Annals of Research in Oncology
Vol. 5(3), 144-154, 2025

SHORT REPORT

LIQUID BIOPSY AND DIAGNOSTIC OF IMAGING (LIDIA) FOR THE DEFINITION OF PROGNOSTIC BIOMARKERS AND PERSONALISED THERAPIES IN LUNG CANCER: A CLINICAL TRIAL

Maria Paola **Belfiore** [1], Michele **Caraglia** [1], Carminia **Della Corte** [1], Floriana **Morgillo** [1], Gaetano **Di Guida** [1], Giovanni **Ciani** [1], Carlotta **Genco** [1], Gianluca **Russo** [2], Mario **Sansone** [3], Umberto **Malapelle** [2], Carmela **Nappi** [4], Salvatore **Cappabianca** [1, *]

[1] Department of Precision Medicine, University of Campania “Luigi Vanvitelli”, Naples, Italy
[2] Department of Public Health, University of Naples Federico II, Naples, Italy
[3] Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy
[4] Department of Advanced Biomedical Sciences, University of Naples Federico II, Naples, Italy
* Correspondence to: salvatore.cappabianca@unicampania.it

ABSTRACT: in recent years, imaging techniques have been successfully used to deliver diagnostic biomarkers with even greater accuracy. In particular, radiomic analysis methods (application of artificial intelligence on radiological images), which describe a segmented tumor region, using various quantitative characteristics derived from radiological images, have shown great potential in the identification, characterisation/classification of different types of cancer and in evaluating the response to radiotherapy and chemotherapy. Liquid biopsy is used for both early screening of malignancy and diagnosing minimal residual disease. It is also performed to assess and monitor the response to pharmacological treatments for a personalised therapeutic strategy. The analysis of morphostructural data obtained by imaging, correlated with the genetic/molecular results of liquid biopsy, could provide useful predictive factors for early diagnosis and predicting the response to anti-cancer drugs. The study aims to design and develop a report structured in CT with contrast media, which includes, in addition to the subjective evaluation of the radiologist, a quantitative/objective assessment of lung cancer (LC) with features that describe the texture and morphology of the lesion. Therefore, we present a workflow aimed at extracting the DICOM images acquired with CT using contrast medium from a significant number of patients, and to evaluate their accuracy in characterising the LC lesions. Furthermore, these data will be correlated to gene mutations and epigenetic changes (DNA methylation) evaluated in circulating tumour DNA derived from peripheral blood with a liquid biopsy approach. The correlation between radiomic characteristics, quantitative analysis of tumours performed by CT, structured lesion reports, and liquid biopsies could help avoid many unnecessary biopsy procedures and enable personalised treatment of LC patients.

Doi: 10.48286/aro.2025.112

Impact statement: An integrated radiomics–liquid biopsy model is proposed to support non-invasive molecular assessment and guide personalised therapy in patients with stage III lung cancer.

Key words: *radiomic; liquid biopsy; texture analysis; stage III lung cancer; biomarkers.*

Received: May 29, 2025/**Accepted:** Sept 25, 2025

Published: Oct 14, 2025

BACKGROUND

Lung cancer (LC) represents a leading cause of death worldwide, contributing to a high percentage of cancer-related deaths in both sexes (1). However, over the last decade, novel targeted therapies and immunotherapies have been developed due to the discovery of different mutations and aberrations in driver oncogenes (2, 3). Based on this, molecular testing and clinical biomarkers are now routinely used in clinical practice for the management of advanced LC, including the search for activating mutations

© 2025 Annals of Research in Oncology - ARO. Published by EDRA SpA. All rights reserved.

of epidermal growth factor receptor (EGFR), BRAF, HER2, MET, ERBB2 and KRAS, rearrangements of anaplastic lymphoma kinase (*ALK*) and ROS proto-oncogene1 (*ROS1*) as well as fusions of *NTRK1-3* genes (4, 5). Unfortunately, a large proportion of patients have an advanced disease, especially stage III, at diagnosis, thus excluding tumor resection opportunities (6, 7). In these cases, the available tumour tissue that can be used for molecular testing is often limited, being represented by small needle core biopsies or cytology specimens due to the risks associated with biopsy procedures involving the lung. However, the collection of sufficient material for diagnosis, subtyping, and characterisation is mandatory, and when not available, repeating the diagnostic procedures can lead to delayed decision-making in creating an algorithm. This can lead to detrimental effects on the clinical outcome of the patients, especially in stage I, II LCs in which the molecular diagnosis is essential for treatment decisions. In fact, in stage III LC, concurrent chemoradiotherapy (CRT) associated with the immunotherapy represents the preferred treatment, being proven to increase the overall survival of the patients if compared to radiotherapy alone. The rapidity of the best therapy choice is essential in these cases, considering that the optimal and personalised treatment of stage III LC patients is critical for achieving disease downstaging, which allows for subsequent surgical resection. The diagnostic delay can lead to a loss of the therapeutic window and subsequent opportunities for neoadjuvant treatment.Therefore, the availability of a diagnostic technique that enables a rapid definition of the diagnostic workflow is essential for informed treatment decision-making. Methodologies based upon liquid biopsy fulfil this role by allowing a wide range of molecular assessments through a minimally invasive procedure (8, 9). Different body fluids can be used for liquid biopsy, including saliva, cerebrospinal fluid, and, more often, peripheral blood. The latter is collected to obtain intact circulating tumour cells or their products, including circulating cell-free DNA (cfDNA), circulating tumour DNA (ctDNA), circulating miRNA, exosomes, extracellular vesicles, and others (10, 11). These products could subsequently be used in diagnosis, prediction of response, monitoring of treatment, and assessment of mutational status before and during the various treatments to which patients are subjected. Therefore, liquid biopsy is highly attractive for assessing both the tumour biology and molecular status of LC, both at single or multiple time points (*e.g.*, at diagnosis or relapse). For instance, it is established that LC demonstrates genomic instability with the progressive acquisition of genetic alterations (including point mutations, chromosomal instability and epigenetic alterations), although at varying rates, resulting in the development of genetic changes during the clinical evolution of the disease, as well as due to the effects of the different treatments. In stage III LC, these alterations can contribute to clonal evolution and resistance development, emphasizing the need for the

