74)	5-fold cross-validation	RNA-seq of 1232 genes with clinical covariates	Integrates RNA-seq features and clinical info for risk prediction	Sample size small in some subgroups; unbalanced data
Alharbi et al. [89]	Linear SVM; polynomial-kernel SVM; RBF-kernel SVM; KNN;	203	RNA-seq	AUC (0.8733–0.9980)	5-fold cross-validation	RNA-seq of selected genes using RFE and UAT	Different gene selection algorithms improve classification	Selected genes vary across training cohorts

	NN							
Aliferis et al. [91]	DT; KNN; linear SVM; polynomial-kernel SVM; RBF-kernel SVM; NN	37	CNV measured by CGH	Accuracy (0.892)	Leave-one-out cross-validation	Copy number of 80 selected genes based on linear SVM	Compares models for lung cancer subtype classification	Small sample size

3. Apply ML to lung cancer treatment response and survival prediction

3.1 Prognosis and therapy response prediction

Advanced machine learning algorithms have been used to enhance the evaluation and prediction of cancer intervention response [98 - 99] and have shown progress in maximizing treatment choices factors improve the chances of a profitable recovery (Fig. 4; Table. 2, contains factors [100 - 101]). One of the many metrics available for evaluating the response to cancer therapy is the response assessment criteria in Solid RECIST cancers [102]. To determine how cancers grow or decrease in people, the RECIST definition uses imaging data, mostly from CT and MRI [103]. A thorough CNN method was created by Jiang et al. [104] to monitor changes in tumor volume from CT scans. Dual big networks built from a full-resolution persistent system [105], for instance model were employed in their CNN model. combining several residual streams of different resolutions to at once incorporate data at different resolutions for lung tumor segmentation. In light of the patient's Rtl genetic mutation profile, Qureshi [106] developed an RF model based on the RECIST criterion to predict the RECIST level after EGFR tyrosine kinase inhibitor (TKI) therapy. The approach used clinic data, structural features, and energy properties from a patient's EGFR mutant medication combo to train the classifiers and enhance prediction performance. A novel metric known as the tumor proportional score (TPS), which is calculated by estimating the share of cells associated with tumors in digital histology pictures was created by the authors of a recent study that evaluated the response to radiation treatment for lung cancer [107]. Scientists identified positive tumor cell areas (technical issues+) and negative tumor cell regions (TC) by combining the Otsu threshold with an auxiliary classifier generative adversarial network (AC-GAN) model [108].

The TPS number was ultimately calculated by dividing the number of photos in each tumor cell region that were discovered by the image counts of the counsellor + regions. In a different study, Geeleher et al. [109] measured medication response using half-maximal inhibitory dosage (IC50).

Using a ridge regression model, the authors calculated the IC50 percentages for various lines of cells according to their whole-genome expression level [110]. A phenotype representation learning (PRL) system was more recently developed through Quiros et al. [111] for spatial grouping cell type labeling on histopathology pictures, using community detection and self-supervised learning. Their clustering results can also be utilized to detect tumor recurrence and track histological tumor growth patterns. In fact, their model has performed well in both the LUAD and LUSC classes.

3.2 Survival prediction

Inclinical oncology, predicting survival and prognosis is a challenging but crucial task for doctors since patients can benefit from cost management and treatment decisions based on the survival duration [112 – 114]. Predictions for the majority of medical history were mostly reliant on the medical professional's knowledge and experience, which was based on medical information and prior patient histories. But studies have revealed that physicians frequently overestimate survival time and perform badly when estimating prognosis and survival expectancy [115 – 117]. Numerous studies [119–122] have used statistical methods, including the Predictions made by doctors using the Cox proportional hazards model [118] are not very accurate [12]. across contrast, machine learning has proven to be capable of predicting a patient's prognosis and survival across a variety of datasets, including genetic, transcriptomic, proteomic, radiomic, and others (as shown in Fig. 4; Fig. 3).

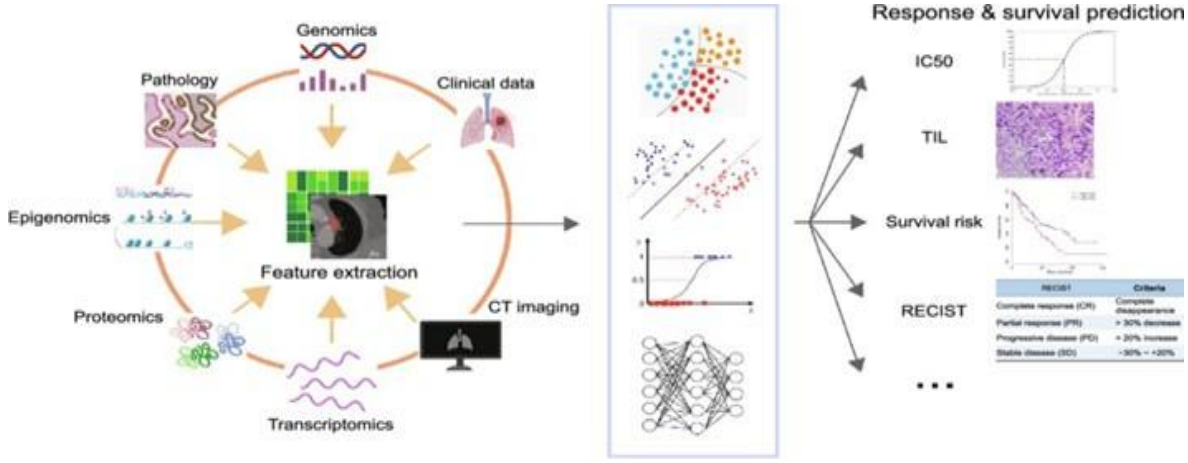

Fig. 4: An example of predicting lifetime and responsiveness to therapy using machine learning

Utilizing a 3-decade survival accepted, Chen and others [123] separated patients into low-risk (presence lifetime > 36 years) and high-risk (longevity time is 36 a period or time groups. After that, they developed a neural network (NN) model to use patient characteristics and gene expression data to assess a patient's risk. Their strategy examined four aspects.

3.3 Applications of Machine Learning in Lung Cancer

Information of sensor gene expression and only five genes were found to be linked with survival time, yielding an overall accuracy of 83.0%. Gene expression data was also used by Liu et al. [124] to classify survival after three years. In contrast to Chen et al. [123], the authors selected a total of 22 genes to increase the stability of their model by combining three different forms of sequencing data: RNA sequencing, DNA methylation, and DNA mutation.

Cho et al. [126] and LUADpp [125] modelled an estimation of 3-year mortality risk based on the somatic mutations as input features. Cho et al. [126], for instance employed chi-squared tests, whereas LUADpp [125] employed a published genome-wide rate comparison test [127] that could measure gene mutation rates while identifying a trade-off between statistical power and accuracy. To determine which genes had to do with the highest significant mortality. Multi-omics information on tumors has been used in numerous experiments for analysis. Due to the intricacy of survival prediction. It is more difficult to precisely identify the most important genes for prediction from multi-omics data than from single-omics data. Several studies [128 – 131] created a comparable workflow to address the problem. Using their multi-omics data, they first created a matrix that showed how similar the patients were to one another.

They next divided the patients into two clusters using the acquired matrix and using an informal clustering approach, generally a self-encoding model with K-means algorithm clustering. Based on the different results of survival between the two clusters in Kaplan–Meier analysis, the two clusters were designated as "both high-risk" and "low-risk." A statistical model such as [128 - 129] or an ML model [130 - 131] was used to extract the genes linked to mortality after the variations in survival outcomes for further analysis

4. Utilize ML in immunotherapy for lung cancer

4.1 Predicting the response to immunotherapy

The importance of treatment has increased recently. T cells are often stimulated, aiding an individual's defence mechanism in its fight against cancer. Numerous novel immunotherapy treatments for lung cancer that are presently being evaluated have developed into standard immunotherapy elements. Immune checkpoint inhibitors (ICIs), a particular form of therapy for programmed mortality gene 1 (PD-1)/programmed cell death protein ligand 1 (PD-L1), are believed to have proven to be beneficial for patients with non-small cell lung cancer, also called NSCLC [132 - 134]. Immunotherapy is not yet as popular as radiation, chemotherapy, or surgery, though. One explanation is that because each Every individual has a distinct tumor immune microenvironment and not all patients benefit from it.

Table. 3: ML-related publications on treatment response and survival forecasting

Publication	Feature Extraction Method	Prediction Model	Sample Size	Data Type	Performance	Validation Method	Feature Selection/Input	Highlight / Advantage	Shortcoming
Jiang et al. [104]	MRRN-based model	MRRN-based model	1210	CT Images	DSC (0.68–0.75)	5-fold cross-validation	3D image features	Accurately tracks tumor volume changes from CT images across resolutions	Does not predict accurately with small tumor size
Qureshi [106]	NA	RF; SVM; KNN; LDA; CART	201	Molecular structure and somatic mutations of EGFR	Accuracy (0.975)	10-fold cross-validation	4 clinical features + 4 protein drug interaction features + 5 geometrical features	Integrates multiple features for better benchmarking performance	Only considers 33 of 594 EGFR mutations from COSMIC database
Kapil et al. [107]	AC-GAN	AC-GAN	270	Digital pathology images	Lcc (0.94); Pcc (0.95); MAE (8.03)	Hold-out	PD-L1-stained tumor section histological slides	Better performance than benchmarked, fully supervised models	PD-L1 staining in TPS evaluation may lack accuracy
Chen et al. [123]	Chi-square test + NN	NN	440	RNA-seq	Accuracy (0.83)	Hold-out	RNA-seq of 5 genes	Uses multiple lab datasets for robust training	Doesn't consider demographic and clinical features that may affect prediction
Yu, et.al., [125]	Top genes with most significant mutation frequency difference	SVM	371	Somatic mutations	Accuracy (0.81)	5-fold cross-validation	85 somatic mutation features	High accuracy using only 7 gene mutation features	Sampling bias and linkage disequilibrium may affect feature selection
Cho et al. [126]	Info gain, Chi-squared test, min redundancy max relevance, correlation	NB; KNN; SVM; DT	471	Somatic mutations	Accuracy (0.68–0.88)	5-fold cross-validation	19 somatic mutation features	Uses 4 methods for feature selection to improve performance	Only one dataset used for training
Yu et al. [128]	Info gain ratio; hierarchical clustering	RF	538	Multi-omics (histology, pathology)	AUC (> 0.8)	Leave-one-out cross-validation	15 gene set features	Integrative omics-pathology model to improve classification accuracy	Cox models may be over fitted in multi-omics settings

Predicting a patient's reaction to immunotherapy is therefore essential when treating cancer. Recently, AI-based techniques constructed for estimating immunotherapy reactions based on immunological sequencing signatures and medical imaging (Fig. 4; Table 3 as refer) [135]. In order for foreseeing the response to PD-1/PD-L1 blocking therapy, Wiesweg et al. [136] trained an SVM classifier of RECIST

classification using gene expression profiles of seven key genes extracted from ML models and 25 tumor type-specific variables as input features.

