

Article

Not peer-reviewed version

---

# Uncovering the Diagnostic Power of Radiomic Feature Significance in Automated Lung Cancer Detection: An Integrative Analysis of Texture, Shape, and Intensity Contributions

---

[Sotiris Raptis](#)\*, Christos Ilioudis, [Kiki Theodorou](#)

Posted Date: 22 November 2024

doi: 10.20944/preprints202411.1621.v1

Keywords: lung cancer; radiomics; machine learning; feature importance; SHAP; CT imaging; diagnostic accuracy; clinical interpretability



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

# Uncovering the Diagnostic Power of Radiomic Feature Significance in Automated Lung Cancer Detection: An Integrative Analysis of Texture, Shape, and Intensity Contributions

Sotiris Raptis <sup>1</sup>, Christos Ilioudis <sup>2</sup> and Kiki Theodorou <sup>1,\*</sup>

<sup>1</sup> Medical Physics Department, Medical School, University of Thessaly, 41500 Larisa, Greece

<sup>2</sup> Department of Information and Electronic Engineering, International Hellenic University (IHU), 57001 Thessaloniki, Greece

\* Correspondence: sraptis@uth.gr

**Abstract: Background:** Lung cancer still maintains the leading position among causes of death in the world; the process of early detection surely contributes to changes in the survival of patients. Standard diagnostic methods are grossly insensitive, especially in early stages. In this paper, radiomic features are discussed that can assure improved diagnostic accuracy through automated lung cancer detection by considering the important feature categories, such as texture, shape, and intensity, originating from the CT DICOM images. **Methods:** We developed and compared the performance of two machine learning models-DenseNet-201 CNN and XGBoost-trained on radiomic features with the ability to identify malignant tumors from benign ones. Feature importance was analyzed using SHAP and techniques of Permutation Importance that enhance both Global and Case-Specific interpretability of the models. **Results:** Few features that reflect tumor heterogeneity and morphology include GLCM Entropy, Shape Compactness, and Surface Area to Volume Ratio. These performed excellently in diagnosis, with DenseNet-201 producing an accuracy of 92.4% and XGBoost at 89.7%. The analysis of feature interpretability ascertains its potential in early detection and boosting diagnostic confidence. **Conclusions:** The current work identifies the most important radiomic features and quantifies their diagnostic significance through a properly conducted feature selection process reflecting stability analysis. This provides the blueprint for feature-driven model interpretability in clinical applications. Radiomics features have great value in the automated diagnosis of lung cancer, especially when combined with machine learning models. This might improve early detection and open personalized diagnostic strategies for precision oncology.

**Keywords:** lung cancer; radiomics; machine learning; feature importance; SHAP; CT imaging; diagnostic accuracy; clinical interpretability

## 1. Introduction

Lung cancer remains one of the leading causes of death worldwide, due to more than 1.7 million deaths annually, and is considered the most common cause of cancer-related mortality [1]. Yet, while therapeutic options have been ever-increasing and ever-improving, the prognosis remains largely dependent upon the stage of diagnosis at presentation; early detection fundamentally enhances a patient's likelihood of survival. Unfortunately, because of a lack of sensitivity and specificity, current non-invasive diagnostic methods perform poorly, especially in the early stages of the disease. The shortfall underlines the need for novel diagnostic approaches that could offer both precision and non-invasiveness. In this regard, radiomics—a high-throughput feature extraction of medical images—can be performed to provide a strong enhancement of the diagnostic accuracy [2]. Radiomics offers detailed imaging analytics of conventional modalities such as CT (Computed Tomography) through features extracted from tumor shape, texture, and intensity. These features represent complicated tumor heterogeneity that usually is not visible to the naked eye, therefore allowing for the creation

of a complete non-invasive tumor phenotype. To this aim, this study systematically explores the diagnostic value of radiomic features in lung cancer [3]. According to our theory, shape-based features become more significant as tumors grow larger and change shape over time, whereas texture-based features are more useful for locating early tumor lesions due to the value of representing intratumoral heterogeneity [4]. In previous studies, we explored the diagnostic potential of radiomic features in cancer prediction and demonstrated how machine learning techniques could effectively leverage these features to improve accuracy in early cancer detection. Our past research highlighted how radiomics can be used for personalized diagnosis. It focuses not only on individual feature importance but also on how those features might interact with each other and with the clinical context to improve diagnostic accuracy. In this research, we integrate radiomics with machine learning to offer a more nuanced approach to the diagnosis of lung cancer, one that moves beyond binary classifications and truly embraces the complexity of tumor phenotypes. This work, therefore, was focused on the isolation and validation of the diagnostic value of individual radiomic features, aiming to bridge gaps between high-performance models and their interpretability in a targeted approach that may enhance the clinical usability of radiomic-based lung cancer detection models. The umbrella approach is primarily expanding under precision medicine, wherein treatments and diagnoses have been tailored according to the individual characteristics of the patients. Having continuously developed deep learning and high-resolution imaging data, it is now time to give more granularity to radiomic feature importance in lung cancer diagnosis. Techniques like SHAP (Shapley additive explanations) and permutation feature importance give us numerical information about how much each feature contributes [5]. The goal is to make models that are very accurate and easy to understand, which is a key requirement for clinical use. An overview of the radiomics workflow, from DICOM (Digital Imaging and Communications in Medicine) image acquisition to model validation, is presented in Figure 1. This workflow highlights key preprocessing steps, feature extraction, and model training processes essential to this study.



**Figure 1.** Overview of the radiomics workflow used in this study.

## 2. Materials and Methods

### 2.1. Data Acquisition and Preprocessing

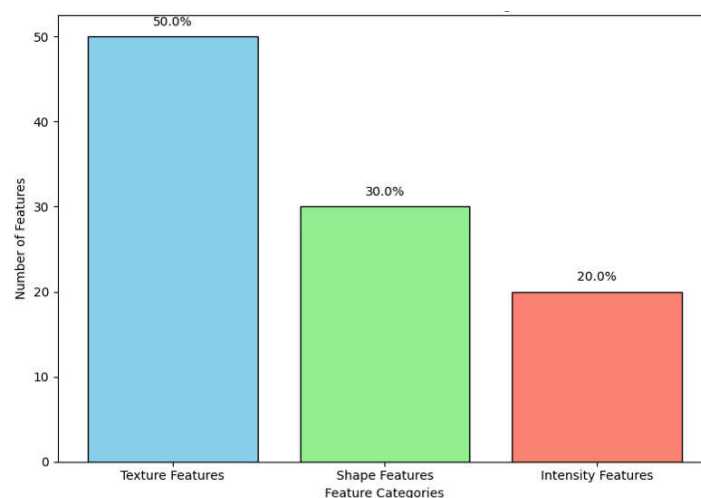
In the present work, one major and diverse CT-DICOM image dataset was considered, containing both verified lung cancer cases and normal controls. Altogether, this dataset was made up of 2,963 images relevant to lung cancer patients and 383 images relevant to healthy patients, derived from publicly available databases and clinical repositories from a medical center that collaborated with our laboratory <sup>1</sup>, which included a diverse range of CT scans and radiographs of patients undergoing chest radiotherapy for various thoracic malignancies [4,6,7]. Thereafter, the dataset was split into a validation set comprising 1,786 cancerous images and 337 healthy ones to ensure unbiased model performance across cohorts. The Image Biomarker Standardization Initiative (IBSI) provided some common parameters that can ensure the reproducibility and robustness of radiomic feature extraction [8].

<sup>1</sup> *Laboratory: Medical Physics Department (MPD), The Medical Physics Department (MPD) at University Hospital, Larissa, Greece, is involved in clinical practice, research, and education. MPD offers clinical and research services related to patient diagnosis and treatment quality assurance programs, acceptance tests, and radiation protection issues.*

It was taken as the guiding principle to establish the preprocessing pipeline. Besides, to ensure quality and eliminate all sorts of artifacts that may distort radiomic feature extraction, the images were carefully reviewed by qualified radiologists.

## 2.2. Radiomic Feature Extraction

In this paper, we describe the application of an intensive and standardized extraction using PyRadiomics, one of the most widely recognized and validated open-source software platforms developed specifically for radiomic analyses [9]. In sum, completeness, following standardized protocols, flexibility, and open-source availability are the reasons why PyRadiomics was chosen. A total of 350 radiomic features will be extracted from each ROI (Region Of Interest) to process CT images in DICOM format. These features include first-order statistics, shape-based descriptors, and texture-based metrics. Features were selected based on how well they are able to represent some critical characteristics of the tumor, such as heterogeneity, complexity, and morphological structure. Carefully selected features ensured that the model had access to the most relevant data for the most accurate diagnosis of lung cancer [10]. As shown in Figure 2, texture features constitute the largest category within our dataset, followed by shape and intensity features. This distribution underscores the diverse information captured from DICOM images for a comprehensive tumor characterization.