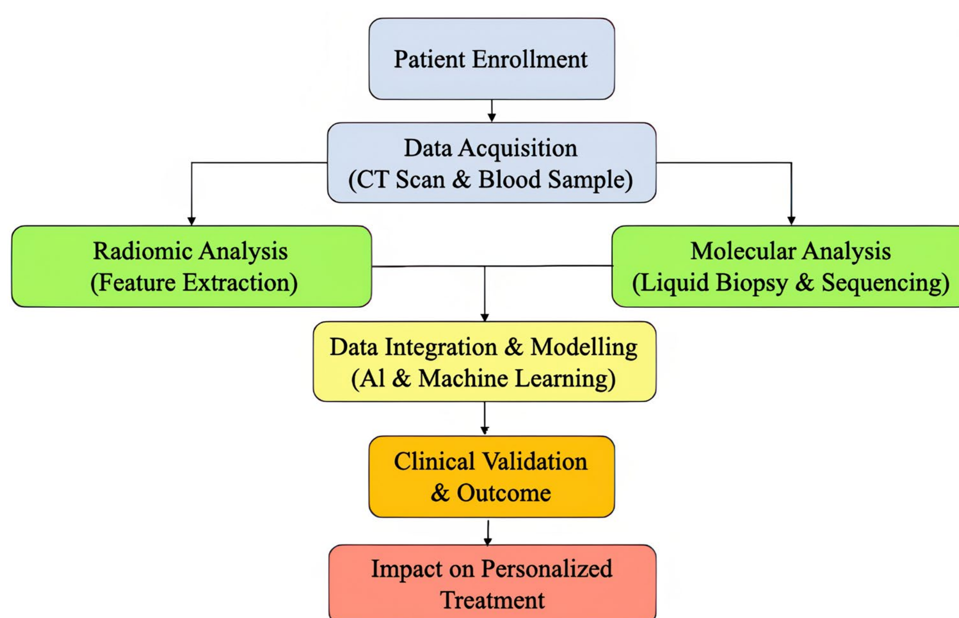

Figure 1. *Workflow of the LIDIA project, from patient enrollment to data analysis, AI modelling, clinical validation, and personalized treatment.*

continuous tracking of LC molecular profile during treatment. The selection of resistant clones during therapy represents a significant mechanism for the development of treatment resistance and disease progression. Moreover, given the increasing role of targeted therapy, the monitoring of the molecular profile of the disease is of paramount importance to identify resistance mechanisms. Therefore, all testing strategies that offer a safe modality for assessing LC biology are likely to be of significant clinical interest. An example is provided by the observation that approximately 15% of LC, particularly advanced non-small cell lung cancers (NSCLCs), display activating EGFR mutations, which can be targeted by tyrosine kinase inhibitors (TKIs). Liquid biopsy can be used not only to determine the presence of activating EGFR mutations in treatment-naïve patients for whom the tissue sample is insufficient or inadequate for molecular analysis, but also in patients who, after disease progression to first- or second-generation TKIs, develop resistance mechanisms.

The identification of EGFR mutations is also crucial for the management of early-stage NSCLC patients who may be candidates for adjuvant therapy with the latest-generation TKI, Osimertinib. Moreover, the potential incorporation of liquid biopsy techniques into screening algorithms, both for routine population screening and for therapy monitoring, represents an extremely attractive approach and an area of active investigation with promising early results. Liquid biopsy has definite and clinically relevant applications for the management of LC, particularly in stage III and other advanced stages, as well as in early-stage disease; however, its use is limited by cost, technical challenges, and availability. Therefore, although it is highly predictable that liquid biopsy will play a significant role in diagnosis, response assessment, and ongoing surveillance in the future, the available data are still inconclusive.

Liquid biopsy techniques offer an excellent combination of convenience and safety for molecular profiling, reducing the need for invasive and technically complicated tissue sampling. This information can be combined with data from imaging of tumour lesions to improve the diagnostic definition of the disease, allowing for molecular subtyping and predicting response to therapies.