Evaluation of tumor-infiltrating lymphocytes. For assessing the response to immunotherapy, the proportion of lymphocytes infiltrating tumors (TILs) is another crucial. To carry out technique, DeepTIL [139] improved the cell deconvolution model CIBERSORT [140]. Two recent studies [137 - 138] trained a machine learning algorithm, incorporating the results of the and RF, for instance for RECIST classification using radiomic biomarkers and other imaging characteristics specific to tumor lesions using contrast-enhanced computing tomography (CE-CT) scans employing transcriptomic data to automatically determine the proportion of leucocyte subsets, such as lymphocytes, T cells that are CD4+, CD8+ lymphocytes, CDsGranulocytes, T cells, lymphocytes, and Mo-Ma-DC cells cancerous sample. An alternative method [141] used the sequencing of RNA of 20,530 genes as biomarkers and 84 radioactive characteristics from the CE-CT scans to build a straight-line elastic-net regression model to estimate amount of CD8+ T cells. In another study, a DL model was created [142] to identify TILs in digitally preserved hydrogen peroxide (pictures, see Table 3). Features are extracted using a convolutional autoencoder (CAE) [143], and TIL areas are classified using VGG 16-layer network [144]. The "cellular death CNN" was designed to find TILs in a necrotic region. They used the DeconvNet [145] model for the TIL segmentation in "cell death CNN" since it has shown good accuracy using different benchmark imaging datasets.

4.2 Neo antigen prediction

ML algorithms have provided insight into immunotherapy predicting response to immunotherapy as well as neoantigens. Neoantigens are tumor specific modified peptides generated by somatic mutations in tumor cells that have the ability to elicit anticancer immune responses [146–148]. development and improving neoantigen-targeted immunity medicines benefit from immunogenic neoantigens, according to recent studies [149–152]. In accordance with neoantigen studies in clinical trials, neoantigens were developed utilizing current algorithms that employ machine learning based on the processing and presentation of human leukocyte antigen (HLA) class I and II [153 – 157]. ML models can use the identified somatic mutations to determine the binding affinity of encoded changed peptides to the patient's HLA alleles (peptide–HLA binding affinity). It is possible to further forecast the neoantigens using the predicted peptide–HLA binding affinity. NetMHC [158 - 159] used a receptor–ligand dataset with 528 peptide–HLA binding interactions assessed by Buus et al. [160] to train a combination of different NNs for neo-peptide affinity prediction. Inorder to increase the prediction's accuracy, NetMHCpan [161-162] trained their NN model using a bigger dataset that included 37,384 distinct peptide–HLArelationships between 18 HLA-B and 24 HLA-A alleles (26,503 and 10,881 for A and B alleles, respectively). Both tools have been used to study the neoantigen architecture in lung cancers [146, 163–165].

5. Challenges and opportunities towards the decades to come

There are still issues to be resolved even with the extensive application of ML studies in clinical treatment research treating lung cancer. In order to educate readers about lung cancer therapeutic analyses and the difficulties facing ongoing lung cancer research, we have included a few examples of recent machine learning algorithms here, particularly the increasingly significant and well-liked DL algorithms of the last ten years.

5.1 Imaging data analysis

For clinical usage, it is essential to learn techniques to successfully glean information from data from images. In the prior machine learning-based Computer-aided design system, feature extractions were often based on texture, shape, and intensity of a dubious region in the image, along with other clinical variables [166]. However, these techniques are subjective and may not be able to restore a dubious nodule's natural features. This can be achieved via a DL-based CAD system that use models generated by CNN for extraction of features from raw imaging data with multilevel representations and hierarchical abstraction [167–169]. Unlike previous methods, CNN model features are not generated through other people. accurately and thoroughly represent the nodule's inherent characteristics. The most recent

advancement in computer vision technology is the Vision Transformer (ViT) [170-171]. ViT was more reliable while training on fewer datasets and performed better than CNN in terms of accuracy and computational economy, outperforming nearly 4 [172]. ViT models have demonstrated their Although they weren't utilized in any cancer screening imaging studies that we are aware of, they show potential as an antagonist to CNN in imaging information analysis.

5.2 Analysis of Omics datasets

Ddl is an extension of machine learning that relies on programmable neural networks to produce precise judgments. When it comes to challenging issues like picture classification, it excels. We examined the applicability of DL models with imaging datasets in this paper. In clinical trials of lung cancer utilizing omics data, DL algorithms were less common than in imaging datasets. Nonetheless, in other areas of omics analysis, computational models of DL have been extensively employed. For example, as genomic CNN models [174] and recurrent artificial neural network (a neural network) algorithms [173] are useful tools for researching population genetics when the data are continuous sequences. Furthermore, because the input dimension of omics data is typically very large, auto encoders or deep generative models have been utilized in a number of studies for feature extraction and dimensionality reduction in order to increase efficiency and prevent overfitting [175]. Meanwhile, by combining multi-omics data, self-supervised representation learning models can overcome the curse of dimensionality and aggregate knowledge about several aspects of the same tissue samples [176]. Alongside the introduction of geographically based DL models, several DL models are becoming more and more popular for computationally intensive analysis. [178] and single cell-based [177] methods that are being deployed in molecular studies.

To manage complexity of massive genomics data, unsupervised deep clustering methods developed for cell population subtype annotation [180–183] and population structure identification [179]. Additionally, to manage complex structure of multi-omics data, graph neural network (GNN) models are increasing being applied in biomedical classification [185], forecasting prognostics [186], dataset integration [184], and other domains. Although these researches have not been directly used for clinical analysis of lung cancer, they are a good source of inspiration for applying DL approaches to complex cancer related omics datasets.

5.3 Multi-view data and multi-database integration

Large volumes of clinical records, data from images, and multiomics data from an individual patient are frequently accessible these days. Combining this data offers a thorough understanding of. Platform noise is unavoidable amongst these data kinds, nevertheless, because they are usually gathered from various platforms. For instance, complex data standardization, data fusion, and data integration are typically obstacles in imaging data processing, particularly radiomics. To manage versatility together, multimodality healthcare segmentation networks have recently designed. Medical pictures in order to get around this restriction [187]. Likewise, batch noise (also known as the batch effect) occurs between databases for sequencing data types.

It is crucial for cancer treatments to eliminate, we can better investigate the processes behind cancer resistance to treatment and recurrence using datasets from numerous systems and batch effects. Despite the fact that integrative tools have been tested and/or benchmarked in biomedical investigations [188–191], these studies are insufficiently thorough and discriminating to address tool selection in relation to biologically relevant problems.

5.4 Model generalizability and robustness

According to this assessment, an ML algorithm's performance typically changes depending on the dataset. It is the database volume effect's involvement, that we identified previously covered could be one explanation. The lack of robustness and generalizability, however, may be further issues that prevent these ML models from being used in clinical trials. Additionally, the majority of research chose marker genes prior to classification using statistical models or machine learning models in order to minimize overfitting. These marker genes, however, typically vary significantly between research, suggesting that the identified marker genes are not biologically interpretable or generalizable. To increase a model's robustness and generalizability, gaining a

deeper comprehension of robustness concerns in various machine learning it is essential to bridge gap in robustness techniques across domains and develop new designs. By employing transfer learning to use model that was already trained when learning their own datasets in lung cancer imaging data processing, for example, current research has improved both the effectiveness and the durability of their CNN-based models [38, 55, 192]. The use of transfer learning in deep neural networks (NNs) to provide a generalizability technique for sequencing datasets can be used as an example of developing a universal and trustworthy model for the interpretation of cancer gene sequencing data [193]. Moreover, DL is an advanced black-box model. An understanding of the mechanics behind a DL system in clinical research may facilitate the building of a standardized and unites deep learning infrastructure to improve its efficiency and reliability has been made available by explainable AI (XAI) models [194 - 195]. These techniques may explain a model both locally and globally, which aids researchers in fine-tuning hyper parameters from various models with great effectiveness [196, 197].

5.5 Performance evaluation metrics

The creation of algorithms for clinical research is typically the subject of investigations. Though it typically plays a significant role in ML systems, the choice of measures for assessing these algorithms' performance is frequently disregarded [198]. Based on this review, accuracy and under the curve (also known as AUC) are the two most commonly used metrics (see Tables 1-3). However, these measures ought to be transformed into ones that can be discussed clinically. Since they do not always represent the clinical needs. In some situations, such as with patients who are at a high risk of visiting the emergency room, sensitivity or specificity may be more related to clinical needs than accuracy [199].