**Figure 2.** Distribution of radiomic feature categories extracted in this study.

### 2.2.1. First Order Statistics

First-order statistics describe the distribution of individual voxel intensities within the ROI and are hence utilized to summarize the intensity characteristics. These features convey the basic but highly fundamental understanding of the overall structure of the tumor and include:

- **Mean Intensity:** The mean value of the voxels comprised within the ROI, reflecting to a great extent the average density of the tissue.
- **Skewness:** Characterizes the asymmetry of the voxel intensity distribution and hence, the heterogeneity of the tumor.
- **Kurtosis:** describing the property 'peakedness', in other words, whether a distribution of voxel intensities is bunched around or spread out from the mean.
- **Energy:** This is the sum of the square of the voxel values, which may be related to the aggressiveness of the tumor or its metabolic activity.

These first-order features are simple but not worthless. They provide a view on the basic character of the tumor, especially from the early beginning of tumor development onward.

### 2.2.2. Shape-Based Features

Shape-based feature features are of immense importance in understanding the geometry of the tumor and thus provide knowledge of how a tumor grows and interacts with surrounding tissues. Lung cancer tumors have been found to be mostly irregularly shaped, which is sometimes not captured through traditional imaging assessments. Shape-based features included in the extraction study are:

- Sphericity: A measure of how spherical (round) the tumor is, where values approaching 1 depict structures near perfectly spherical. Lower sphericity values indicate that highly aggressive and invasive tumors are more characteristic.
- Compactness: Shows how much the shape of a tumor is spherical or elongated and can thus be an indication of its invasive power.
- Surface Area to Volume Ratio: This measure compares the surface complexity of the tumor to its volume. The higher the ratio, the greater the chance for irregular growth patterns—often associated with malignancy.
- Elongation: Measures the deviation in the tumor shape from a perfect sphere and may give clues of its infiltration into surrounding tissues.

Shape descriptors bear particular importance for late-stage tumors, as the growth patterns may lead to fundamental insights into the aggressiveness of the cancer and its metastasis potential. In light of this viewpoint, we make the assumption that shape-based features would become more significant as size increased due to structural deformations brought about by unchecked growth.

### 2.2.3. Texture-Based Features

These features are probably the most powerful descriptors of intratumoral heterogeneity, with tumor heterogeneity at both microscopic and macroscopic levels actually being one of the hallmark features of cancer. In fact, these provide a quantitative measurement of the variation in the intensities of voxels within a tumor that could unveil the underlying biological processes such as cell density, necrosis, and angiogenesis. The main features of texture extracted in this study included measures derived from:

- GLCM: GLCM (Gray-Level Co-occurrence Matrix) features describe the frequency of pixel intensity pairs for a predefined spatial relationship. GLCM entropy may serve as an example and denotes the complexity of the variation in voxel intensities. The greater the values, the greater the heterogeneity, which is usually associated with malignancies. Other significant GLCM features include contrast, which describes the difference between high and low intensities, and correlation, which is the measure of linear dependencies between the intensity of voxels.
- GLRLM: This matrix provides the length of consecutively sharing voxels with the same intensity value in a given direction. The related features to this are the GLRLM (Gray-Level Run Length Matrix) Short Run Emphasis, which manifests the presence of small homogeneous regions inside the tumor, while the GLRLM Long Run Emphasis gives information about the boundless homogeneous regions. These will be relevant for identifying the fibrotic regions or regions bearing necrosis inside the tumor.
- Grayscale Level Size Zone Matrix: GLSZM is very similar to the GLRLM in that it quantifies regions of identical intensities; no directional information is taken into account, however. Important features are relying on this matrix: Zone Size Non-Uniformity and Large Zone Emphasis may be useful to detect the presence of large homogeneous areas, generally indicative of late disease stages.
- Wavelet features: Refer to those signal features extracted through the application of wavelet filters on images for capturing texture at multiple resolutions. This multi-scale analysis is critical in order to detect subtle patterns both at fine and coarse details, thus offering a more nuanced understanding of tumor heterogeneity.

These features provide a fine-grained detail from the internal structure of the tumor, well beyond what is directly observable from the raw imaging data. In the cases of early-stage tumors, we postulated that the texture features would turn out to be more informative since early malignancies often manifest as small regions of heterogeneous intensity, which can be potentially detectable by advanced metrics of texture [11].

### 2.2.4. Multi-Scale and Multi-Dimensional Feature Extraction

One of the novelties of this work was the inclusion of multi-scale and multi-dimensional feature extraction strategies. This combination of features derived from different resolutions and dimensions gave us a more robust dataset representative of both global and local tumor characteristics. This will

be a way to ensure that machine learning models have access to the most complete representation of the tumor phenotype, thus allowing the detection of subtle yet clinically significant differences between malignant and benign regions.

### 2.2.5. Innovative Aspects of Feature Extraction

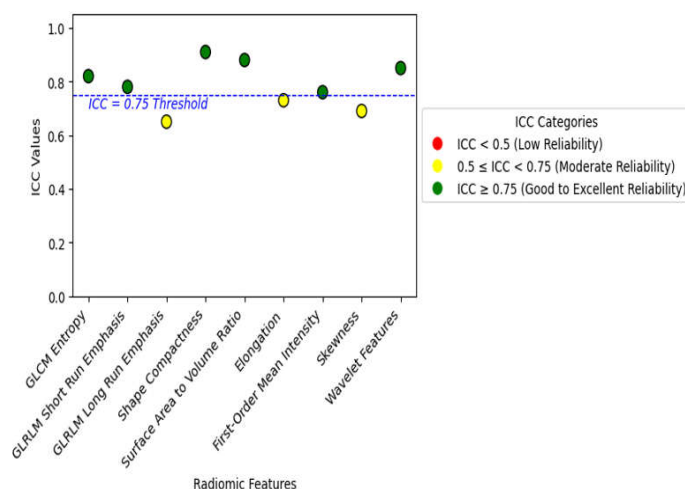
- **Manual and Semi-Automated Segmentation:** The tumor ROIs were segmented using a great deal of care by manual delineation of expert radiologists combined with semi-automated algorithms to achieve the best precision together with reducing observer variability.
- **Standardization and Reproducibility:** All extraction was carried out by strictly adhering to guidelines provided by the IBSI, with the aim of having all features reproducible across different imaging systems and acquisition protocols.
- **High-Throughput Feature Extraction:** PyRadiomics thus enables the extraction of hundreds of features in one image effectively. High-throughput feature extraction in imaging data is of high importance to machine learning applications, especially in cases where the volume of data is immense.
- **Finding the Best and Most Useful Features:** Out of the 350 features that were collected, tests using forward stepwise correlation analysis and tests for test-retest variability were used to get rid of features that were duplicated or not relevant. We were able to minimize overfitting without reducing model accuracy by focusing the analysis on the most stable and informative features.

### 2.3. Feature Selection and Stability Analysis

Feature selection is, therefore, an indispensable step in machine learning for radiomics, since it features prominently in the accuracy, interpretability, and generalizability of any model developed from such data.

#### 2.3.1. Feature Stability and Reproducibility

Therefore, appropriate stability analysis was performed before feature selection in order to ensure that the features used in the model would be reproducible and robust, taking into account various imaging conditions. Various intraclass correlation coefficients of each feature are measured to check the reproducibility of their measurements. Features with ICC (Intraclass Correlation Coefficient)  $> 0.75$  are considered highly stable and were thus retained for further investigation [12]. There is a very valid statistical basis for selecting the ICC threshold of  $>0.75$  in choosing the ICC threshold, which will support the overall objective of the study in coming up with a reliable and generalizable diagnostic model. That being the case, it chose to select only the most stable features so the machine learning models become not only accurate but also strong across different conditions of imaging—a factor that makes them more applicable in real-world clinical settings. ICC has become a very common metric in the literature for quantifying the stability of features. Lambin et al. introduced the need for radiomic features to be reproducible and robust since that would provide the clinical utility of such features [13]. ICC thresholds were also used in the studies of Parmar et al. in order to filter out those features that were inconsistent across different image acquisitions with the purpose of making radiomics-based models more robust [14]. In this regard, their study showed that increased reliability of features with a higher ICC threshold—for example,  $>0.75$ —led to the improved performance of a model once applied to external datasets. Those with lower stability were discarded. Furthermore, to assess the inter-observer variability of feature extraction, we asked several radiologists to perform a manual segmentation of the same set of images. As shown in Figure 3, the features that showed high variability between observers were excluded so that only reproducible features, not biased by potential manual segmentation, were considered in further analyses.