Another critical challenge of the present study is determining the epigenetic alteration of cfDNA based on its methylation profile. Epigenetic modifications are considered a hallmark of cancer and are found in early stages of disease, tumour progression, and metastasis formation. DNA methylation is a tissue- and cancer-specific modification and, in contrast to the heterogeneity of gene mutations, appears to be similar in cancer cells of the same type and tissue origin (12, 13). Genome-wide methylation analysis using the bisulfite conversion method of cfDNA has been previously employed for cancer diagnosis (14). However, this method is expensive, time-consuming, and requires large amounts ofcfDNA. An innovative and highly sensitive alternative is offered by using cell-free methylated DNA immunoprecipitation with anti-5mC antibodies and subsequent high-throughput sequencing (cfMeDIP-seq) (15) to assess the methylation profile, even with low cfDNA input. Differentially methylated regions (DMRs) have been used to construct classifiers that can identify patients with several cancers (15, 16). Therefore, one of the objectives of the present study will be to use cfMeDip for the early diagnosis, determination of minimal residue disease, and histological subtyping of patients with LC, and to correlate these results with radiological imaging.

Based on these advances, this study aims to evaluate the diagnostic accuracy of chest CT in the morpho-structural characterisation of stage III LC. By extracting radiomic capabilities related to the structure and morphology of the lesions, the observation aims to correlate this information with the results of genetic, epigenetic, and molecular analyses obtained through liquid biopsy.

METHODS/DESIGN

The concept of a single-site biopsy to monitor disease dynamics during therapy is practically unfeasible, as it is invasive and may result in an underestimation of heterogeneity. On the other hand, a liquid biopsy based on the analysis of circulating tumour cells or tumour macromolecular products reflects the mutational status of the overall disease sites, allowing for the identification of emerging subclones responsible for treatment resistance. Additionally, radiomics has emerged as a novel field of research dealing with the extraction and analysis of specific features from diagnostic images, potentially reflecting the pathophysiological processes and the heterogeneity of tumour genetics.

The combined approach of radiomics and liquid biopsy has the potential to elucidate the dynamics of molecular lesions, thereby supporting informed clinical decision-making (17, 18).

Table 1. *Overview of Study Phases and Methodologies.*

PHASE	DESCRIPTION	TECHNIQUES / TOOLS
Data Collection	Collection of clinical data, imaging, and biopsy samples	Clinical Records, CT Imaging, Liquid Biopsy
Radiomic Analysis	Extraction of radiomic features from imaging data	Pyradiomics, ITK-SNAP
Genetic Analysis	Study of genetic mutations through liquid biopsy	PCR, Sequencing
Prediction of Outcomes	Combination of data to predict therapeutic outcomes	Machine Learning Models, AUC

Aims and objectives

The project aims to create a structured report in CT with contrast media, which includes, in addition to the subjective evaluation of the radiologist, a quantitative/objective assessment of the lung tumour with several features that describe the texture and morphology of the lesion (19, 20).

Therefore, as shown in **Figure 1**, the project aims to extract the DICOM images acquired with CT with contrast medium, from a significant number of patients, with a particular focus on stage III LC, for the subsequent evaluation of their accuracy in the characterisation of LC malignancy. Furthermore, in the same patients, the molecular analysis of the genes involved in LC will be performed on DNA extracted from a peripheral blood sample (21, 22).

The correlation between radiomic characteristics, quantitative analysis of tumours performed by CT, structured lesion reports, and liquid biopsies could help avoid many unnecessary biopsy procedures (23, 24).

Study design

This section outlines the techniques and protocols of our experimental study, aimed at integrating scientific, molecular, and imaging records to assess the diagnostic accuracy of CT scans and liquid biopsy in LC management, with a selected emphasis on stage III cases.

Inclusion criteria:

- Age ≥18 years
- Full understanding of the study and signed informed consent
- Presence of a neoplasm requiring further diagnostic evaluation
- Availability to undergo liquid biopsy.

Exclusion criteria:

- Allergies to contrast media
- Inability to maintain immobility during the exam
- Pregnancy or breastfeeding
- Risk factors for contrast nephropathy (GFR <60 ml/dl)
- Known allergy to contrast agent.

Recruitment Process

In the first year, from the first bimonthly period to the sixth, the two Diagnostic Imaging Units will be responsible for enrolling hospitalised patients who undergo CT-guided biopsy for suspected lung cancer. The CT investigation will be performed before histopathological sampling, to obtain information regarding morpho-densitometric characteristics of the lesion and to plan the subsequent biopsy procedure (25-28).

Prior to treatment administration and molecular pathology assessments, all patients provided written informed consent. The study was approved by the Ethics Committee "Comitato Etico Università degli Studi della Campania Luigi Vanvitelli" (approval No. 24997/2020) on 11th November 2020.

The recruitment and collection phase of clinical anamnestic data will be performed in a specific DICOM file (structured report) and will start after Informed Consent has been signed by the patient. Informed Consent will be accurately prepared for this study by the PI and substitute PI.

Imaging Acquisition and Analysis

Once the CT imaging has been acquired, the DICOM images will be evaluated by the PI and the Deputy PI from the first to sixth bimonthly period of the first year. From the second to the sixth bimonthly period of the first year a quantitative analysis of the lung lesion will be performed with an artificial intelligence system capable of identifying the tumor on the CT image, calculating its diameters and volume in a semi-automatic way.