5.6 Making clinical decisions

According to the study, most patients' total expenses for lung cancer treatment would surpass $ 50,000 [200], and most families would find the expense to be prohibitive. Therefore, precise prognostic prediction and decision-making will open the door to individualized care. Current DL models have been applied to optimize the combination of various therapies and pharmaceuticals as well as forecast the efficacy of a therapy or medication [201-202]. Nevertheless, the majority of DL models now in use for clinical decision-making struggle to keep up with the rapid changes in health care data and/or knowledge [203]. Clinical decision support systems, such as Google DeepMind Health and IBM Watson Health, have recently been used in the treatment of lung cancer [204-205]. Even though these technologies have helped to increase the efficiency of clinical work, in terms of clinical trials, they are still below ideal and are currently unable to satisfy the duties of clinicians [205].

6. Conclusion

Intelligence enables us to research carcinoma of the lung from a different perspective and explores the potential applications of decision support systems in precision oncology. In this analysis, we examined the most recent developments in machine learning algorithms in a number of lung cancer treatment domains, such as Immunotherapy practice, rapid assessment, prognostic forecasting, and medication response evaluation. We carefully compiled datasets (Table S1), baseline techniques (Table S2), and method attributes (Tables 1–3) to support future machine learning advancements in lung cancer treatments. Finally, we outlined the issues that still require attention, including the deficiency of medical data labels for training in terms of both quantity and quality, the significance of biomedical need for data integration and batch removal, anxiety about performance evaluation criteria, and the explanations and model robustness for clinical usage. Plans According to this study, imaging and omics data will be utilized during lung cancer treatments thus, a multi-modal clinical analysis tool for machine learning should be employed. System that considers how multiple information kinds should be unified as well as the treatment of both imaging and omics data. Lastly, we anticipate that these difficulties may spur additional research aimed at lung cancer.

Conflict of interest

The authors declared 'No conflict of interest'.

References

[1] J. D. Minna, J. A. Roth, and A. F. Gazdar, "Focus on lung cancer", *Cancer cell*, Vol. 1, No. 1, pp. 49-52, 2002. https://doi.org/10.1016/S1535-6108(02)00027-2

[2] E. Svoboda, "Artificial intelligence is improving the detection of lung cancer", *Nature*, Vol. 587, No. 7834, pp. S20–S22, 2020. https://doi.org/10.1038/d41586-020-03157-9

[3] S. Ling, Z. Hu, Z. Yang "Extremely high genetic diversity in a single tumor point to the prevalence of non-Darwinian cell evolution", *Proceedings of the National Academy of Sciences of the USA*, Vol. 112, No. 47, pp. E6496–E6505, 2015. https://doi.org/10.1073/pnas.1519556112

[4] T. J. Hudson et al., "International network of cancer genome projects," *Nature*, Vol. 464, No. 7286, pp. 993–998, 2010. https://doi.org/10.1038/nature08987

[5] J. N. Weinstein et al., "The Cancer Genome Atlas Pan-Cancer analysis project", *Nature Genetics*, Vol. 45, No. 10, pp. 1113–1120, 2013. https://doi.org/10.1038/ng.2764

[6] K. Clark et al., "The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository", *Journal of Digital Imaging*, Vol. 26, No. 6, pp. 1045–1057, 2013. https://doi.org/10.1007/s10278-013-9622-7

[7] A. Pavlopoulou, D. A. Spandidos, and I. Michalopoulos, "Human cancer databases (Review)", *Oncology Reports*, Vol. 33, No. 1, p. 3-18, 2014. https://doi.org/10.3892/or.2014.3579

[8] Y. Luo, F. Wang, and P. Szozovits, "Tensor factorization toward precision medicine", *Briefings in Bioinformatics*, Vol. 18, No. 3, pp. 511–514, 2017. https://doi.org/10.1093/bib/bbw026

[9] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications", *SIAM Review*, Vol. 51, No. 3, pp. 455–500, 2009. https://doi.org/10.1137/07070111X

[10] G. Chao, C. Mao, F. Wang, Y. Zhao, and Y. Luo, "Supervised nonnegative matrix factorization to predict ICU mortality risk", *2018 IEEE International Conference on Bioinformatics Biomedicine*, pp. 1189–1194, 2018. https://doi.org/10.1109/BIBM.2018.8621403

[11] M. Ghafoorian, A. Mehrtash, T. Kapur, C. R. G. Guttaman, C. M. Tempany, W. M. Wells, et al., "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation", International Conference on Medical image *Computing and computer Assisted intervention*, pp. 516–524, 2017. https://doi.org/10.1007/978-3-319-66179-7_59

[12] L. Xu, Z. Huang, Z. Zeng, J. Li, H. Xie and C. Xie, "An Integrative analysis of DNA methylation and gene expression to predict lung adenocarcinoma prognosis", *Frontiers in Genetics*, Vol. 13-2022, art. no. 970507, 2022. https://doi.org/10.3389/fgene.2022.970507

[13] X. Y. Wang, S.H. WU, J. Ren, L. Y. Zeng, Li. Li. Guo "Predicting Gene Comutation of the EGFR and TP53 by Radiomics and Deep Learning in Patients with lung adenocarcinomas", *Journal of Thoracic Imaging*, Vol. 40, No. 3, art. no. e0817, 2025. https://doi.org/10.1097/rti.0000000000000817

[14] E. Vassella, S. Langsch, M. S. Dettmer, C. Schlup, M. Frattini et al., "Molecular Profiling of lung adenosquamous carcinoma: hybrid or genuine type?", *Oncotarget*, pp. 23905-23916, 2015. https://doi.org/10.18632/oncotarget.4163

[15] R. Ramirez, Y. C. Chiu, S. Zhang, J. Ramirez et al., "Prediction and interpretation of cancer survival using graph convolution neural networks", *Methods*, Vol. 192, pp. 120-130, 2021. https://doi.org/10.1016/j.ymeth.2021.01.004

[16] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. McConnell, G. Corrado, et al., "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning", *Nature Biomedical Engineering*, Vol. 2, pp. 158–164, 2018. https://doi.org/10.1038/s41551-018-0195-0

[17] R. Li, W. Zhang, J. Li et al., "Deep learning-based imaging data completion for improved brain disease diagnosis", *Medical image Computing and computer – Assisted intervention*, pp. 305–312, 2014. https://doi.org/10.1007/978-3-319-10443-0_39

[18] W. Bai, C. Chen, G. Tarroni, J. Duan, F. Guitton, P. M. Matthews, et al., "Self-supervised Learning for cardiac MR Image segmentation by Anatomical position prediction", *International*

Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 541-549, 2019. https://doi.org/10.1007/978-3-030-32245-8_60

[19] F. Prior, K. Smith, A. Sharma, J. Kirbyet al., "The public cancer radiology imaging collections of the cancer imaging achieve", *Scientific Data*, Vol. 4, art. no. 170124, 2017. https://doi.org/10.1038/sdata.2017.124

[20] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. B. Sloan, J. E. V. Vega, et al., "Predicting cancer outcomes from histology and genomics using convolutional networks", *Biophysics and Computational Biology*, Vol. 115, No. 13, pp. E2970–E2979, 2018. https://doi.org/10.1073/pnas.1717139115

[21] Y. Li and Y. Luo, "Optimizing the Evaluation of gene-targeted panels for tumor mutational burden estimation", *Scientific Reports*, Vol. 11, art. no. 21072, 2021. https://doi.org/10.1038/s41598-021-00626-7

[22] S. Bhalla, A. Lagana, "Artificial intelligence for precision oncology", *Advances in Experimental Medicine and Biology*, Vol. 1361, pp. 249-268, 2022. https://doi.org/10.1007/978-3-030-91836-1_14

[23] Z. Zeng, S. Espino, A. Roy, X. Li, S. A. Khan, S. E. Clare, et al., "Using natural language processing and machine learning to identify breast cancer local recurrence", *BMC Bioinformatics*, Vol. 19, art. no. 498, pp. 498, 2018. https://doi.org/10.1186/s12859-018-2466-x

[24] Y. Luo, A. R. Sohani, E. P. Hochberg, and P. Szolovits, "Automatic lymphoma classification with sentence subgraph mining from pathology reports", *Journal of the American Medical Informatics Association*, Vol. 21, No. 5, pp. 824–832, 2014. https://doi.org/10.1136/amiajnl-2013-002443

[25] C. Luchini, A. Pea, and A. Scarpa, "Artificial intelligence in oncology: Current applications and future perspectives", *British Journal of Cancer*, Vol. 126, No. 4, pp. 4–9, 2022. https://doi.org/10.1038/s41416-021-01633-1

[26] Z. Zeng, A. Vo, X. Li, A. Shidfar, P. Saldana, L. Blanco, et al., "Somatic genetic aberrations in benign breast disease and the risk of subsequent breast cancer", *NPJ Breast Cancer*, Vol. 6, art. no. 24, 2020. https://doi.org/10.1038/s41523-020-0165-z

[27] J. Na, N. Zong, C. Wang, D. E. Midthun, Y. Luo, P. Yang, et al., "Characterizing phenotypic abnormalities associated with high-risk individuals developing lung cancer using electronic health records from the All of Us research workbench", *Journal of the American Medical Informatics Association*, Vol. 28, No. 11, pp. 2313–2324, 2021. https://doi.org/10.1093/jamia/ocab174

[28] H. Fujita, "AI-based computer-aided diagnosis (AI-CAD): the latest review to read first", *Radiological Physics and Technology*, Vol. 13, pp. 6 - 9, 2020. https://doi.org/10.1007/s12194-019-00552-4

[29] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments", *Expert Systems with Applications*, Vol. 138, art. no. 112821, 2019. https://doi.org/10.1016/j.eswa.2019.112821

[30] Y. Abe, K. Hanai, M. Nakano, Y. Ohkubo, T. Hasizume, T. Kakizaki, et al., "A computer-aided diagnosis (CAD) system in lung cancer screening with computed tomography", *Anticancer Research*, Vol. 25, No. 6, pp. 483–488, 2005. [Cross Ref]