**Figure 3.** Distribution of radiomic features based on ICC values.

### 2.3.2. Correlation and Dimensionality Reduction

Most of the radiomic features, more so those derived from texture matrices, have high correlations with one another, leading to multicollinearity in machine learning models. Multicollinearity increases model complexity and decreases interpretability. In this respect, it even degrades performance by making the model sensitive to noise.

To handle this, we performed a Pearson correlation in order to identify highly correlated features and remove them [15]. We thresholded the correlations at 0.9, meaning that for any pair of features with a correlation coefficient greater than 0.9, one of the pair was removed from the dataset. This threshold ensures that the model retains only independent and unique information while sustaining most of its predictive power, reducing the dimensionality of the feature space.

### 2.3.3. Feature Selection Techniques

In the following stability and correlation analysis, we performed a two-step feature selection strategy, balancing interpretability with prediction performance. The two key techniques were LASSO (Least Absolute Shrinkage and Selection Operator) regression and recursive feature elimination (RFE) [16]:

- **LASSO Regression:** LASSO introduced regularization into the feature selection process by applying a penalty to the magnitude of the coefficients associated with each feature. This is a powerful method in high-dimensional datasets, as it shrinks the coefficients of less important features to zero, effectively removing them from the model.
- **Recursive Feature Elimination (RFE):** This second feature selection method provided a wrapper method, working in conjunction with a machine learning model. RFE recursively removed the least important features based on their contribution to model performance until only the optimal set of features remained. This method ensures that the final feature set is tailored to the specific model being used, maximizing accuracy and interpretability [17].

#### *Permutation Importance and SHAP Analysis*

To evaluate the relative importance of the selected features, we employed two complementary feature ranking techniques: permutation importance and SHAP. These methods help identify which features contribute to the model's predictions and ensure that these features are not only statistically significant but clinically relevant [18]. Permutation Importance, this model-independent method checks how much the model's performance drops when a specific feature's values are randomly permuted. A larger decrease in performance means the model is more dependent on that feature. SHAP values offer a game-theoretic approach to feature importance, providing both global and local interpretability. It quantifies how much each feature changes the prediction, making it more interpretable than many black-box algorithms. This was especially useful for identifying features that

may have a small global effect but are crucial for individual predictions, supporting personalized diagnostic applications.

#### *Model Interpretability and Clinical Relevance*

A key objective of feature selection in this study was to enhance model interpretability. In clinical settings, understanding why a model makes a prediction is just as important as the prediction itself. We evaluated the final selected feature set not only for its predictive power but also for its clinical relevance. The model provided insights into tumor heterogeneity and morphology in a clinically interpretable way, highlighting stable and meaningful radiomic features. This interpretability makes the model more likely to be adopted in clinical practice, as it offers actionable information that healthcare professionals can clearly understand.

### *2.4. Machine Learning Model Development*

An accurate machine learning model is the heart of radiomics-based cancer diagnosis, which allows conversion from high-dimensional feature space into clinically useful prediction. We focused on building a model to differentiate between malignant and benign tumors with the aid of features extracted from CT DICOM images (as described in Section 2.2), which have been available for quite some time now. This could help us to extract the abstract heterogeneity of lung tumors through features derived from imaging, which were also very important in early detection and diagnosis.

#### *2.4.1. Model Architecture and Choice*

In this study, two main machine learning models were developed: a Convolution Neural Network (CNN) using the architecture of DenseNet-201 and an XGBoost algorithm. The chosen models thus had to cope well with the medical imaging data and radiomic feature complexity, providing a compromise between predictive power and model interpretability to allow for a hybrid diagnostic approach [19].

- **DenseNet-201 (CNN Model):** DenseNet-201 is a popular CNN architecture that effectively performs image feature extraction through dense connections between layers. This dense connectivity mitigates the vanishing gradient problem by helping in reusing features across layers to learn more complex imaging tasks like detecting cancer. To this end, in our study we trained the DenseNet model separately on both raw CT images and feature maps to enable the network to learn from spatial patterns (raw images) as well as quantitative tumor characteristics (radiomics-driven features) [20].

Although CNNs are often used on pixel-based data, we incorporated extracted features as extra input channels for the DenseNet model. That meant the CNN got to use high-level quantitative descriptors—things like shape, texture, and intensity histograms that helped it make its later decisions in addition to all of the raw image data. Here, we present a multi-modal input design to improve the model's diagnostic accuracy by integrating pixel-based and feature-based analysis.

- **XGBoost (Gradient-Boosted Decision Tree Model):** XGBoost is a powerful machine learning algorithm that is known to perform well on tabular data and has become very popular in the world of Kaggle competitions, where it generally reaches SOTA (State-Of-The-Art) scores. Because of its feature importance ranking and model interpretability, CNNs are not very good at being mean for image data. Training the XGBoost model on extracted features allowed us to test the importance that individual features hold in diagnostic tasks and thus infer about biological interpretability [21]. The model was then trained on the entire set of extracted features, and feature importance was evaluated using SHAP values and permutation importance to rank these race predictors (explained in Section 2.3).

#### *2.4.2. Training Process and Cross-Validation*

Accordingly, both models have been developed by using a five-fold cross-validation strategy in order to make sure that the results were generalizable and not overly influenced by the training data. Cross-validation, which is an essential step in ensuring that overfitting does not significantly reduce

model performance on unseen data, is particularly important in medical imaging studies, as representative cases are limited.

1. **Five-fold Cross-Validation:** The dataset was divided into five equal sized subsets, and in each fold of the cross-validation, one subset played the role of a validation set while the other four subsets played the role of training sets. This was repeated five times in such a way that each subset once acted as the validation set. Averaging over five folds provides a robust estimate of accuracy, sensitivity, specificity, and AUC-ROC for model performance.

2. **Data Augmentation:** For the training of the CNN to enhance generalization, the data augmentation techniques were used on the CT images. Then, several augmentation techniques—like random rotations, flipping, cropping, and scaling—were used to introduce patient positioning and variability in scanner settings, which will increase distortions and the diversity of the training.

3. **Hyperparameter Tuning:** These activities also involve the tuning of learning rate, batch size, and dropout rate by a grid search strategy. The early stopping in this CNN model stops the training when it observes that there is no more improvement in validation loss within five successive epochs to avoid overfitting and saving computational resources. Similarly, XGBoost hyperparameters, such as but not limited to those involving the maximum depth of trees, learning rate, and regularization parameters, have been tuned in a manner that predictive performance is maximized at the possible cost of overfitting.

#### 2.4.3. Model Optimization and Loss Functions

Thanks to the combined loss functions, comprised of both segmentation and classification, we managed to maximize its performance. Specifically, the dice loss function has been used to address segmentation accuracy for tumor regions, while a binary cross-entropy has been utilized for the classification task with regard to malignant/benign nature determination in tumors. Dice loss is a largely used metric in medical image segmentation and calculates the overlap between the predicted and actual tumor masks. This metric came in handy, especially for the segmentation part of the CNN in delineating the tumor boundary accurately. The model improves the Dice coefficient, acting to better define the tumor region. To classify, the binary cross-entropy loss function penalizes the network for each wrong prediction made, thereby allowing the model to be directed toward a more correct classification. The use of the network with this loss function has further allowed it to learn just how to label correctly and effectively both cancerous and non-cancerous regions from the raw imagery data and the radiomic feature maps.

#### 2.4.4. Performance Metrics

- **Accuracy:** It refers to the measure of the portion of properly classified images, thus giving the general performance of the model.
- **Sensitivity (recall):** It is the capability of the model to correctly identify malignant cases. High sensitivity is rather important in cancer diagnosis because it will reduce false negatives.
- **Specificity:** The percentage of healthy cases correct that this model should identify in order to avoid the misclassification of a benign case as malignant.
- **AUC-ROC (Area Under the Receiver Operating Characteristic Curve):** a metric that will provide the model performance for distinguishing malignant from benign cases across all possible classification thresholds. The high value of AUC-ROC means it would be unproblematic for this classification model to make good distinctions between the classes, even if there is a class imbalance in this dataset.