Subsequently, the radiologists assisted by the engineer of the second research unit, will export the CT images. This phase will take place from the third

to the sixth bimonthly period of the first year. The images exported by the individual research units will be archived using a 'GDPR compliant' Cloud system that will be developed ad hoc for the study.
The cloud platform comprises:

- Storage Section: Secure archiving of DICOM files, structured reports, and genetic data.
- Computing Section: An online application for structured reporting, integrating clinical and anamnestic information, ensuring standardized procedures and simplifying radiomics/radiogenomics analysis.

Following the data extraction, during the period from the fifth to the fourth two-month period of the first year to the second year, the Department of Electrical Engineering and Information Technologies at the University of Federico II will carry out the computational analysis of the tumour volume to extract the radiomics features. Subsequently, from the sixth bimonthly period of the first year to the fourth bimonthly period of the second year, the same Department will carry out the classification of radiomic features with Machine Learning techniques. Finally, from the sixth bimonthly period of the first year to the third bimonthly period of the second year, these data will be processed and analysed to predict tumour characteristics.
Peripheral venous blood samples can be gathered throughout imaging acquisition to evaluate liquid biopsy molecular data. Plasma samples will be stored in two laboratories to maintain ctDNA integrity:

- Molecular and Precision Oncology Laboratory (Vanvitelli University and Biogem scarl)
- Cytology and Predictive Molecular Pathology Laboratory (Federico II University).

Furthermore, from the fifth two-month period of the first year to the third two-month period of the second year, the extraction of the ctDNA and the preparation of the genetic library will be performed by using OncomineTM Lung ctDNA Assay (Thermofisher, Massachusetts, USA). Afterwards, from the first two-month period of the second year to the fourth two-month period of the second year, the sequencing by Next Generation Sequencing (NGS) technique on the Ion Torrent GeneStudio S5Plus system (Thermofisher, Massachusetts, USA) will be run. Regarding data analysis, NGS technology involves various processes, which are very expensive from the point of view of the computational resources used. Gene sequencing of cfDNA samples using the NGS technique will be analysed on ThermoFisher systems and software.
The analysis of the characteristic driver mutations of lung cancer, as included in the OncomineTM Lung ctDNA Assay, provides sequencing of 11 genes (*ALK*, *BRAF*, *EGFR*, *ERBB2*, *KRAS*, *MAP2K1*, *MET*, *NRAS*, *PIK3CA*, *ROS1*, and *TP53*) and more than 150 hotspots. The analysis has a high specificity and sensitivity, along with an efficient workflow that enables the rapid generation of results. In addition to being inclusive of clinical-laboratory information, patient data will also contain information obtained from genetic analysis and will always be archived within the same cloud platform created *ad ho*c by the Department of Biomedical Engineering of Federico II.
Subsequently, from the third two-month period of the first year to the sixth two-month period of the second year, the two Research Units will undertake to correlate the data obtained from radiomic analysis with the data obtained from genetic analysis.
Finally, from the sixth two-month period of the first year to the sixth two-month period of the second year, the Engineering department of the second Unit will carry out the Radiomic analysis of the segmented volume using Imaging with the aim of obtaining a number of significant features that can be correlated with the genetic data of the liquid biopsy.

Techniques for the Analysis of Liquid Biopsy

A liquid biopsy will be performed only for patients who have previously undergone a CT study for diagnosis and staging. All patients enrolled in the study will undergo a peripheral venous blood sample collection in two test K2 tubes.
The ctDNA will be extracted from the plasma for molecular analysis, which will be performed using NGS technology, based on Ion Torrent technology. Unlike other fluorescence-based platforms, Ion Torrent uses an electrochemical approach to detect nucleotides, eliminating the need for optical labels and thereby increasing sequencing speed and accessibility. After genomic library preparation, DNA molecules are fragmented and ligated to oligonucleotide adapters, allowing immobilization onto specific beads. Each bead is then placed into an oil droplet containing emulsion PCR (emPCR) reagents, ensuring that each bead carries a single amplified DNA molecule. After amplification, the beads are loaded into a semiconductor chip, with each well containing a single bead with multiple copies of the same DNA fragment. Sequencing occurs through the sequential introduction of nucleotides. When a

complementary nucleotide is incorporated by DNA polymerase, a proton (H+) is released, causing a pH shift. This change is detected by chip sensors, which convert the electrochemical signal into digital data.

Routine Sample Processing Strategy

Circulating free nucleic acids are purified from 1 mL of clarified plasma. In particular, cfDNA is isolated by using the Qiamp Circulating Nucleic Acid Kit (Qiagen) and eluted with 50µL of Nuclease-free Water, following manufacturer instructions. The extracted cfDNA is stored at -20°C. The concentration of cfDNA is evaluated using a Qubit 4 fluorometer (ThermoFisher) with the Qubit 1X dsDNA High Sensitivity (HS) kit.

Extracted cfDNA samples are tested on Genexus (Thermo Fisher Scientific) system. The platform enables entire NGS workflows (from library preparation to data interpretation) within 24 hours. The OPA assay includes the most clinically relevant actionable genes for solid tumour patients. Firstly, samples are created on a dedicated server and assigned to a new run. Then, the Genexus platform is loaded with OPA primers, strip solutions, strip reagents, and supplies according to manufacturer instructions. A total of 10ng is required by the OPA assay on the Genexus platform. Accordingly, each sample is dispensed on a 96-well plate, following manufacturer instructions. Finally, nucleic acids are sequenced on a GX5™ chip that allows for the simultaneous processing of n = 4 samples in a single line with an OPA assay, for a maximum of 4 lanes (16 samples) in a row. Data analysis is performed using proprietary IonTorrent Genexus software (6.8.2.0). Particularly, detected alterations are annotated by adopting Oncomine Knowledgebase Reporter Software (Oncomine Reporter 5.0).