[31] A. McWilliams, M. C. Tammemagi, J. R. Mayo, H. Roberts, G. Li, K. Soghrati, et al., "Probability of cancer in pulmonary nodules detected on first screening CT", *The New England Journal of Medicine*, Vol. 369, No. 10, pp. 910–919, 2013. https://doi.org/10.1056/nejmoa1214726

[32] S. J. van Riel, F. Ciompi, M. M. W. Wille, A. Dirksen, S. Lam, E. T. Scholten, et al., "Malignancy risk estimation of pulmonary nodules in screening CTs: Comparison between a computer model and human observers", *PLoS One*, Vol. 12, No. 11, art.no. e0185032, 2017. https://doi.org/10.1371/journal.pone.0185032

[33] M. M. W. Wille, S. J. van Riel, Z. Saghir, A. Dirksen, J. H. Pedersen, B. de Hoop, et al., "Predictive accuracy of the PanCan lung cancer risk prediction model—External validation based on CT from the Danish Lung Cancer Screening Trial", *European Radiololgy*., Vol. 25, No. 10, pp. 3093–3099, 2015. https://doi.org/10.1007/s00330-015-3689-0

[34] M. Kriegsmann, R. Casadonte, J. Kriegsmann, H. Dienemann, P. Schirmacher, J. H. Kobarg, et

al., "Reliable entity subtyping in non-small cell lung cancer by matrix-assisted laser desorption/ ionization imaging mass spectrometry on formalin-fixed paraffin-embedded tissue specimens", *Molecular& cellular Proteomics*, Vol. 15, No. 10, pp. 3081–3089, 2016. https://doi.org/10.1074/mcp.m115.057513

[35] B. A. Mohammed, P. C. Brennan, and C. Mello-Thoms, "A review of lung cancer screening and the role of computer-aided detection", *Clinical Radiololgy*, Vol. 72, No. 6, pp. 433–442, 2017. https://doi.org/10.1016/j.crad.2017.01.002

[36] K. Murphy, B. V. Ginneken, A. M. R. Schilham, B.J.de hoop, H. A. Gietema and M. Prokop, "A Large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k – nearest neighbor classification", *Medical Image Analysis*, Vol. 13, No. 5, pp. 757-770, 2007. https://doi.org/10.1016/j.media.2009.07.001

[37] M. Buty, M. Gao, U. Bagci, A. Wu, and D. J. Mollura, "Characterization of lung nodule Malignancy using hybrid shape and appearance features", *Medical Image Computing and Computer Assisted Intervention – MICCAI* 2016, pp. 662–670, 2016. https://doi.org/10.1007/978-3-319-46720-7_77

[38] J. Zhang, M. Yang, W. Guo, B. A. Xavier, M. Bolen, and X.Li, "Detection-Guided Deep Learning – Based model with spatial regularization for lung nodule segmentation," *Quantitative Imaging in Medicine and Surgery*, Vol. 15, No. 5, pp. 4204-4216, 2025. https://doi.org/10.21037/qims-2024-2511

[39] F. Ciompi, B. de Hoop, S. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, et al., "Automatic classification of pulmonary perifissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out of the box," *Medical Image Analysis*, Vol. 26, No. 1, pp. 195-202, 2015. https://doi.org/10.1016/j.media.2015.08.001

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, No. 6, pp. 84–90, 2017. https://doi.org/10.1145/3065386

[41] F. Z. Ouadiay, H. Bouftaih, E. H. Bouyakhf, "Simultaneous object detection and localization using convolutional neural networks," *International conference on intelligent system and computer vision*, pp. 1-8, 2018. https://doi.org/10.1109/ISACV.2018.8354045

[42] J. Hussain, M. Bath, J. Ivarsson, "Generative adversarial networks in medical image reconstruction: A systematic literature review", *Computers in biology and medicine*, Vol. 191, art. no. 110094, 2025. https://doi.org/10.1016/j.compbiomed.2025.110094

[43] X. Gu, Y. Wang, T. F. Chan, P. M. Thompson, and S. T. Yau, "Genus zero surface conformal mapping and its application to brain surface mapping", *Information Processing Medical. Imaging*, Vol. 18, pp. 172–184, 2003. https://doi.org/10.1007/978-3-540-45087-0_15

[44] K. V. Venkadesh, A. A. A. Setio, A. Schreuder, E. T. Scholten, K. M. Chung, M. M. W. Wille, et al., "Deep learning for malignancy risk estimation of pulmonary nodules detected at low-dose screening CT", *Radiology*, Vol. 300, No. 2, pp. 438–447, 2021. https://doi.org/10.1148/radiol.2021204433

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", *IEEE Conference Computer Vision Pattern Recognition*, pp. 770–778, 2015. https://doi.org/10.48550/arXiv.1512.03385

[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going deeper with convolutions", *IEEE Conference Computer Vision Pattern Recognition*, pp. 1–9, 2015. https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594

[47] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, et al., "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography", *Nature Medicine*, Vol. 25, pp. 954–961, 2019. https://doi.org/10.1038/s41591-019-0447-x

[48] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN", *Computer Vision Pattern Recognition*, 2017. https://doi.org/10.48550/arXiv.1703.06870

[49] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection",

IEEE International conference on computer vision, pp. 2999–3007, 2017. https://doi.ieeecomputersociety.org/10.1109/ICCV.2017.324

[50] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the Kinetics dataset", *IEEE International conference on computer vision and*Pattern Recognition, pp. 4724–4733, 2017. https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.502

[51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, pp. 2818-2826, 2016. https://doi.org/10.1109/CVPR.2016.308

[52] K. AbdulJabbar et.al "Geospatial immune variability illuminates differential evolution of lung adenocarcinoma", *Nature Medicine*, Vol. 26, pp. 1054–1062, 2020. https://doi.org/10.1038/s41591-020-0900-x

[53] S. E. A. Raza et.al., "Micro-Net: A unified model for segmentation of various objects in microscopy images", *Medical Image Analysis*, Vol. 52, pp. 160–173, 2019. https://doi.org/10.1016/j.media.2018.12.003

[54] K. Sirinukunwattana et. al., "Locality-sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images", *IEEE Transactions Medical Imaging*, Vol. 35, No. 5, pp. 1196–1206, 2016. https://doi.org/10.1109/tmi.2016.2525803

[55] N. Coudray et al., "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning", *Nature medicine*, Vol. 24, pp. 1559-1567, 2018. https://doi.org/10.1038/s41591-018-0177-5

[56] Q. Yao, L. Xiao, P. Liu, and S. K. Zhou, "Label-free segmentation of COVID-19 lesions in lung CT", *IEEE Transactions Medical Imaging*, Vol. 40, pp. 2808–2819, 2021. https://doi.org/10.1109/tmi.2021.3066161

[57] M. J. M. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? A visual Turing test for lung cancer diagnosis", *Computer vision and pattern Recognition,* pp. 240–244, 2018. https://doi.org/10.48550/arXiv.1710.09762

[58] J. Li, J. Jia, and D. Xu, "Unsupervised representation learning of image-based plant disease with deep convolutional generative adversarial networks", *Proceedings of Chinese Control Conference,* pp. 9159–9163, 2018. https://doi.org/10.23919/ChiCC.2018.8482813

[59] C. H. Lin, C. J. Lin, Y. C. Li, and S. H. Wang, "Using generative adversarial networks and parameter optimization of convolutional neural networks for lung tumor classification", *Applied Sciences*, Vol. 11, No. 2, art. no. 480, 2021. https://doi.org/10.3390/app11020480

[60] Z. Ren, Y. Zhang, and S. Wang, "A hybrid framework for lung cancer classification", *Electronics*, Vol. 11, No. 10, art. no. 1614, 2022. https://doi.org/10.3390/electronics11101614

[61] P. F. Pinsky et al., "Performance of Lung-RADS in the National Lung Screening Trial: A retrospective assessment", *Annals of Internal Medicine*, Vol. 162, No. 7, pp. 485–491, 2015. https://doi.org/10.7326/m14-2086

[62] National Lung Screening Trial Research Team et al., "Results of initial low-dose computed tomographic screening for lung cancer", *The New England Journal of Medicine*, Vol. 368, No. 21, pp. 1980–1991, 2013. https://doi.org/10.1056/nejmoa1209120

[63] S. Herathet al., "The role of circulating biomarkers in lung cancer," *Frontiers in Oncology*, Vol. 11, art. no. 801269, 2021. https://doi.org/10.3389/fonc.2021.801269

[64] K. Politi and R. S. Herbst, "Lung cancer in the era of precision medicine", *Clinical Cancer Research*, Vol. 21, pp. 2213–2220, 2015. https://doi.org/10.1158/1078-0432.ccr-14-2748

[65] V. Relli, M. Trerotola, E. Guerra, and S. Alberti, "Abandoning the notion of non small cell lung cancer", *Trends in Molecular Medicine*, Vol. 25, No. 7, pp. 585–594, 2019. https://doi.org/10.1016/j.molmed.2019.04.012

[66] J. W. Chen and J. Dhahbi, "Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods", *Scientific Reports*, Vol. 11, art. no. 13323, 2021. https://doi.org/10.1038/s41598-021-92725-8

[67] Z. Zenget al., "Cancer classification and pathway discovery using non-negative matrix factorization", *Journal of Biomedical Informatics*, Vol. 96, art. no. 103247, 2019. https://doi.org/10.1016/j.jbi.2019.103247

[68] W. Jiao et al., "A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns", *Nature Communications*, Vol. 11, art. no. 728, 2020. https://doi.org/10.1038/s41467-019-13825-8

[69] Y. Li and Y. Luo, "Performance-weighted-voting model: An ensemble machine learning method for cancer type classification using whole exome sequencing mutation", *Quantitative Biology*, Vol. 8, pp. 347–358, 2020. https://doi.org/10.1007/s40484-020-0226-1