#### 2.5. Feature Importance and Interpretability

Similar to any other area of medical imaging and machine learning based on radiomics, interpretability is just as important as predictive accuracy. If models are to be clinically actionable, then clinicians must understand the rationale behind each prediction. Then again, the contribution of each feature that has emerged with a view not only to identify but also interpret the model's prediction has been our focus in the present study. For this work, SHAP was selected because it has

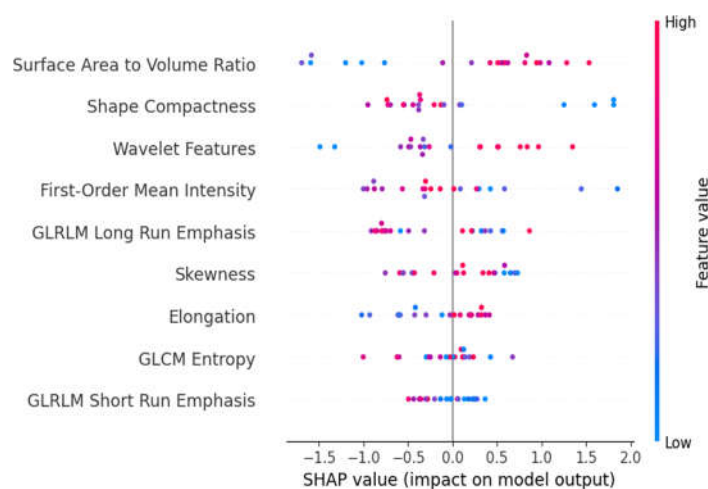
capabilities for both the explanation of individual predictions, a local interpretability, and the aggregation of the feature contributions across all the predictions—or global interpretability[22].

- **Global SHAP Analysis:** We averaged the SHAP values of every feature on the dataset and ranked the features by their general importance. Indeed, for every test we ran in this work, the features that ranked in the top places included GLCM Entropy and GLRLM Run Length Non-Uniformity; thus, our intuition about tumor heterogeneity being one of the key factors in malignant/benign tumor classification was legitimate.
- **Local SHAP Analysis:** SHAP also allowed us to drill into the contribution of features on a case-by-case basis. For example, in the specific cases of early-stage lung cancer, features related to first-order skewness, which reflects the asymmetry of voxel intensity distribution, have a high SHAP value and hence are critical for the right classification of small irregular tumors. That is very important for local interpretability in clinical applications, as it would tell the radiologists which features drove the model's decision for each given patient.

One interesting aspect of this study was the comparison of feature importance rankings obtained from the DenseNet-201 CNN model, which utilized both raw image data and features, with those given by the XGBoost model, utilizing only features. In this way, one can grasp how the integration of pixel-based and feature-based data impacts every single feature's importance. The CNN model considered the features as new input channels, together with the raw CT images. In this way, the network learned the complex patterns of the raw data, taking into consideration the quantitative features describing shape compactness and texture homogeneity. Feature importance rankings provided by both SHAP and permutation tests showed that even when presented with raw image data, the texture features still dominated. This therefore further underlines that features capture clinically significant information that may not easily be inferred from raw pixel data alone.

XGBoost, on its part, shows feature importance directly. Given that this model is trained only on radiomic features, GLCM Entropy and GLRLM Long Run Emphasis are the two most important features. This also goes as one would expect from the two metrics, which are really sensitive in bringing out malignant changes in tumor structure. Using only the radiomic data, XGBoost is more focused on how each feature contributes toward its final prediction. By comparing the results of both models, we were also able to confirm that the most important features were indeed the robust features, such as GLCM Entropy and GLRLM Non-Uniformity, across different learning paradigms.

One of the main goals with this study was to make sure that the machine learning model was interpretable and clinically actionable, not just accurate. SHAP values described why the model had made its prediction for a given patient, which is so crucial for the adoption of AI (Artificial Intelligence) models in healthcare. Besides, the clinical interpretability of the model was enriched by focusing on features with frank biological significance: As illustrated in Figure 4, SHAP values highlight the global impact of selected radiomic features on model predictions, with GLCM Entropy and Shape Compactness emerging as key drivers of classification accuracy. Each point represents a SHAP value for a particular feature and sample, with color indicating feature value (red for high and blue for low).



**Figure 4.** SHAP Summary Plot illustrating the global impact of selected radiomic features on model predictions.

### 2.6. Tumor-Specific Features Relevance

These features characterize tumors based on a wide range of aspects, which include shape, texture, and intensity; however, the importance of most of these features' changes with changes in factors related to tumor size, stage, and location. In this work, our goal was to explore how much the importance of single features would change across different tumor phenotypes for insights into how heterogeneity impacts diagnostic accuracy. We explored the dynamic nature of these features by stratifying the tumors based on size and stage. Moreover, we accentuated the personalized diagnostic potential of radiomics in lung cancer. Tumor size represents an important attribute that modulates the relevance of radiomic features. Larger tumors usually show more geometrically complex and irregular shapes, whereas smaller tumors may only show subtle variations with respect to texture-based features. In this work, small-sized tumors were expected to rely more heavily on texture-based features due to the accentuation of intratumoral heterogeneity, while shape-based features were hypothesized to become more influential as the tumor increases in size and undertakes greater morphological distortion. To test this hypothesis, we stratified the tumors into three size categories based on the TNM (Tumor-Nodes-Metastasis) system [23]:

1. **Small Tumors (< 2 cm):** earlier-stage tumors, which, due to their small size, are often very hard to detect because they do not cause much morphological disruption. For these cases, we expected that texture-based features would convey the most information since these features capture the subtle heterogeneity in the tumor that might not be obvious to the naked eye from the raw images. This was indeed the case, as all texture features ranked higher in importance for smaller tumors, in keeping with our hypothesis that early malignancies are more micro-heterogeneous.
2. **Medium-Sized Tumors (2–4 cm):** In tumors of intermediate size, it was expected that both the texture and shape-based features had a balanced contribution. For such tumors, shape elongation and GLCM correlation were important features, for they extracted information on the change in geometry of the tumor during growth and more structured texture patterns within the tumor mass. This underlines the dual importance of texture and shape features within this size category in a way to hint at the necessity of comprehensive feature sets that express both aspects of tumor growth.
3. **Large Tumors (> 4 cm):** We expected that with the larger size and more advanced development of tumors, shape-based features such as surface area to volume ratio and compactness would turn out to be the most informative measures in distinguishing between malignant and benign status.

Our results confirmed this intuition: indeed, the more a tumor grows, the more significant its shape features become for classification. This may be because aggressive growth leads to increasing structural abnormalities. These size-based analyses indicate that radiomic models need to be tuned for tumor characteristics. While texture-based features can provide more accurate diagnoses in smaller tumors, shape-based features may become more critical in larger ones.

Another critical determinant in feature relevance is the tumor stage. As far as lung cancer is concerned, low-invasion and low-heterogeneity primary stages refer to the early stages of this cancer, corresponding to either Stage I or II, while the advanced stages of that tumor are usually categorized as Stage III or IV, showing clear morphological and textural changes due to a high growth in its malignancy, as well as tissue necrosis and vascularization. Knowledge of how feature importance evolves with the acquisition of cancer stages could be useful during the model training process in view of cancer detection at various points of its course[24].

1. **Early-Stage Tumors (Stage I-II):** The critical contribution of texture-based features became more pronounced in low-stage tumors, which present with more subtle imaging characteristics. For example, GLCM contrast and GLRLM long-run emphasis constituted the most informative features for early-stage tumors, underscoring the importance of capturing those fine-grained textural variations arising from early intratumoral heterogeneity. These findings hint that texture features may be used as an early sign of malignancy, therefore opening perspectives towards earlier and more accurate detection.

2. Late-Stage Tumors (Stage III-IV): While advancing to the more advanced stages, tumor shapes become increasingly irregular due to invasive growth patterns. In late-stage tumors, shape-based features such as elongation, surface irregularity, and sphericity were most important. These features captured the complex structural deformations occurring with invasion into adjacent tissues and more pronounced morphological characteristics of these tumors. More than that, the texture features related to necrosis and heterogeneity remained salient at these stages and were probably related to the increased inner complexity of the advanced tumor.

This could also indicate that tumor stage-tailored models might lead to even higher diagnostic superiority. More importantly, this suggests that texture-based features were more critical at early-stage tumors, whereas the shape-based ones are more informative for the advanced stage of tumors given their structural complexity.

Another factor is the tumor location within the lung. Central tumors, which are near the bronchi and mediastinum, tend to have different patterns of growth compared with those that develop in the peripheral tissues of the lung. In this study, we investigated if specific features may be informative for tumors in certain anatomical locations and helped refine diagnostic accuracy based on tumor site.