In addition, BAM files are also visually inspected with the Golden Helix Genome Browser v.2.0.7 (Bozeman, MT, USA) in hotspot regions in *EGFR*, *KRAS*, and *BRAF* lung cancer-addicted molecular alterations.

cfMeDIP-seq

cfMeDIP-seq is conducted following previously published protocols. In short, cfDNA libraries are generated using the Kapa Hyper Prep Kit (Roche) according to the manufacturer's guidelines. After performing end-repair and A-tailing, adaptors from the NEBNext Multiplex Oligos for Illumina (NEB) are ligated to the samples, followed by purification using AMPure XP beads.

To achieve a final quantity of 100 ng, Lambda DNA—comprising both methylated and unmethylated amplicons with varying CpG content—is added to the libraries. 0.3 ng of methylated and unmethylated *Arabidopsis thaliana* DNA is added for quality control purposes (Diagenode). One small part of the library is kept aside for input control (IC), and the remaining part was used for immunoprecipitation (IP).

MeDIP is carried out with the MagMeDIP Kit (Diagenode) and Antibody anti5mC* (33D3 clone) as per the manufacturer's protocol. The efficiency of the immunoprecipitation is verified via qPCR by detecting the recovery of the spiked-in Arabidopsis thaliana DNA (both methylated and unmethylated), following Diagenode's instructions. All samples with a specificity of reaction are sequenced at the resolution with a mean of 54.7 million reads per sample, resulting in ~5.1X depth per sample.

Processing of cfMeDIP-seq data

The quality of raw reads is evaluated using FastQC version 0.11.9 and MultiQC version 1.11. Low-quality reads and adaptors are removed with Trim Galore version 0.6.6. The trimmed reads are aligned to hg38 with Bowtie2 version 2.3.4.3. SAMTools version 1.9 is used to convert the SAM alignment files to BAM files, sort and index reads, and remove duplicates. Samples with <10M mapped reads are excluded. Tumour fraction is estimated using IchorCNA on the low-pass WGS of IC samples.

Processing of cfMeDIP-seq data

The quality of raw reads is evaluated using FastQC version 0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) and MultiQC version 1.11 (29). Then, low-quality reads and adaptors are removed with Trim Galore version 0.6.6 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore). The trimmed reads are aligned to hg38 with Bowtie2 version 2.3.4.3 (30). SAMTools version 1.9 (31) is used to convert the SAM alignment files to BAM files, sort and index reads, and remove duplicates. Samples with <10 M mapped reads are excluded. Tumour fraction is estimated using IchorCNA (20) on the low-pass WGS of IC samples.

Identification and annotation of differentially methylated regions (DMRs)

The filtered BAM files are processed using MEDIPS (32) to identify the Differentially Methylated Regions (DMRs) between LC patients with different hystotypes and stages. The enrichment scores relH and GoGe are estimated for each sample to express the grade of CpG enrichment in the DNA fragments compared

to the reference genome. The enrichment score relH is the ratio between the relative frequency of CpGs within the regions and the reference genome. The enrichment score GoGe is the observed/expected ratio of CpGs within the regions and the reference genome. Samples with a relH value less than 2.7 and/or a GoGe value less than 1.75 are excluded. Then, the genome of each sample is binned into 300-bp windows, and the methylation status of each bin is compared between the two groups. Regions with an absolute value of log2 fold change (FC) greater than or equal to 2 and a p-value less than 0.01 are selected as differentially methylated. The identified DMRs are annotated with the annotatr (33) R package. Gene set enrichment with DAVID and gene ontologies with a p-value less than 0.05 is selected.

Base Calling

In base calling, nucleotide sequences are "extracted" from the image data generated by sequencing platforms. Base-calling algorithms convert image information into sequence data. The process also corrects for artefacts such as crosstalk and phase errors. Crosstalk occurs due to overlapping fluorescence emissions of different nucleotides, while phasing is caused by signal dispersion and diffusion between cycles. Each base is assigned a quality score, called the "Phred quality score" (Q), which indicates the accuracy of base identification.

Alignment

Short DNA reads (200–8000 bp) are sequenced from either one or both ends of DNA fragments (single-end or paired-end reads), with typical lengths around 400 bp on platforms like 454. Alignment aims to locate these reads on a reference sequence, but challenges arise in regions that diverge significantly from the reference. Using longer or paired-end reads, which sequence DNA in both 5'-3' and 3'-5' directions, can improve alignment accuracy. A critical factor for successful assembly is coverage, defined as the number of times a sequence aligns with the reference, ensuring reliability and completeness in the reconstructed sequence.

Calibration of Quality Scores

Phred quality scores derived from alignment algorithms do not always accurately reflect real errors in base calling. Therefore, recalibration is performed, considering factors such as raw quality scores, the relative position of the base within the read, and the dinucleotide context.