[70] G. Eraslan, Z. Avsec, J. Gagneur, and F. J. Theis, "Deep learning: New computational modeling techniques for genomics", *Nature Review Genetics*, Vol. 20, pp. 389–403, 2019. https://doi.org/10.1038/s41576-019-0122-6

[71] Y. Luo and C. Mao, "Panther: Pathway-augmented nonnegative tensor factorization for higher-order feature learning", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 1, pp. 37–180, 2021. https://doi.org/10.1609/aaai.v35i1.16113

[72] M. A. Shipp et al., "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning", *Nature medicine*, Vol. 8, pp. 68–74, 2002. https://doi.org/10.1038/nm0102-68

[73] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective", *Artificial Intelligence in Medicine*, Vol. 23, No. 1, pp. 89–109, 2001. https://doi.org/10.1016/S0933-3657(01)00077-X

[74] E. Ferrari, A. Retico, D. Bacciu "Measuring the effects of cofounders in medical supervised classification problems: the confounding index (CI)", *Artificial Intelligence in Medicine*, Vol. 103, pp. 101804, 2020. https://doi.org/10.1016/j.artmed.2020.101804

[75] Z. Zenget al., "Deep learning for cancer type classification and driver gene identification", *BMC Bioinformatics*, Vol. 22, art. no. 491, 2021. https://doi.org/10.1186/s12859-021-04400-4

[76] Y. Zhang et al., "Genetic load and potential mutational meltdown in cancer cell populations", *Molecular Biology and Evolution*, Vol. 36, No. 3, pp. 541-552, 2019. https://doi.org/10.1093/molbev/msy231

[77] L. B. Alexandrov et al., "Signatures of mutational processes in human cancer", *Nature*, Vol. 500, pp. 415–421, 2013. https://doi.org/10.1038/nature12477

[78] D. Mathios et al., "Detection and characterization of lung cancer using cell-free DNA fragmentomes", *Nature Communications*, Vol. 12, pp. 5060, 2021. https://doi.org/10.1038/s41467-021-24994-w

[79] J. J. Chabon et al., "Integrating genomic features for non-invasive early lung cancer detection", *Nature*, Vol. 580, pp. 245–251, 2020. https://doi.org/10.1038/s41586-020-2140-0

[80] W. Liang et al., "Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA)", *Theranostics*, Vol. 9, No. 7, pp. 2056–2070, 2019. https://doi.org/10.7150/thno.28119

[81] L. Raman, M. van der Linden, K. van der Eecken, K. Vermaelen, I. Demedts, V. Surmont, et al., "Shallow whole-genome sequencing of plasma cell-free DNA accurately differentiates small from non-small cell lung carcinoma", *Genome Medicine*, Vol. 12, art. no. 35, 2020. https://doi.org/10.1186/s13073-020-00735-4

[82] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Molecular Biology of the Cel4th edition", *Annals botany*, Vol. 91, No. 3, pp. 401, 2002. https://doi.org/10.1093/aob/mcg023

[83] K. Kobayashi, A. Bolatkan, S. Shiina, and R. Hamamoto, "Fully-connected neural networks with reduced parameterization for predicting histological types of lung cancer from somatic mutations", *Biomolecules*, Vol. 10, pp. 1249, 2020. https://doi.org/10.3390/biom10091249

[84] M. S. Lawrence, P. Stojanov, C. H. Mermel, J. T. Robinson, L. A. Garraway, T. R. Golub, et al., "Discovery and saturation analysis of cancer genes across 21 tumor type", *Nature*, Vol. 505, pp. 495–501, 2014. https://doi.org/10.1038/nature12912

[85] M. Uhlen, C. Zhang, S. Lee, E. Sjostedt, L. Fagerberg, G. Bidkhori, et al., "A pathology atlas of the human cancer transcriptome", *Science*, Vol. 357, No. 6352, 2017. https://doi.org/10.1126/science.aan2507

[86] D. H. Whitney, M. R. Elashoff, K. Porta-Smith, A. C. Gower, A. Vachani, J. S. Ferguson, et al., "Derivation of a bronchial genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy," *BMC Medical Genomics*, Vol. 8, No. 18, 2015, https://doi.org/10.1186/s12920-015-0091-3

[87] M. D. Podolsky, A. A. Barchuk, V. I. Kuznetcov, N. F. Gusarova, V. S. Gaidukov, and S. A. Tarakanov, "Evaluation of machine learning algorithm utilization for lung cancer classification based on gene expression levels", *Asian Pacific journal of Cancer Prevention*, vol. 17, pp. 835–838, 2016. https://doi.org/10.7314/apjcp.2016.17.2.835

[88] Y. Choi, J. Qu, S. Wu, Y. Hao, J. Zhang, J. Ning, et al., "Improving lung cancer risk stratification leveraging whole transcriptome RNA sequencing and machine learning across multiple cohorts", *BMC Medical Genomics*, Vol. 13, No. 151, 2020. https://doi.org/10.1186/s12920-020-00782-1

[89] F. Alharbi, and A. Vakansi, "Machine learning methods for cancer classification using array gene expression data: A review", *Bioengineering*, Vol.10, No. 2, art. no. 173, 2023. https://doi.org/10.3390/bioengineering10020173

[90] X. Shao, N. Lv, J. Liao, J. B. Long, R. Xue, N. Ai, et al., "Copy number variation is highly correlated with differential gene expression: A pan-cancer study", *BMC Medical Genetics*, Vol. 20, No. 175, 2019. https://doi.org/10.1186/s12881-019-0909-5

[91] C. F. Aliferis, D. Hardin, and P. P. Massion, "Machine learning models for lung cancer classification using array comparative genomic hybridization", *AMIA Symposium*, pp. 7–11, 2002. https://pmc.ncbi.nlm.nih.gov/articles/PMC2244172/

[92] A. Daemen, O. Gevaert, K. Leunen, E. Legius, I. Vergote, and B. De Moor, "Supervised classification of array CGH data with HMM-based feature selection", *Pacific Symposium on Biocomputing*, pp. 468–479, 2009. https://doi.org/10.1142/9789812836939_0045

[93] P. Jurmeister, M. Bockmayr, P. Seegerer, T. Bockmayr, D. Treue, G. Montavon, et al., "Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases", *Science Translational Medicine*, Vol. 11, No. 509, 2019. https://doi.org/10.1126/scitranslmed.aaw8513

[94] Y. Luo, G. Riedlinger, and P. Szolovits, "Text mining in cancer gene and pathway prioritization", *Cancer Informatics*, Vol. 13, pp. 69–79, 2014. https://doi.org/10.4137/cin.s13874

[95] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines", *Machine Learning*, Vol. 46, pp. 389–422, 2002. https://doi.org/10.1023/A:1012487302797

[96] S. Mirhadi, S. Tam, Q. Li, N. Moghal, N. A. Pham, J. Tong, et al., "Integrative analysis of non-small cell lung cancer patient-derived xenografts identifies distinct proto types associated with patient outcomes", *Nature Communications*, Vol. 13, art. no. 1811, 2022. https://doi.org/10.1038/s41467-022-29444-9

[97] J. Y. Xu, C. Zhang, X. Wang, L. Zhai, Y. Ma, Y. Mao, et al., "Integrative proteomic characterization of human lung adenocarcinoma", *Cell*, Vol. 182, No. 1, pp. 245–261, 2020. https://doi.org/10.1016/j.cell.2020.05.043

[98] Y. Mardor, Y. Roth, A. Ocherashvilli, et al., "Pretreatment prediction of Brain Tumors Response to radiation therapy using high b-Value Diffusion-Weighted MRI1", *Neoplasia*, Vol. 6, No. 2, pp. 136-142, 2004. https://doi.org/10.1593/neo.03349

[99] Z. Zeng, A. Amin, A. Roy, N. E. Pulliam, L. C. Karavites, S. Espino, et al., "Preoperative magnetic resonance imaging use and oncologic outcomes in premenopausal breast cancer patients", *NPJ Breast Cancer*, Vol. 6, No. 49, 2020. https://doi.org/10.1038/s41523-020-00192-7

[100] Y. Chang, H. Park, H. J. Yang, S. Lee, K. Y. Lee, T. S. Kim, et al., "Cancer drug response profile scan (CDRscan): A deep learning model that

predicts drug effectiveness from cancer genomic signature", *Scientific Reports*, Vol. 8, No. 8857, 2018.
https://doi.org/10.1038/s41598-018-27214-6

[101] M. P. Menden, F. Iorio, M. Garnett, U. McDermott, C. H. Benes, P. J. Ballester, et al., "Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties", *PLoS ONE*, Vol. 8, art. no. 0061318, 2013.
https://doi.org/10.1371/journal.pone.0061318

[102] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, et al., "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)", *European Journal of Cancer*, Vol. 45, No. 2, pp. 228–247, 2009.
https://doi.org/10.1016/j.ejca.2008.10.026

[103] G. Adam, L. Rampasek, Z. Safikhani, P. Smirnov, B. Haibe-Kains, and A. Goldenberg, "Machine learning approaches to drug response prediction: Challenges and recent progress", *NPJ Precision Oncology*, Vol. 4, art. no. 19, 2020.
https://doi.org/10.1038/s41698-020-0122-1

[104] J. Jiang, Y. C. Hu, C. J. Liu, D. Halpenny, M. D. Hellmann, J. O. Deasy, et al., "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images", *IEEE Transactions on Medical Imaging*, Vol. 38, No. 1, pp. 134–144, 2019.
https://doi.org/10.1109/TMI.2018.2857800

[105] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3309–3318, 2017.
https://doi.org/10.1109/CVPR.2017.353

[106] R. Qureshi, "Personalized drug response prediction model for lung cancer patients using machine learning", *TechRxiv*, *preprint*, 2020.
https://doi.org/10.36227/techrxiv.13273319.v1