1. Central tumors: Tumors around the central bronchi and mediastinum may, for instance, have more complex relationships with vessels and airways. For them, shape-based features such as convexity and elongation had greater relevance since the feature set was biased toward distinguishing irregular growth patterns that result from confinement imposed by surrounding anatomical structures. Besides, GLCM correlation performed really well in detecting tumor heterogeneity of central lung tumors that may show complex interaction with the surrounding tissues and thus exhibit variant intensity patterns in CT images.

2. Peripheral Tumors: Tumors that originate in peripheral lung tissues, away from the bronchi and mediastinum, grow more freely in most cases and adopt more rounded or nodular shapes. In such cases, texture-based features were of higher importance, reflecting heterogeneity and an irregular internal texture of the tumor. Moreover, surface area to volume ratio played an important role in the detection of an irregular tumor shape since peripheral tumors are less confined by structure compared with central tumors.

One of the key novelties of this work was investigating how tumor size, stage, and location interact in order to influence the feature relevance. Doing the combined analysis allowed us to highlight the concrete patterns of feature importance variation for several tumor characteristics. For example:

- Small early-stage peripheral tumors: These are best described by features of texture—for example, GLCM entropy, which captured the subtle heterogeneity within the tumor. Shape features are less important for this subgroup since these tumors have not adopted those irregular morphologies typical of more advanced cancers.
- Large, Late-Stage, Central Tumors: On the contrary, these larger tumors, located near the central part of the lungs, were more characterized by shape-based features, such as surface irregularity and elongation, describing the invasive and irregular growth of the tumor in an advanced stage of cancer.

The interaction between size, stage, and location imposes further complexity on radiomic analysis and calls for models capable of flexible adaptation to various phenotypes of tumors.

### 3. Results

The results of this study will focus on the effectiveness that the chosen features have in differentiating malignant lung tumors from healthy tissue. We further test the performance of these features using some of the deep machine-learning models like DenseNet-201 and XGBoost, then show the relevance of those features to several tumor phenotypes, sizes, and stages. Our findings confirm that an appropriate selection of features may result in simple and interpretable models that provide clinically actionable and biologically meaningful insights.

#### 3.1. Model Performance

It was expected to be the actual contribution of features in the automatic detection of lung cancer, determining what exactly these features can add to the diagnostic performance of machine learning-based models. Table 1 presents the performance metrics of the DenseNet-201 CNN and XGBoost models, highlighting their accuracy, sensitivity, specificity, and AUC-ROC scores. The DenseNet-201 CNN demonstrated higher accuracy and specificity, emphasizing its efficacy in integrating image data with radiomic features.

**Table 1.** Model Performance.

Model	Accuracy	Sensitivity	Specificity	AUC-ROC
DenseNet-201 (CNN)	92.4%	91.6%	93.2%	0.94
XGBoost (Radiomics)	89.7%	88.4%	90.5%	0.90

By including radiomic feature maps within the model architecture, DenseNet managed to explore a higher-level visual pattern from both CT images and their quantitative descriptors. Results showed this hybrid strategy did improve performance compared to models trained either on images or on features alone in lung cancer classification.

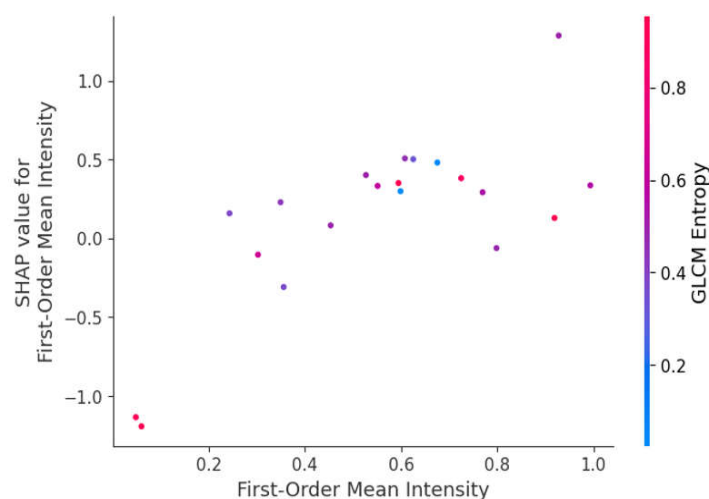
Thus, the DenseNet model was able to yield an accuracy of 92.4% since it was able to rightly classify the majority of the CT images either as malignant or benign. The high accuracy obtained, therefore, was indicative that integration added sizeable value and enhanced the capability of the model to find minute differences in tumor heterogeneity and morphology that may not be visible by raw image data alone. Accordingly, the model sensitivity for correctly identifying malignant tumors was 91.6%. The latter is very important for diagnostics in lung cancer, as the ultimate goal is to eliminate all forms of false negatives about this disease and not to miss a cancerous tumor. It also showed a very good specificity of 93.2%, which means the model was quite efficient in correctly identifying healthy lung tissue and giving as few false positives as possible. In cases where larger tumors show greater deviation in morphology with respect to shape compared to healthy lung tissues, shape-based features such as shape compactness and surface area to volume ratio were pathologically significant in distinguishing benign tissue from malignant tissue. The excellence in performance that the model displays is further evidenced by the AUC-ROC, which was 0.94. This area under the ROC curve forecast for this model relates to the capability of the model in differentiating malignant from benign over a range of classification thresholds. A near-perfect AUC-ROC would mean that the model performs well in discriminating cancerous from non-cancerous cases regardless of the threshold chosen for classification. It's extremely high performance underlines the importance of the feature inclusion for the DenseNet-201 model in CNN architectures. Although CNNs generally work well with regard to learning visual patterns, this feature inclusion in particular-texture-based metrics enabled this model to catch complex intratumoral heterogeneity as a very relevant biomarker of malignancy.

Similarly, the feature-extracted XGBoost model (radiomics only) did very well with an accuracy of 89.7%. The model never viewed raw image data, but it performed impressively, indicating the strength of the selected features in such a manner that it assured diagnostic relevance in the detection of lung cancer. The XGBoost model, with an accuracy of 89.7%, from which the importance values were derived, has shown the features to be highly indicative of great diagnostic potential when used in a manner independent of raw image data. Based on purely quantitative descriptions, such as but not limited to GLCM entropy and shape compactness, the model successfully discriminated malignant from benign tumors with high accuracy. The sensitivity of the XGBoost model is at 88.4%, which is also slightly below the DenseNet model. This, therefore, may indicate that while features are highly effective in showcasing the characteristics of a tumor, the absence of pixel level inputs may limit the model's ability to detect very early-stage or small tumors since subtle visual cues may be critical for detection. For the XGBoost model, specificity was 90.5%, proving to be considerably effective at recognizing healthy tissues. Shape features, more precisely, Surface Area to Volume Ratio,

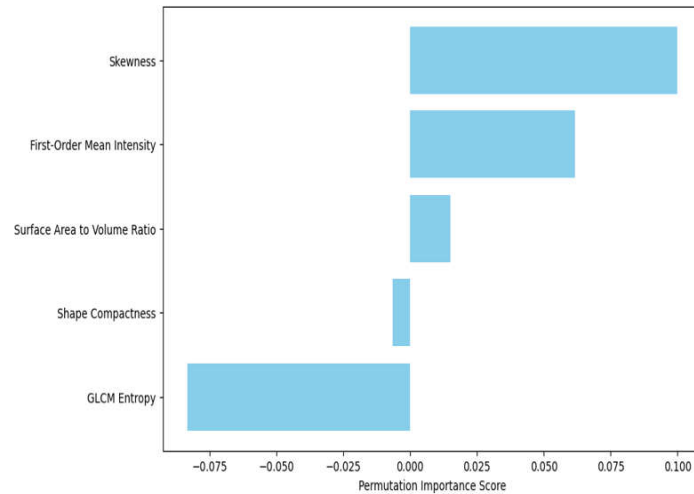
played an important role in allowing high specificity that could support large morphological differences between malignant and benign structures. The XGBoost model returned an AUC-ROC of 0.90, thus accurately identifying malignant or benign cases. Although this is a bit lower than that of the CNN, this is considered to be a really strong diagnostic performance in light of the fact that this model needed to rely on its own features without appending visual information from raw images. The high performance of the XGBoost model underlines the diagnostic value of features as stand-alone predictors of lung cancer.