Clinical Applications

The clinical applications of liquid biopsy depend on the approach used to study circulating tumour cells or ctDNA. A quantitative approach provides prognostic information, while a qualitative approach enables the analysis of predictive mutations, monitoring of clonal evolution, and adjustment of therapeutic strategies. ctDNA, released by apoptotic or necrotic tumour cells, provides DNA information from both primary lesions and metastases.

Imaging Techniques

CT will be performed using multidetector equipment (GE Revolution GSI 128 MDTC).
Clinical and radiological data will be collected to correlate with molecular and genomic data.

Radiologists' Responsibilities

Radiologists will be required to:

- Collect clinical information using a structured report (see sheet).
- Obtain informed consent from patients.

Extraction of Quantitative Features for Radiomics

Textural Features

Plot features will be obtained from manually segmented ROIs on CT images. They will include first-order features (mean, mode, median, standard deviation (std), median absolute deviation (MAD), range, kurtosis, skewness, and interquartile range (IQR) and second-order characteristics. For the latter, bandpass, wavelet, isotropic resampling, discretisation length corrections and different quantisation tools will be implemented. The first three sets are based on the grey-level co-occurrence matrix (GLCM), the grey-level run-length matrix (RLM), and the size zone matrix (SZM), all of which belong to the family of statistical matrices. Once these matrices have been constructed, it is possible to derive texture features (such as Haralick features and moments).
To improve robustness, advanced techniques like bandpass filtering, wavelet transformations, isotropic resampling, and quantisation corrections will be applied. Multi-grey-level SZM variants will also be utilised to compute texture features across various quantisation levels, combining the results using weighted averages to enhance sensitivity to subtle texture variations.

The formula for calculating the multi-gray-level SZM is as follows:

$$MSZM_f(s,g) = \sum_{k=1}^{8} w_k SZM_f^{N_k}(s,g)$$

The integration of these features enables detailed and multi-scale texture characterisation, optimising the ability to differentiate and classify lung lesions in CT imaging, even across diverse morphological and pathological presentations.

Morphological Features

A set of morphological features will be considered, including mean radial length, radial length entropy, irregularity, diameter, circularity, compactness, smoothness, roughness, rectangularity, convexity, eccentricity, and eulogy.

Classification Methods

Classification involves assigning an individual (such as a lesion or patient) to a specific class based on extracted features. This is done using a feature vector $x = [x(1),x(2), \ldots ,x(N)]$, where the classifier assigns the individual to one of K possible classes.
The process includes several steps:

- Choosing the Classification Criterion:
  This decision (linear or nonlinear) depends on the problem and the available data.
- Training:
  The classifier is trained on a data subset, typically using supervised learning with cross-validation techniques.
- Validation:
  The classifier's performance is tested on a separate dataset to evaluate its generalisation ability.

The performance of a classifier depends on the combination of features, algorithms, and training methods used. In recent years, deep learning techniques have gained popularity for their ability to identify critical features from large datasets automatically. eeThe following sections examine some of the most popular classification techniques and methods. In this study, all currently available techniques will be applied with the aim of finding the best combination in terms of classification performance.

Classifier Types

Classification techniques can be essentially divided into linear and nonlinear. Linear techniques adopt a linear combination (sum) of features to try to classify the individual. Such techniques (*e.g.*, Linear Discriminant Analysis, LDA) are helpful when features are chosen such that the problem is linearly separable. More often, the problem is not linearly separable, and therefore nonlinear techniques (such as neural networks, k-nearest-neighbours, and support vector machines) are more useful. Trees are a special type of nonlinear classifier that is based on successive dichotomous processes. At each step, the algorithm creates a binary separation, and each leaf is further divided into two at the next step. This type of algorithm is generally chosen for its 'human' comprehensibility. Dichotomies are binary decisions of the yes/no type on individual features, and thus their interpretation is transparent. In contrast, classification rules generated by linear or nonlinear algorithms are generally not understandable.

Cross-Validation

Cross-validation is an essential aspect of classifier training and aims to reduce possible overfitting, i.e., the tendency of training to select parameters that make the classifier very good at classifying individuals used as a training set, while the ability to generalise, i.e., classify individuals not belonging to the training set, is limited. This issue is related to the fact that, often, as in the present case, it is not possible to examine a significant sub-population that is representative of the entire population (all possible breast cancers, in this case). Therefore, it is necessary for the classifier to be able to have reasonable performance on the entire population.

Evaluation Metrics

Performance metrics for classifiers will include standard measures like True Positives, False Positives, ROC curves, and confusion matrices. For binary classifiers, the confusion matrix provides insight into misclassifications; for multi-class classifiers, more complex metrics are used.

Implementation in the Present Protocol

All pre-processing, DICOM image handling, and feature extraction will be conducted using Matlab (The MathWorks Inc., Natick, MA) or R (R Core Team, 2018). Matlab is widely used for scientific data processing and classification. At the same time, R is an open-source statistical analysis tool that has grown to support advanced techniques in machine learning and radiomics.