[107] A. Kapil, A. Meier, A. Zuraw, K. E. Steele, M. C. Rebelatto, G. Schmidt, et al., "Deep semi-supervised generative learning for automated tumor proportion scoring on NSCLC tissue needle biopsies", *Scientific Reports*, Vol. 8, art. no. 17343, 2018.
https://doi.org/10.1038/s41598-018-35501-5

[108] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions System, Man and Cybernetics*, Vol. 9, No. 1, pp. 62–66, 1979.
https://doi.org/10.1109/TSMC.1979.4310076

[109] P. Geeleher, N. J. Cox, and R. S. Huang, "Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines", *Genome Biology*, Vol. 15, No. R47, 2014.
https://doi.org/10.1186/gb-2014-15-3-r47

[110] E. Cule and M. De Iorio, "Ridge regression in prediction problems: Automatic choice of the ridge parameter", *Genetic Epidemiology*, Vol. 37, No. 7, pp. 704–714, 2013.
https://doi.org/10.1002/gepi.21750

[111] A. C. Quiros, N. Coudray, A. Yeaton, X. Yang, L. Chiriboga, A. Karimkhan, et al., "Self-supervised learning unveils morphological clusters behind lung cancer types and prognosis", *arXiv preprint arXiv: 2205.01931*, 2022.
https://doi.org/10.48550/arXiv.2205.01931

[112] M. F. Gensheimer, S. Aggarwal, K. R. K. Benson, J. N. Carter, A. S. Henry, D. J. Wood, et al., "Automated model versus treating physician for predicting survival time of patients with metastatic cancer", *Journal of the American Medical Informatics Association*, Vol. 28, No. 6, pp. 1108–1116, 2021.
https://doi.org/10.1093/jamia/ocaa290

[113] S. Doppalapudi, R. G. Qiu, and Y. Badr, "Lung cancer survival period prediction and understanding: Deep learning approaches", *International Journal of Medical Informatics*, Vol. 148, art. no. 104371, 2021.
https://doi.org/10.1016/j.ijmedinf.2020.104371

[114] M. Nair, S. S. Sandhu, and A. K. Sharma, "Prognostic and predictive biomarkers in cancer", *Current Cancer Drug Targets*, Vol. 14, pp. 477–504, 2014.
https://doi.org/10.2174/1568009614666140506111118

[115] E. Chow, L. Davis, T. Panzarella, C. Hayter, E. Szumacher, A. Loblaw, et al., "Accuracy of survival prediction by palliative radiation oncologists", *International Journal Radiation Oncololgy Biololgy Physics*, Vol. 61, No. 3, pp. 870–873, 2005.

https://doi.org/10.1016/j.ijrobp.2004.07.697

[116] J. R. Lakin, M. G. Robinson, R. E. Bernacki, B. W. Powers, S. D. Block, R. Cunningham, et al., "Estimating 1-year mortality for high-risk primary care patients using the 'surprise' question", *JAMA Internal Medicine*, Vol. 176, No. 12, pp. 1863–1865, 2016. https://doi.org/10.1001/jamainternmed.2016.5928

[117] N. White, F. Reid, A. Harris, P. Harries, and P. Stone, "A systematic review of predictions of survival in palliative care: How accurate are clinicians and who are the experts?", *PLoS ONE*, Vol. 11, art. no. 0161407, 2016. https://doi.org/10.1371/journal.pone.0161407

[118] D. R. Cox, "Regression models and life tables", *Journal of the Royol Statistical Society Series B (Methodological)*, Vol. 34, pp. 187–202, 2018. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[119] X. Wang, S. Yao, Z. Xiao, J. Gong, Z. Liu, B. Han, et al., "Development and validation of a survival model for lung adenocarcinoma based on autophagy-associated genes", *Journal of Translational Medicine*, Vol. 18, art. no. 149, 2020. https://doi.org/10.1186/s12967-020-02321-z

[120] Y. H. Zhang, Y. Lu, H. Lu, and Y. M. Zhou, "Development of a survival prognostic model for non-small cell lung cancer", *Frontiers in Oncology*, Vol. 10, art. no. 362, 2020. https://doi.org/10.3389/fonc.2020.00362

[121] K. H. Yu, C. Zhang, G. J. Berry, R. B. Altman, R. E. Re, D. L. Rubin, et al., "Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features", *Nature Communications*, Vol. 7, art. no. 12474, 2016. https://doi.org/10.1038/ncomms12474

[122] P. Hatlen, B. H. Gronberg, A. Langhammer, S. M. Carlsen, and T. Amundsen, "Prolonged survival in patients with lung cancer with diabetes mellitus", *Journal of Thoracic Oncology*, Vol. 6, No. 11, pp. 1810–1817, 2011. https://doi.org/10.1097/jto.0b013e31822a75be

[123] Y. C. Chen, W. C. Ke, and H. W. Chiu, "Risk classification of cancer survival using ANN with gene expression data from multiple laboratories", *Computers in Biology and Medicine*, Vol. 48, No. 1, pp. 1–7, 2014. https://doi.org/10.1016/j.compbiomed.2014.02.006

[124] Y. Liu, M. Yang, W. Sun, M. Zhang, J. Sun, W. Wang, et al., "Developing prognostic gene panel of survival time in lung adenocarcinoma patients using machine learning", *Translational Cancer Research*, Vol. 9, No. 6, pp. 3860–3869, 2020. https://doi.org/10.21037/tcr-19-2739

[125] J. Yu, Y. Hu, Y. Xu, J. Wang, J. Kuang, W. Zhang, et al., "LUADpp: An effective prediction model on prognosis of lung adenocarcinomas based on somatic mutational features", *BMC Cancer*, Vol. 19, art. no. 263, 2019. https://doi.org/10.1186/S12885-019-5433-7

[126] H. J. Cho, S. Lee, Y. G. Ji, and D. H. Lee, "Association of specific gene mutations derived from machine learning with survival in lung adenocarcinoma", *PLoS ONE*, Vol. 13, No. 11, art. no. e0207204, 2018. https://doi.org/10.1371/journal.pone.0207204

[127] X. Hui, Y. Hu, M. A. Sun, X. Shu, R. Han, Q. Ge, et al., "EBT: A statistic test identifying a moderate size of significant features with balanced power and precision for genome-wide rate comparisons", *Bioinformatics*, Vol. 33, No. 17, pp. 2631–2641, 2017. https://doi.org/10.1093/bioinformatics/btx294

[128] K. H. Yu, G. J. Berry, D. L. Rubin, R. E. Re, R. B. Altman, and M. Snyder, "Association of omics features with histopathology patterns in lung adenocarcinoma", *Cell Systems*, Vol. 5, No. 6, pp. 620–627, 2017. https://doi.org/10.1016/j.cels.2017.10.014

[129] D. Ramazzotti, A. Lal, B. Wang, S. Batzoglou, and A. Sidow, "Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival", *Nature Communications*, Vol. 9, art. no. 4453, 2018. https://doi.org/10.1038/s41467-018-06921-8

[130] K. Asada, K. Kobayashi, S. Joutard, M. Tubaki, S. Takahashi, K. Takasawa, et al., "Uncovering prognosis-related genes and pathways by multi-omics analysis in lung cancer", *Biomolecules*, Vol. 10, No. 4, art. no. 524, 2020. https://doi.org/10.3390/biom10040524

[131] S. Takahashi, K. Asada, K. Takasawa, R. Shimoyama, A. Sakai, A. Bolatkan, et al., "Predicting deep learning-based multi-omics

parallel integration survival subtypes in lung cancer using reverse phase protein array data", *Biomolecules*, Vol. 10, No. 10, art. no. 1460, 2020. https://doi.org/10.3390/biom10101460

[132] L. Xia, Y. Liu, and Y. Wang, "PD-1/PD-L1 blockade therapy in advanced non-small-cell lung cancer: Current status and future directions", *The Oncologist*, Vol. 24, No. S1, pp. S31–S41, 2019. https://doi.org/10.1634/theoncologist.2019-io-s1-s05

[133] D. B. Doroshow, M. F. Sanmamed, K. Hastings, K. Politi, D. L. Rimm, L. Chen, et al., "Immunotherapy in non-small cell lung cancer: Facts and hopes", *Clinical Cancer Research*, Vol. 25, No. 15, pp. 4592–4602, 2019. https://doi.org/10.1158/1078-0432.ccr-18-1538

[134] S. M. Lim, M. H. Hong, and H. R. Kim, "Immunotherapy for non-small cell lung cancer: Current landscape and future perspectives", *Immune Network*, Vol. 20, No. 1, art. no. e10, 2020. https://doi.org/10.4110/in.2020.20.e10

[135] Z. Xu, X. Wang, S. Zeng, X. Ren, Y. Yan, and Z. Gong, "Applying artificial intelligence for cancer immunotherapy", *Acta Pharmaceutica Sinica B*, Vol. 11, No. 11, pp. 3393–3405, 2021. https://doi.org/10.1016/j.apsb.2021.02.007

[136] M. Wiesweg, F. Mairinger, H. Reis, M. Goetz, J. Kollmeier, D. Misch, et al., "Machine learning reveals a PD-L1-independent prediction of response to immunotherapy of non-small cell lung cancer by gene expression context", *European Journal of Cancer*, Vol. 140, pp. 76–85, 2020. https://doi.org/10.1016/j.ejca.2020.09.015

[137] S. Trebeschi, S. G. Drago, N. J. Birkbak, I. Kurilova, A. M. Calin, A. D. Pizzi, et al., "Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers", *Annals of Oncology*, Vol. 30, No. 6, pp. 998–1004, 2019. https://doi.org/10.1093/annonc/mdz108

[138] T. P. Coroller, V. Agrawal, V. Narayan, Y. Hou, P. Grossmann, S. W. Lee, et al., "Radiomic phenotype features predict pathological response in non-small cell lung cancer", *Radiotherapy Oncology*, Vol. 119, No. 3, pp. 480–486, 2016. https://doi.org/10.1016/j.radonc.2016.04.004