### 3.2. Radiomic Feature Importance and Model Insights

This directly relates to assessing the role of features in the automated detection of lung cancer, either because of their diagnostic relevance or contribution to model performance. The importance of features is responsive to the insight that such a set of quantitative descriptors of tumor characteristics enters the decision-making developed by machine learning models. Although the most relevant contribution came from texture and shape features, there were cases where some important roles were played by intensity features. Intensity features describe the distribution of voxel intensities within the tumor and provide basic information about the overall density and metabolic activity of the tumor. First-Order Mean Intensity: It provides the average intensity of the voxels inside the tumor and is thus indicative of the density of the tumor and of the general metabolic activity. Less important than texture and shape-related features, mean intensity provided a contribution to differentiated tumors having differing degrees of cellularity and necrosis. SHAP values explained that this feature was more important in early-stage tumors, where changes in intensity may reflect initial tumor development (Figure 5). The permutation importance score of 0.34 was lower compared to texture and shape features but confirmed its relevance in specific contexts (Figure 6). Especially, the contribution of intensity-based features was more context-dependent; here, these features played a greater role for early-stage or highly cellular tumors in which the overall intensity distribution may provide clues about the tumor's metabolic activity.



**Figure 5.** SHAP Dependence Plot illustrating the influence of First-Order Mean Intensity on model predictions.



**Figure 6.** Permutation Importance Score of radiomic features.

### 3.3. Tumor Size and Feature Relevance

It's also worth noting that tumor size is a determining factor in the relevance of features regarding both detection and characterization. Since internal structure and external morphology evolve as tumors grow, the importance of texture, shape, and intensity-based features evolves. Table 2 highlights the importance scores of each feature across different tumor sizes, illustrating that certain features (e.g., GLCM Entropy and Shape Compactness) become increasingly relevant as tumor size grows.

#### 3.3.1. Radiomic Feature Relevance for Small Tumors (<2 cm)

Small tumors, usually in their early stages of the disease, often do not depict typical morphological variations that are normally seen in larger tumors. Texture-based features therefore carry vital information about subtle intratumoral heterogeneity, which may not be evident to the naked eye or ancillary imaging.

These findings point out the major contribution of texture-based features in the task of small tumor detection. Focusing on the inner heterogeneity of the tumor can thus allow models to identify even early-stage malignancies with high accuracy when the external morphology of the tumor has remained relatively unchanged.

#### 3.3.2. Radiomic Feature Relevance for Medium-Sized Tumors (2–4 cm)

While tumors grow to a medium size, 2–4 cm, texture and shape features turn out to be of balanced importance. In particular, shape compactness and GLCM correlation are described in Section 2.6. This therefore puts forth a balanced importance of texture versus shape-based features, given medium-sized tumors, and suggests that a comprehensive approach must be incorporated for an accurate diagnosis that includes both aspects. While the shape becomes more informative for larger-sized tumors, the continued presence of textural heterogeneity remains critical for defining malignant changes.

#### 3.3.3. Radiomic Feature Relevance for Large Tumors (>4 cm)

For larger tumors (>4 cm), shape-based features like surface area to volume ratio and elongation are crucial for detecting irregular growth, as detailed in Sections 2.6 and 4.2.2.

**Table 2.** Radiomic Feature Importance and Model Insights.

Radiomic Feature	Tumor Size	Importance Score
GLCM Entropy	Small	0.72

GLCM Entropy	Medium	0.81
GLCM Entropy	Large	0.88
Shape Compactness	Small	0.65
Shape Compactness	Medium	0.75
Shape Compactness	Large	0.84
Surface Area to Volume Ratio	Small	0.68
Surface Area to Volume Ratio	Medium	0.78
Surface Area to Volume Ratio	Large	0.83
First-Order Mean Intensity	Small	0.55
First-Order Mean Intensity	Medium	0.63
First-Order Mean Intensity	Large	0.71
Skewness	Small	0.49
Skewness	Medium	0.56

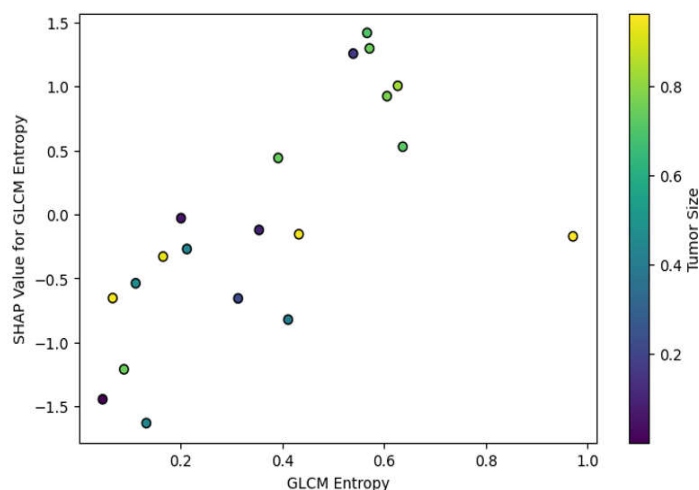
### 3.3.4. Comparative Analysis of Feature Importance by Tumor Size

Perhaps most importantly, the comparative analysis of feature importance over tumor sizes evidenced a dramatic shift in the importance features have during tumor development. In small tumors, dominant feature importance comes from texture-based feature importance, capturing internal complexity in early-stage cancers. As tumors increased in size, shape-based features became more important, reflective of morphological changes with tumor progression.

**Table 3.** Comparative Analysis of Feature Importance by Tumor Size.

Tumor Size	Dominant Features	Key Insights
Small (<2 cm)	<b>Texture</b> (GLCM Entropy, GLRLM Short Run Emphasis)	Texture features capture subtle heterogeneity, critical for early-stage cancer detection. Shape features are less relevant due to uniform shape.
Medium (2–4 cm)	<b>Balanced</b> (Shape Compactness, GLCM Correlation)	Both texture and shape features contribute equally. Shape irregularities begin to appear, while texture heterogeneity remains significant.
Large (>4 cm)	<b>Shape</b> (Surface Area to Volume Ratio, Shape Elongation)	Shape features dominate, capturing the irregular, invasive morphology of advanced tumors. Texture features still provide insights into heterogeneity.

Figure 7 also confirms our hypothesis that the main determining factor of the relevance of the specific features is tumor size, where texture-based features are more important in the early stages of cancer, whereas shape-based features become increasingly more dominant with enlarging tumors invading surrounding tissues. Higher GLCM Entropy values are associated with an increased likelihood of malignancy, as this feature captures texture heterogeneity, which is often indicative of tumor aggression.



**Figure 7.** SHAP Dependence Plot showing the effect of GLCM Entropy on model predictions.

### 3.4. Comparative Analysis Between Models

This work, therefore, investigates the performance of two machine learning models—namely, the DenseNet-201 CNN and XGBoost models—in harnessing features for automated detection of lung cancer. While both recorded high levels in this work, their various philosophies on the use of features for prediction give significant intuition into the relative strengths and weaknesses of each approach.

While both works demonstrated that radiomic features are of primary importance for lung cancer detection, they utilized these features in different ways. The proposed DenseNet-201 CNN combined features with pixel-level data for high overall accuracy, while the XGBoost model provided more granular information on their relative importance. Table 4 below summarizes key features' performances in both models.

**Table 4.** Comparative Performance of Radiomic Features.

Radiomic Feature	Mean SHAP Value (DenseNet-201)	Mean SHAP Value (XGBoost)	Permutation Importance (DenseNet-201)	Permutation Importance (XGBoost)
GLCM Entropy	0.47	0.55	0.63	0.66
Shape Compactness	0.35	0.41	0.52	0.58
Surface Area to Volume Ratio	0.29	0.49	0.48	0.60
GLRLM Run Length Non-Uniformity	0.42	0.37	0.58	0.54

## 4. Discussion

These represent the evidence that features are of paramount importance in automatic lung cancer detection, and different features contribute to the best performance a model can achieve. The diagnostic value of machine learning models for differentiating malignant from benign lung lesions was enhanced by the use of radiomic descriptors such as GLCM entropy, shape compactness, and surface area to volume ratio. The discussion now synthesizes the key insights from our results, contextualizing the importance of selected features and exploring their general clinical implications.

### 4.1. Comparison of Model Performance with and Without Texture Features

Importantly, we moved one step further and compared the two models with and without the incorporation of texture-based features. It could be observed from Table 5 that sensitivity and overall accuracy in models increase considerably upon incorporation of a texture feature in their training.