Statistical Analysis

The characteristics of the study population and other relevant variables will be described using the appro-

priate descriptive statistics for both continuous and categorical data.
Data will be presented with absolute frequencies and percentages, reporting the respective confidence limits. The mean and standard deviation will be reported for discrete parameters, following a Gaussian curve. Medians and interquartile ranges will be reported in cases where parameters are not distributed according to a Gaussian curve. Parametric and nonparametric tests for paired and unpaired data will be used, regardless of whether the data distribution is Gaussian or non-Gaussian, to detect statistically significant differences between groups. For continuous variables, the difference between median values for different groups will be calculated and tested using a two-sided Student t-test (if the differences are normally distributed) or the Mann-Whitney test (if the differences are generally not distributed. Assessment of inter-observer variability will be performed by calculating Cohen's Kappa index. Mixed-effects regression models will adjust for covariates in longitudinal data. Multivariate analysis, including linear classifiers, support vector machines, and decision trees, will explore feature combinations to optimise classification accuracy of lung lesions.
A p-value <0.05 will indicate statistical significance, with Bonferroni correction for multiple comparisons. Analyses will use Matlab Statistics Toolbox and R.

DISCUSSION

Precision medicine enables the targeted treatment of LC, including stage III, by applying multimodal omic strategies tailored to individual groups based on their genetics (34, 36).
Radiogenomics aims to correlate imaging phenotypes with gene and epigenetic modifications. Radiomics has recently emerged as a promising tool for discovering new imaging biomarkers. It can be applied to any field of diagnostic imaging and is used in various clinical settings. Radiogenomics is a specialised evolution of oncology radiomics that utilises imaging capabilities to non-invasively identify or predict tumour-specific genomic alterations (37-38).
The biopsy of the suspected cancer is today the gold standard for the characterisation of LC. However, it is expensive, invasive and evaluates only the sampled section of a heterogeneous tumour. The applicative and ambitious goal of the present study is to develop a new protocol and mathematical algorithm based on the imaging of the entire tumour or of a multifocal tumor load in a single patient, with the possibility of providing a non-invasive diagnosis correlating also the data derived from liquid biopsy on the gene mutations and epigenetic changes of the tumour.
Currently, there is no universal image acquisition protocol and no structured reporting standards (39) The method and application of the structured report could be adopted as a reporting method not only in LC but more generally in all cancers. The algorithm derived from the present study should be validated by scientific agencies and societies to transfer the obtained diagnostic procedures into the clinical setting and real-world practice.
Suppose the goals of the present project are successful. In that case, they will result in a significant reduction of health system expenses, allowing for highly personalised LC treatment and enabling early-stage diagnosis, thereby avoiding unnecessary treatments.
All this would bring enormous benefits to patients in terms of quality of life and social and productive contribution.

COMPLIANCE WITH ETHICAL STANDARDS

Funding

Funded by European Union-Next Generation EU, Missione 4 Componente 1 CUP_B53D23031840001.

Conflicts of interests

The authors declare no competing interests.

Data availability

Data collection and analysis are ongoing; data will be made available from the corresponding author upon reasonable request after study completion.

Author's contributions

All authors contributed equally to this work, read and approved the final version of the manuscript.

Ethical approval

The study was approved by the Ethics Committee "Comitato Etico Università degli Studi della Campania Luigi Vanvitelli" (approval No. 24997/2020) on 11th November 2020.

Publication ethics

The study was conducted in accordance with publication ethics standards.

Plagiarism

The authors declare that this manuscript is an original work and has not been published elsewhere.

Data falsification and fabrication

The authors declare that no data were fabricated or manipulated in this study.

REFERENCES

1. Wang J, Chang S, Li G, Sun Y. Application of liquid biopsy in precision medicine: opportunities and challenges. Front Med. 2017;11(4):522-7. doi: 10.1007/s11684-017-0526-7.
2. Hench IB, Hench J, Tolnay M. Liquid Biopsy in Clinical Management of Breast, Lung, and Colorectal Cancer. Front Med (Lausanne). 2018;5:9. doi: 10.3389/fmed.2018.00009.
3. Mader S, Pantel K. Liquid Biopsy: Current Status and Future Perspectives. Oncol Res Treat. 2017;40(7-8):404-8. doi: 10.1159/000478018.
4. Jaffe TA, Wickersham NW, Sullivan DC. Quantitative imaging in oncology patients: Part 2, oncologists' opinions and expectations at major U.S. cancer centers. AJR Am J Roentgenol. 2010;195(1):W19-30. doi: 10.2214/AJR.09.3541.
5. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol. 2015;60(14):5471-96. doi: 10.1088/0031-9155/60/14/5471.
6. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics. 1973;SMC-3(6):610-21. doi: 10.1109/TSMC.1973.4309314.
7. Wang J, Zhou X, Bradley PL, Chang SF, Perrimon N, Wong ST. Cellular phenotype recognition for high-content RNA interference genome-wide screening. J Biomol Screen. 2008;13(1):29-39. doi: 10.1177/1087057107311223.
8. Galloway MM. Texture analysis using gray level run lengths. Computer Graphics and Image Processing. 1975;4(2):172-9. doi: 10.1016/S0146-664X(75)80008-6.
9. Chu A, Sehgal CM, Greenleaf JF. Use of gray value distribution of run lengths for texture analysis. 1990;11:415-9. doi:10.1016/0167-8655(90)90112-F.
10. Dasarathy BV, Holder EB. Image characterizations based on joint gray level—run length distributions. Pattern Recognition Letters. 1991;12(8):497-502. doi: 10.1016/0167-8655(91)80014-2.
11. Thibault G, Fertil B, Navarro C, Pereira S, Lévy N, Sequeira J, et al. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. 2009.
12. Luo H, Wei W, Ye Z, Zheng J, Xu RH. Liquid Biopsy of Methylation Biomarkers in Cell-Free DNA. Trends Mol Med. 2021;27(5):482-500. doi: 10.1016/j.molmed.2020.12.011.
13. Li S, Tollefsbol TO. DNA methylation methods: Global DNA methylation and methylomic analyses. Methods (San Diego, Calif). 2021;187:28-43. doi: 10.1016/j.ymeth.2020.10.002.
14. Li S, Tollefsbol TO. DNA methylation methods: Global DNA methylation and methylomic analyses. Methods. 2021;187:28-43. doi: 10.1016/j.ymeth.2020.10.002.
15. Nuzzo PV, Berchuck JE, Korthauer K, Spisak S, Nassar AH, Abou Alaiwi S, et al. Detection of renal cell carcinoma using plasma and urine cell-free DNA methylomes. Nat Med. 2020;26(7):1041-3. doi: 10.1038/s41591-020-0933-1.
16. Chen S, Petricca J, Ye W, Guan J, Zeng Y, Cheng N, et al. The cell-free DNA methylome captures distinctions between localized and metastatic prostate tumors. Nat Commun. 2022;13(1):6467. doi: 10.1038/s41467-022-34012-2.
17. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. IEEE Trans Biomed Eng. 2014;61(3):630-7. doi: 10.1109/TBME.2013.2284600.
18. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition. 1996;29(1):51-9. doi: 10.1016/0031-3203(95)00067-4.
19. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002;24(7):971-87. doi: 10.1109/TPAMI.2002.1017623.
20. Serra JP, Serra J. Image Analysis and Mathematical Morphology: Academic Press; 1982.