[139] M. Tosolini, F. Pont, M. Poupot, F. Vergez, M. L. Nicolau-Travers, D. Vermijlen, et al., "Assessment of tumor-infiltrating TCR Vγ9Vδ2 T lymphocyte abundance by deconvolution of human cancers microarrays", *Oncoimmunology*, Vol. 6, No. 3, art. no. e1284723, 2017. https://doi.org/10.1080/2162402X.2017.1284723

[140] A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, et al., "Robust enumeration of cell subsets from tissue expression profiles", *Nature Methods*, Vol. 12, pp. 453-457, 2015. https://doi.org/10.1038/nmeth.3337

[141] R. Sun, E. J. Limkin, M. Vakalopoulou, L. Dercle, S. Champiat, S. R. Han, et al., "To assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: An imaging biomarker, retrospective multicohort study", *The Lancet Oncology*, Vol. 19, No. 9, pp. 1180–1191, 2018. https://doi.org/10.1016/s1470-2045(18)30413-3

[142] J. Saltz, R. Gupta, L. Hou, T. Kurc, P. Singh, V. Nguyen, et al., "Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images", *Cell Reports*, Vol. 23, No. 1, pp. 181–193, 2018. https://doi.org/10.1016/j.celrep.2018.03.086

[143] L. Hou, V. Nguyen, A. B. Kanevsky, D. Samaras, T. M. Kurc, T. Zhao, et al., "Sparse auto encoder for unsupervised nucleus detection and representation in histopathology images", *Pattern Recognition*, Vol. 86, pp. 188–200, 2019. https://doi.org/10.1016/j.patcog.2018.09.007

[144] Y. Xu, Z. P. Jia, Y. Ai, F. Zhang, M. Lai, and E. I. C. Chang, "Deep convolutional activation features for large-scale brain tumor histopathology image classification and segmentation", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015)*, pp. 947–951, 2015. https://doi.org/10.1109/ICASSP.2015.7178109

[145] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation", *Computer Vision and pattern recognition*, pp. 1520–1528, 2015. https://doi.org/10.48550/arXiv.1505.04366

[146] L. D. M. Arruda, M. Vazquez, F. Finotello, R. Lepore, E. Porta, J. Hundal, et al., "Neoantigen prediction and computational perspectives towards clinical benefit: Recommendations from the ESMO precision medicine working group", *Annals of Oncology*, Vol. 31, No. 8, pp. 978–990, 2020.
https://doi.org/10.1016/j.annonc.2020.05.008

[147] V. Roudko, B. Greenbaum, and N. Bhardwaj, "Computational prediction and validation of tumor-associated neoantigens", *Frontiers in Immunology*, Vol. 11, art. no. 27, 2020.
https://doi.org/10.3389/fimmu.2020.00027

[148] Z. Zhang, M. Lu, Y. Qin, W. Gao, L. Tao, W. Su, et al., "Neoantigen: A new breakthrough in tumor immunotherapy", *Frontiers of Immunology*, Vol. 12, art. no. 672356, 2021.
https://doi.org/10.3389/fimmu.2021.672356

[149] N. Hilf, S. Kuttruff-Coqui, K. Frenzel, V. Bukur, S. Stevanovic, C. Gouttefangeas, et al., "Actively personalized vaccination trial for newly diagnosed glioblastoma", *Nature*, Vol. 565, pp. 240–245, 2019.
https://doi.org/10.1038/s41586-018-0810-y

[150] B. M. Carreno, V. Magrini, M. Becker-Hapak, S. Kaabinejadian, J. Hundal, A. A. Petti, et al., "A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen specific T cells", *Science*, Vol. 348, No. 6236, pp. 803–808, 2015.
https://doi.org/10.1126/science.aaa3828

[151] P. A. Ott, Z. Hu, D. B. Keskin, S. A. Shukla, J. Sun, D. J. Bozym, et al., "An immunogenic personal neoantigen vaccine for patients with melanoma", *Nature*, Vol. 547, pp. 217–221, 2017.
https://doi.org/10.1038/nature22991

[152] D. B. Keskin, A. J. Anandappa, J. Sun, I. Tirosh, N. D. Mathewson, S. Li, et al., "Neoantigen vaccine generates intratumoral T cell responses in phase Ibglioblastoma trial", *Nature*, Vol. 565, pp. 234–239, 2019.
https://doi.org/10.1038/s41586-018-0792-9

[153] W. Zhao and X. Sher, "Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes", *PLoS Computational Biology*, Vol. 14, No. 11, art. no. e1006457, 2018.
https://doi.org/10.1371/journal.pcbi.1006457

[154] J. Racle, J. Michaux, G. A. Rockinger, M. Arnaud, S. Bobisse, C. Chong, et al., "Robust prediction of HLA class II epitopes by deep motif deconvolution of immune peptidomes", *Nature Biotechnology*, Vol. 37, pp. 1283–1286, 2019.
https://doi.org/10.1038/s41587-019-0289-6

[155] B. B. Chen, M. S. Khodadoust, N. Olsson, L. E. Wagar, E. Fast, C. L. Liu, et al., "Predicting HLA class II antigen presentation through integrated deep learning", *Nature Biotechnology*, Vol. 37, pp. 1332–1343, 2019.
https://doi.org/10.1038/s41587-019-0280-2

[156] T. J. O'Donnell, A. Rubinsteyn, M. Bonsack, A. B. Riemer, U. Laserson, and J. Hammerbacher, "MHC flurry: Open-source class I MHC binding affinity prediction", *Cell Systems*, Vol. 7, No. 1, pp. 129–132, 2018.
https://doi.org/10.1016/j.cels.2018.05.014

[157] B. Bulik-Sullivan, J. Busby, C. D. Palmer, M. J. Davis, T. Murphy, A. Clark, et al., "Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification", *Nature Biotechnology*, Vol. 37, pp. 55–63, 2018.
https://doi.org/10.1038/nbt.4313

[158] C. Lundegaard, K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen, "NetMHC-3.0: Accurate web-accessible predictions of human, mouse, and monkey MHC class I affinities for peptides of length 8–11", *Nucleic Acids Research*, Vol. 36, pp. W509–W512, 2008.
https://doi.org/10.1093/nar/gkn202

[159] M. Nielsen, C. Lundegaard, P. Worning, S. L. Lauemoller, K. Lamberth, S. Buus, et al., "Reliable prediction of T-cell epitopes using neural networks with novel sequence representations", *Protein Science*, Vol. 12, No. 5, pp. 1007–1017, 2003.
https://doi.org/10.1110/ps.0239403

[160] S. Buus, A. Stryhn, K. Winther, N. Kirkby, and L. O. Pedersen, "Receptor-ligand interactions measured by an improved spun column chromatography technique: A high-efficiency and high-throughput size separation method", *Biochimica et Biophysica Acta*, Vol. 1243, No. 3, pp. 453–460, 1995.
https://doi.org/10.1016/0304-4165(94)00172-T

[161] V. Jurtz, S. Paul, M. Andreatta, P. Marcatili, B. Peters, and M. Nielsen, "NetMHCpan-4.0: Improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data", *Journal of Immunology*, Vol. 199, No. 9, pp. 3360–3368, 2017. https://doi.org/10.4049/jimmunol.1700893

[162] M. Nielsen, M. Andreatta "NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets", *PLoS One,* Vol. 8, art. no. 33, 2016. https://doi.org/10.1186/s13073-016-0288-x

[163] L. Ye, J. Creaney, A. Redwood, and B. Robinson, "The current lung cancer neoantigen landscape and implications for therapy", *Journal of Thoracic Oncology,* Vol. 16, No. 6, pp. 922–932, 2021. https://doi.org/10.1016/j.jtho.2021.01.1624

[164] L. Gong, R. He, Y. Xu, T. Luo, K. Jin, W. Yuan, et al., "Neoantigen load as a prognostic and predictive marker for stage II/III non-small cell lung cancer in Chinese patients", *Thoracic Cancer,* Vol. 12, No. 15, pp. 2170–2181, 2021. https://doi.org/10.1111/1759-7714.14046

[165] W. Zhang, Q. Yin, H. Huang, J. Lu, H. Qin, S. Chen, et al., "Personal neoantigens from patients with NSCLC induce efficient antitumor responses", *Frontiers in Oncology,* Vol. 11, p. 628456, 2021. https://doi.org/10.3389/fonc.2021.628456

[166] L. Zou, S. Yu, T. Meng, Z. Zhang, X. Liang, and Y. Xie, "A technical review of convolutional neural network-based mammographic breast cancer diagnosis," *Computational and Mathematical Methods in Medicine,* Vol. 2019, art. no. 6509357, 2019. https://doi.org/10.1155/2019/6509357

[167] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, Vol. 521, pp. 436 - 444, 2015. https://doi.org/10.1038/nature14539

[168] C. Mao, L. Yao, and Y. Luo, "ImageGCN: Multi-relational image graph convolutional networks for disease identification with chest X-rays," *IEEE Transactions on Medical Imaging,* Vol. 41, No. 8, pp. 1990–2003, 2022. https://doi.org/10.1109/TMI.2022.3153322

[169] C. Mao, L. Yao, Y. Pan, Z. Zeng, and Y. Luo, "Deep generative classifiers for thoracic disease diagnosis with chest X-ray images", *IEEE International Conference on Bioinformatics and Biomedicine (BIBM),* pp. 1209–1214, 2019. https://doi.org/10.1109/BIBM.2018.8621107

[170] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., "An image is worth 16×16 words: Transformers for image recognition at scale", *arXiv preprint arXiv:2010.11929,* 2020. https://doi.org/10.48550/arXiv.2010.11929

[171] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. Khan, and M. Shah, "Transformers in vision: A survey", *ACM Computing Surveys,* Vol. 54, No. 10s, pp. 1-41, 2022. https://doi.org/10.1145/3505244