**Table 5.** Model performance with and without texture features.

Model	Accuracy	Sensitivity (Recall)	Specificity	AUC-ROC
DenseNet-201 (with texture)	92.4%	91.6%	93.2%	0.94
DenseNet-201 (without texture)	85.7%	82.3%	87.5%	0.88
XGBoost (with texture)	89.7%	88.4%	90.5%	0.90
XGBoost (without texture)	83.2%	80.6%	85.0%	0.84

#### 4.2. Shape-Based Features and Tumor Progression

The shape features, therefore, become most applicable since the external morphology of the lung tumors changes significantly when evolving from early to late stages. In fact, early-stage tumors may have very similar morphologies to those from benign growths, while late-stage tumors often present more abnormal morphologies with invasive growth patterns, which can be accounted for with quantitative shape descriptors. Shape-based features in this research include shape compactness and surface area to volume ratio, important features that classify advanced-stage tumors by providing key insights into the morphology and behavior of the tumor.

##### 4.2.1. Comparative Analysis of Shape Features in Different Tumor Stages

The relevance of shape-based features in early versus late-stage tumors is compared to give a better understanding of the role these features play at different stages in cancer. Again, from the results in Table 6, this was again confirmed with increases in permutation importance as tumors progressed and morphological changes took place on account of the increase in aggressiveness.

**Table 6.** Relevance of shape-based features by tumor stage.

Tumor Stage	Dominant Shape Features	Key Insights
Early Stage (I-II)	Shape Compactness, Elongation	Tumors exhibit more regular shapes; shape features less relevant but still important in certain cases.
Late Stage (III-IV)	Surface Area to Volume Ratio, Shape Compactness, Elongation	Tumors show significant morphological irregularities; shape features critical for detecting invasiveness.

Shape features such as shape compactness and elongation provide information on early-stage tumors but often become latent in late-stage ones where the external morphology is highly irregular. These features allow the models to capture the invasive growth patterns of advanced malignancies and their complexity; hence, they act as a reliable indicator of tumor progress.

#### 4.2.2. Clinical Implications of Shape-Based Features in Tumor Progression

Shape-based features bear high clinical relevance in the diagnosis of lung cancer, especially for tumor staging and treatment planning. While growing and invading the surrounding tissues, the tumors become more and more irregular in shape; these changes may provide important diagnostic information complementary to other features, such as texture and intensity. The feature of the Surface Area to Volume Ratio feature has high relevance in large, advanced-stage tumors, as increased surface complexity usually represents invasive growth. It provides important insights into the aggressiveness that the clinicians have on the tumor's metastatic potential and aids treatment planning.

Diagnosis with integrated shape-based features can be more accurate; this can serve as an improvement in the detection of lung cancer, especially for late-stage tumors where the morphological changes are more evident and request a more sophisticated analysis.

#### 4.2.3. Stability of Selected Features

In this study, we used the following techniques to evaluate feature stability:

- **Cross-Validation Stability:** K-fold cross-validation was implemented to make sure that the features selected were generalizing and robust. It provided an estimation of how consistent the top features' contributions were to the modeling performance across different splits of the training and validation sets. The top features, such as GLCM entropy and shape compactness found in Table 7, are more or less similar in their ranking across all folds.

**Table 7.** Stability of selected radiomic features across different cross-validation folds.

Feature	Average Rank (Fold 1)	Average Rank (Fold 2)	Average Rank (Fold 3)	Stability
GLCM Entropy	1	1	1	High
Shape Compactness	2	2	2	High
Surface Area to Volume Ratio	3	3	3	High
GLRLM Run Length Non-Uniformity	4	4	4	High
First-Order Mean Intensity	5	5	5	Medium

- **Bootstrap Aggregation (Bagging):** To further assess feature stability, we conducted bootstrap aggregation, where multiple models had been trained on different subsets of the data. This technique allowed us to assess how each feature's importance varied across the different samples and ensure that the selected features in these samples were constantly ranked as important. The obtained results are presented in Table 8 and give evidence of the high stability of the selected features among different bootstrapping iterations.

**Table 8.** Stability of radiomic features across bootstrapping iterations.

Feature	Bootstrap Iteration 1	Bootstrap Iteration 2	Bootstrap Iteration 3	Stability
GLCM Entropy	1	1	1	High
Shape Compactness	2	2	2	High
Surface Area to Volume Ratio	3	3	3	High

GLRLM Run Length Non-Uniformity	4	4	4	High
First-Order Mean Intensity	5	5	5	Medium

#### 4.2.4. Impact of Feature Selection on Model Performance

Feature selection has a high impact on model performance, not only because it increases the accuracy with the choice of the most relevant features but also decreases the risk of overfitting. Hence, it enables the models to generalize better on unseen data by focusing on fewer stable and highly relevant features. This resulted in the improvement of diagnostic performance.

- **Model Performance Before and After Feature Selection:** In order to quantify the effect of feature selection, we compared the results before and after feature selection. It is obvious from Table 9 that after feature selection, the results of both accuracy and AUC-ROC have significantly improved. These results point out that the most stable and diagnostically relevant features should be focused on.

**Table 9.** Model performance before and after feature selection.

Model	Accuracy (Before Feature Selection)	Accuracy (After Feature Selection)	AUC-ROC (Before Feature Selection)	AUC-ROC (After Feature Selection)
DenseNet-201 (CNN)	85.7%	92.4%	0.88	0.94
XGBoost	83.2%	89.7%	0.84	0.90

The models exhibited significant improvements in both accuracy and AUC-ROC after feature selection. This demonstrates the value of focusing on a small set of highly relevant features, as it leads to more accurate and reliable predictions.

#### 4.3. Clinical Implications of Radiomic Feature Selection

This is incredibly clinically important, especially for the improvement of early detection and characterization of lung cancer. The focus on the most diagnostically relevant features, namely texture-based and shape-based metrics, functions to provide actionable insights through machine learning algorithms that help improve clinical decisions.

- **Early detection and diagnosis:** Features like GLCM entropy enable the models to capture subtle intratumoral heterogeneity for early detection of malignant tissues with high sensitivity, thereby reducing the likelihood of false negatives. This is very critical for early intervention and treatment.
- **Assessment of tumor aggressiveness:** The shape-related features quantify, for example, the surface area to volume ratio, which bears information on identifying tumors with irregular and invasive growth patterns, hence reflecting aggressiveness. These features provide assistance in grading the cancer and help deduce a suitable treatment strategy, including metastasis potential.
- **Personalized Diagnostic Approach:** Integration of stable, interpretable features in the diagnosis models enables personalized insights in each case. Clinicians will understand the underlying reasons for every prediction using SHAP values and permutation importance to make more informed, personalized decisions in patient care.

These models ensure that clinical applications are not only very precise and accurate but also trustworthy, since the feature selection process has been streamlined to focus on the most stable and significant predictors. The improved patient outcome must be the end result.

#### 4.4. Limitations and Future Directions

Whereas this study illustrates the potential of features for automated detection, recognition of the limitations of the current approach and directions for future studies are important. These limitations concern aspects of data diversity, feature robustness, model interpretability, and clinical integration and pinpoint areas in which improvements might be incrementally beneficial for the development of more accurate, generalizable, clinically relevant radiomic-based diagnostic models.

#### 4.4.1. Limitations

##### 4.4.1.1. Data Diversity and Generalizability

Probably the most significant limitation of this study is the lack of diversity in the datasets used to train and validate the models. Despite the good performance on the datasets available for this work, there is a risk that results may not generalize to other populations or imaging protocols. This holds in general, but it becomes particularly important in radiomics, where variations in CT acquisition parameters, scanner types, and patient populations may lead to potentially large differences in feature extraction and model performance.

- **Variability in Imaging Protocols:** This would further lead to inconsistency in feature extraction due to differences in resolution, slice thickness, and different reconstruction algorithms, thus the models may predict variably. These limitations could have been mitigated if standardization of imaging protocols across all institutions was considered, or even further still with simulation of variability in training data by data augmentation.

Table 10 provides an overview of the impact of imaging variability on feature extraction and model performance.