21. Fusco R, Di Marzo M, Sansone C, Sansone M, Petrillo A. Breast DCE-MRI: lesion classification using dynamic and morphological features by means of a multiple classifier system. Eur Radiol Exp. 2017;1(1):10. doi: 10.1186/s41747-017-0007-4.
22. Flusser J. Moment Invariants in Image Analysis. Transactions on engineering, Computing and Technology. 2011;VII.
23. Maragos P. Pattern spectrum and multiscale shape representation. IEEE Trans Pattern Anal Mach Intell. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1989;11:701-16.
24. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. IEEE Trans Biomed Eng. 2014;61(3):630-7. doi: 10.1109/TBME.2013.2284600.
25. Fusco R, Granata V, Rega D, Russo C, Pace U, Pecori B, et al. Morphological and functional features prognostic factor of magnetic resonance imaging in locally advanced rectal cancer. Acta radiologica (Stockholm, Sweden : 1987). 2019;60(7):815-25. doi: 10.1177/0284185118803783.
26. Fusco R, Sansone M, Filice S, Carone G, Amato DM, Sansone C, Petrillo A. Pattern Recognition Approaches for Breast Cancer DCE-MRI Classification: A Systematic Review. J Med Biol Eng. 2016;36(4):449-459. doi: 10.1007/s40846-016-0163-7.
27. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. Lancet Oncol. 2016;17(12):e542-e551. doi: 10.1016/S1470-2045(16)30406-5.
28. Garofalo A, Sholl L, Reardon B, Taylor-Weiner A, Amin-Mansour A, Miao D, et al. The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine. Genome Med. 2016;8(1):79. doi: 10.1186/s13073-016-0333-9.
29. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8. doi: 10.1093/bioinformatics/btw354.
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9. doi: 10.1038/nmeth.1923.
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi: 10.1093/bioinformatics/btp352.
32. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics. 2014;30(2):284-6. doi: 10.1093/bioinformatics/btt650.
33. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. Bioinformatics. 2017;33(15):2381-2383. doi: 10.1093/bioinformatics/btx183.
34. Campesato LF, Barroso-Sousa R, Jimenez L, Correa BR, Sabbaga J, Hoff PM, et al. Comprehensive cancer-gene panels can be used to estimate mutational load and predict clinical benefit to PD-1 blockade in clinical practice. Oncotarget. 2015;6(33):34221-7. doi: 10.18632/oncotarget.5950.
35. Champiat S, Ferté C, Lebel-Binay S, Eggermont A, Soria JC. Exomics and immunogenics: Bridging mutational load and immune checkpoints efficacy. Oncoimmunology. 2014;3(1):e27817. doi: 10.4161/onci.27817.
36. Zou W, Wolchok JD, Chen L. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations. Sci Transl Med. 2016;8(328):328rv4. doi: 10.1126/scitranslmed.aad7118.
37. Pitt JM, Vétizou M, Daillère R, Roberti MP, Yamazaki T, Routy B, et al. Resistance Mechanisms to Immune-Checkpoint Blockade in Cancer: Tumor-Intrinsic and -Extrin-sic Factors. Immunity. 2016;44(6):1255-69.
38. Morrison C. 'Financial toxicity' looms as cancer combinations proliferate. Nature biotechnology. 2015;33(8):783-4. doi: 10.1038/nbt0815-783.
39. Robinson K, Li H, Lan L, Schacht D, Giger M. Radiomics robustness assessment and classification evaluation: A two-stage method demonstrated on multivendor FFDM. Med Phys. 2019 May;46(5):2145-2156. doi: 10.1002/mp.13455.