[172] G. Boesch, "Comparing Vision transformers and convolutional neural networks for image classification: A Literature review", *Applied Sciences*, Vol. 13, No. 9, art. no. 5521, 2023. https://doi.org/10.3390/app13095521

[173] B. Hie, E. D. Zhong, B. Berger, and B. Bryson, "Learning the language of viral evolution and escape", *Science,* Vol. 371, No. 6526, pp. 284–288, 2021. https://doi.org/10.1126/science.abd7331

[174] L. Flagel, Y. Brandvain, and D. R. Schrider, "The unreasonable effectiveness of convolutional neural networks in population genetic inference", *Molecular Biology and Evolution,* Vol. 36, No. 2, pp. 220–238, 2019. https://doi.org/10.1093/molbev/msy224

[175] R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, and N. Yosef, "Deep generative modeling for single-cell transcriptomics", *Nature Methods,* Vol. 15, pp. 1053–1058, 2018. https://doi.org/10.1038/s41592-018-0229-2

[176] S. Hashim, M. Ali, K. Nandakumar, and M. Yaqub, "SubOmiEmbed: Self-supervised representation learning of multi-omics data for cancer type classification", *Proceedings of the 10th International Conference on Bioinformatics and Computational Biology (ICBCB),* pp. 66-72, 2022. https://doi.org/10.1109/ICBCB55259.2022.9802478

[177] J. Lee, D. Y. Hyeon, and D. Hwang, "Single-cell multiomics: Technologies and data analysis methods", *Experimental & Molecular Medicine*, Vol. 52, pp. 1428–1442, 2020. https://doi.org/10.1038/s12276-020-0420-2

[178] A. Rao, D. Barkley, G. S. Franca, and I. Yanai, "Exploring tissue architecture using spatial transcriptomics", *Nature,* Vol. 596, pp. 211–220, 2021. https://doi.org/10.1038/s41586-021-03634-9

[179] Y. Li, Q. Liu, Z. Zeng, and Y. Luo, "Using an unsupervised clustering model to detect the early spread of SARS-CoV-2 worldwide", *Genes,* Vol. 13, No. 4, art. no. 648, 2022. https://doi.org/10.3390/genes13040648

[180] T. Tian, J. Wan, Q. Song, and Z. Wei, "Clustering single-cell RNA-seq data with a model-based deep learning approach", *Nature Machine Intelligence,* Vol. 1, pp. 191–198, 2019. https://doi.org/10.1038/s42256-019-0037-0

[181] J. Hu, X. Li, G. Hu, Y. Lyu, K. Susztak, and M. Li, "Iterative transfer learning with neural network for clustering and cell type classification in single-cell RNA-seq analysis", *Nature Machine Intelligence,* Vol. 2, pp. 607–618, 2020. https://doi.org/10.1038/s42256-020-00233-7

[182] M. Brbic, M. Zitnik, S. Wang, A. O. Pisco, R. B. Altman, S. Darmanis, et al., "MARS: Discovering novel cell types across heterogeneous single-cell experiments", *Nature Methods,* Vol. 17, pp. 1200–1206, 2020. https://doi.org/10.1038/s41592-020-00979-3

[183] H. Shen, Y. Li, M. Feng, X. Shen, D. Wu, C. Zhang, et al., "Miscell: An efficient self-supervised learning approach for dissecting single cell transcriptome", *iScience,* Vol. 24, No. 11, art. no. 103200, 2021. https://doi.org/10.1016/j.isci.2021.103200

[184] Q. Song, J. Su, and W. Zhang, "scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics", *Nature Communications,* Vol. 12, art. no. 3826, 2021. https://doi.org/10.1038/s41467-021-24172-y

[185] T. Wang, W. Shao, Z. Huang, H. Tang, J. Zhang, Z. Ding, et al., "Mogonet integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification," *Nature Communications,* Vol. 12, art. no. 3445, 2021. https://doi.org/10.1038/s41467-021-23774-w

[186] Y. Wang, Z. Zhang, H. Chai, and Y. Yang, "Multi-omics cancer prognosis analysis based on graph convolution network", *IEEE International Conference on Bioinformatics and Biomedicine (BIBM),* pp. 1564–1568, 2021. https://doi.org/10.1109/BIBM52615.2021.9669797

[187] Y. D. Zhang, Z. Dong, S. H. Wang, X. Yu, X. Yao, Q. Zhou, et al., "Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation", *Information Fusion,* Vol. 64, pp. 149–187, 2020. https://doi.org/10.1016/j.inffus.2020.07.006

[188] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, et al., "Comprehensive integration of single-cell data", *Cell,* Vol. 177, No. 7, pp. 1888–1902, 2019. https://doi.org/10.1016/j.cell.2019.05.031

[189] I. Subramanian, S. Verma, S. Kumar, A. Jere, and K. Anamika, "Multi-omics data integration, interpretation, and its application", *Bioinformatics and Biology Insights,* Vol. 14, pp. 1-24, 2020. https://doi.org/10.1177/1177932219899051

[190] J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, and E. Z. Macosko, "Single-cell multi-omic integration compares and contrasts features of brain cell identity", *Cell,* Vol. 177, pp. 1873–1887, 2019. https://doi.org/10.1016/j.cell.2019.05.006

[191] Y. Luo, A. Eran, N. Palmer, P. Avillach, A. Levy-Moonshine, P. Szolovits, et al., "A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia," *Nature Medicine,* Vol. 26, pp. 1375–1379, 2020. https://doi.org/10.1038/s41591-020-1007-0

[192] L. Diao, H. Guo, Y. Zhou, and Y. He, "Bridging the gap between outputs: Domain adaptation for lung cancer IHC segmentation," *IEEE International Conference on Image Processing,* pp. 6-10, 2021. https://doi.org/10.1109/ICIP42928.2021.9506115

[193] M. Lotfollahi, M. Naghipourfar, M. D. Luecken, M. Khajavi, M. Buttner, M. Wagenstetter, et al., "Mapping single-cell data to reference atlases by transfer learning," *Nature Biotechnology*, Vol. 40, pp. 121–130, 2022. https://www.nature.com/articles/s41587-021-01001-7

[194] A. B. Arrieta, N. Diaz-Rodriguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, et al., "Explainable artificial intelligence (XAI):

Concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, Vol. 58, pp. 82–115, 2020. https://doi.org/10.1016/j.inffus.2019.12.012

[195] Z. Salahuddin, H. C. Woodruff, A. Chatterjee, and P. Lambin, "Transparency of deep neural networks for medical image analysis: A review of interpretability methods", *Computers in Biology and Medicine*, Vol. 140, art. no. 105111, 2022. https://doi.org/10.1016/j.compbiomed.2021.105111

[196] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, et al., "Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review", *Applied Sciences*, Vol. 11, No. 11, pp. 5088, 2021. https://doi.org/10.3390/app11115088

[197] K. Kourou, K. P. Exarchos, C. Papaloukas, P. Sakaloglou, T. Exarchos, and D. I. Fotiadis, "Applied machine learning in cancer research: A systematic review for patient diagnosis, classification, and prognosis", *Computational and Structural Biotechnology Journal*, Vol. 19, pp. 5546–5555, 2021. https://doi.org/10.1016/j.csbj.2021.10.006

[198] L. Maier-Hein, A. Reinke, P. Godau, M. D. Tizabi, E. Christodoulou, B. Glocker, et al., "Metrics reloaded: recommendations for image analysis validation", *Nature Methods*, Vol. 21, pp. 195-212, 2024. https://doi.org/10.1038/s41592-023-02151-z

[199] N. J. Meropol, J. Donegan, and A. S. Rich, "Progress in the application of machine learning algorithms to cancer research and care", *JAMA Network Open*, Vol. 4, No. 7, art. no. e2116063, 2021. https://doi.org/10.1001/jamanetworkopen.2021.16063

[200] D. F. Sheehan, S. D. Criss, Y. F. Chen, A. Eckel, L. Palazzo, A. C. Tramontano, et al., "Lung cancer costs by treatment strategy and phase of care among patients enrolled in Medicare", *Cancer Medicine*, Vol. 8, No. 1, pp. 94–103, 2018. https://doi.org/10.1002/cam4.1896

[201] C. Mao, L. Yao, and Y. Luo, "MedGCN: Medication recommendation and lab test imputation via graph convolutional networks", *Journal of Biomedical Informatics*, Vol. 127, art. no. 104000, 2022. https://doi.org/10.1016/j.jbi.2022.104000

[202] J. Zhu, J. Wang, X. Wang, M. Gao, B. Guo, M. Gao, et al., "Prediction of drug efficacy from transcriptional profiles with deep learning", *Nature Biotechnology*, Vol. 39, pp. 1444–1452, 2021. https://doi.org/10.1038/s41587-021-00946-z

[203] Y. Luo, R. G. Wunderink, and D. Lloyd-Jones, "Proactive vs reactive machine learning in healthcare: Lessons from the COVID-19 pandemic", *JAMA*, Vol. 327, No. 7, pp. 623–624, 2022. https://doi.org/10.1001/jama.2021.24935

[204] H. S. Gao, C. X. Wang, H. B. Luo, S. S. Chen, S. Y. Dong, Y. L. et al., "Concordance of treatment recommendations for metastatic non-small-cell lung cancer between Watson for Oncology system and medical team", *Cancer Management and Research*, Vol. 12, pp. 1947–1958, 2020. https://doi.org/10.2147/cmar.s244932

[205] C. Liu, X. Liu, F. Wu, M. Xie, Y. Feng, and C. Hu, "Using artificial intelligence (Watson for Oncology) for treatment recommendations amongst Chinese patients with lung cancer: Feasibility study", *Journal of Medical Internet Research*, Vol. 20, No. 9, art. no. e11087, 2018. https://doi.org/10.2196/11087

***