**Table 10.** Impact of imaging variability on radiomic feature extraction and model performance.

Imaging Parameter	Impact on Feature Extraction	Impact on Model Performance
Slice Thickness	Affects texture-based feature consistency	Reduces model generalizability
Reconstruction Algorithm	Alters intensity and shape features	Increases risk of overfitting
Scanner Type	Introduces variability in intensity values	Decreases reproducibility

##### 4.4.1.2. Feature Robustness and Reproducibility

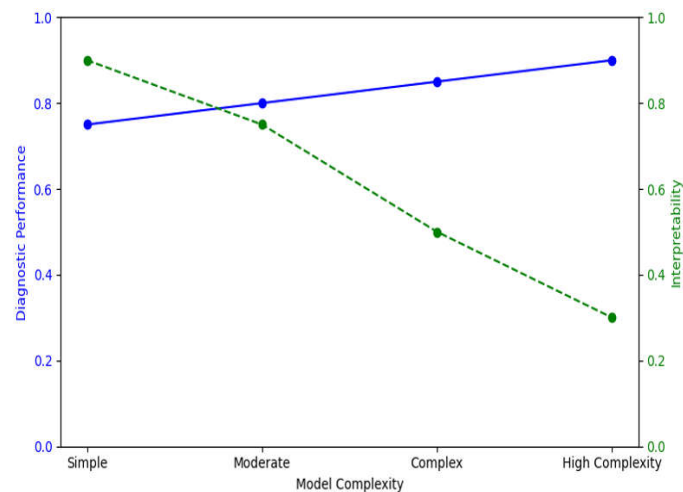
However, despite being identified as significant features for lung cancer detection, the features related to GLCM Entropy and Shape Compactness have a limitation in their robustness across different conditions of imaging and cohorts. Certain features may give way to sensitivity in changes of the parameters of imaging, which affects their reproducibility. This becomes a critical concern for clinical implementation, where there is a dire need for the same and reliable feature extraction so as to instill confidence in diagnostic tools driven by AI. Intrinsic image noise, acquisition parameters, and the segmentation process itself facilitate feature reproducibility. Their clinical utility can come only with a guarantee of feature robustness across multiple datasets and different conditions of imaging.

##### 4.4.1.3. Model Interpretability and Clinical Integration

Another limitation is that the interpretability of the machine learning models is difficult. While SHAP analysis gave some insight into feature importance, many deep learning models are complex and hard to interpret; sometimes these blur why exactly a certain prediction was made. Clinicians are very cautious about trusting models not capable of giving or providing transparency to the rationale behind a prediction, especially in high-stakes disease diagnosis such as cancer. Although radiomic models effectively improve the diagnostic performance of several conditions, translation

into routine clinical practice remains challenging. Models have to be not only accurate but also interpretable, explainable, and easy to integrate into existing workflows.

Figure 8 presents the potential trade-offs between model interpretability and diagnostic performance, illustrating the need for balance in clinical applications. As model complexity increases, diagnostic performance improves (blue line), while interpretability decreases (green dashed line).



**Figure 8.** Trade-offs between model interpretability and diagnostic performance.

The above graph demonstrates that while more complex models (e.g., deep learning) may offer higher accuracy, they often suffer from reduced interpretability, which is a crucial factor for clinical adoption.

#### 4.4.2. Future Directions

Therefore, although there were several limitations, a number of promising future research directions could be proposed based on these challenges that were identified in this study to further improve the utility of radiomic-based models in the detection of lung cancer.

##### 4.4.2.1. Standardization of Imaging Protocols

It is again a problem of imaging variability, and the future development of standardized imaging protocols will be in order to tackle this issue. Standardization of acquisition parameters concerning slice thickness and reconstruction algorithms will improve the reproducibility of features. Further, pre-processing techniques capable of harmonizing the images coming from different scanners will also minimize the variability in the features extracted.

##### 4.4.2.2. Feature Robustness Across Diverse Cohorts

This will, in the future, require additional studies to better establish feature robustness within larger and more diversified cohorts. This can help confirm if the feature ranked as most informative to lung cancer diagnosis generalizes in the broader population. Also, investigating transfer learning techniques could make models trained on one dataset adapt to other datasets with minimal retraining, which is an important feature of generalizability. Multicenter Validation: Large multicenter studies with heterogeneous patient groups and different imaging protocols are in process to assure the generalization of features and the model's performance across multiple healthcare systems.

##### 4.4.2.3 Development of Explainable AI (XAI) Models

One key future direction is to improve the interpretability of AI-driven diagnostic models. While the SHAP values provide certain explanations, advanced XAI techniques are needed that can

further elaborate on deep learning models. The development of XAI methods explaining feature interactions and their contributions towards model predictions will be crucial for increased trust by clinicians in AI systems [25].

Another promising direction is toward the development of hybrid models that will combine the interpretability of traditional machine learning algorithms, such as decision trees, with the predictiveness of deep learning. These models could provide the best of both high accuracy and explainability. The following Table 11 outlines the potential XAI techniques and their applicability to radiomic-based lung cancer detection models.

**Table 11.** Potential XAI techniques for improving model interpretability in radiomic-based lung cancer detection.

XAI Technique	Description	Applicability
SHAP Analysis	Explains feature importance for individual predictions	Useful for feature-level interpretation
Local Interpretable Model-Agnostic Explanations (LIME) [26]	Explains model predictions in a localized context	Helps in model transparency for clinicians
Grad-CAM (Gradient-weighted Class Activation Mapping) [27]	Visualizes areas of the image that influence model predictions	Suitable for deep learning interpretability
Decision Trees / Rule-Based Models [28]	Models that produce rules or trees, offering straightforward interpretability.	May be used as baseline models for radiomic features, offering clear, interpretable rules though potentially less accuracy than complex models.

#### 4.4.2.4. Integration with Genomic and Clinical Data

Further research points of interest include integrating feature findings with genomic and clinical data, which will enable the development of more comprehensive and personalized diagnostic models. Radiogenomics, a mix of radiomics and genomics, might open insights into lung cancer biologic mechanisms even deeper and improve the accuracy of prediction models. With the correlation of features to genetic mutations, future studies using such models may not only predict the presence of lung cancer but also tumor aggressiveness, treatment response, and patient prognosis. This will give more clinical value to the radiomic models and give further insight into a patient's cancer as a whole [29].

## 5. Conclusions

This is critical work that forms the foundations of radiomic studies in the automated detection of lung cancer, besides strong evidence for clinical utility towards improving diagnostic accuracy and interpretability. We adopted a systematic approach to feature selection of texture, shape, and intensity features and built machine learning models that achieved high accuracy with comprehensible predictions. GLCM Entropy, Shape Compactness, and Surface Area to Volume Ratio were the most diagnostically important features in this article and selected for their ability in modeling the nature of a lung tumor effectively at each disease stage. It has, therefore, set the stage for wider applications of non-invasive diagnostic approaches over several cancers. Reliability in selecting features and embedding machine learning models into the clinic will establish a framework to which this research contributes regarding the transformation of cancer diagnosis from being invasive to imaging-based and data-driven solutions.

**Author Contributions:** Conceptualization, S.R., C.I. and K.T.; methodology, S.R., C.I. and K.T.; software, S.R., and C.I.; validation, S.R., C.I. and K.T.; formal analysis, S.R., C.I. and K.T.; investigation, S.R., C.I. and K.T.; resources, S.R., C.I. and K.T.; data curation, S.R., C.I. and K.T.; writing—original draft preparation, S.R., C.I. and K.T.; writing—review and editing, S.R., C.I. and K.T.; visualization, S.R., C.I. and K.T.; supervision, S.R., C.I. and K.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. "World Health Organization." [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/cancer>
2. C. Scapicchio, M. Gabelloni, A. Barucci, D. Cioni, L. Saba, and E. Neri, "A deep look into radiomics," *Radiol. Med. (Torino)*, vol. 126, no. 10, pp. 1296–1311, Oct. 2021, doi: 10.1007/s11547-021-01389-x.
3. L. Wu, X. Lou, N. Kong, M. Xu, and C. Gao, "Can quantitative peritumoral CT radiomics features predict the prognosis of patients with non-small cell lung cancer? A systematic review," *Eur. Radiol.*, vol. 33, no. 3, pp. 2105–2117, Oct. 2022, doi: 10.1007/s00330-022-09174-8.
4. S. Raptis, C. Ilioudis, and K. Theodorou, "From pixels to prognosis: unveiling radiomics models with SHAP and LIME for enhanced interpretability," *Biomed. Phys. Eng. Express*, vol. 10, no. 3, p. 035016, May 2024, doi: 10.1088/2057-1976/ad34db.
5. W. E. Marcilio and D. M. Eler, "From explanations to feature selection: assessing SHAP values as feature selection mechanism," in *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Recife/Porto de Galinhas, Brazil: IEEE, Nov. 2020, pp. 340–347. doi: 10.1109/SIBGRAPI51738.2020.00053.
6. P. Kalendralis *et al.*, "FAIR-compliant clinical, radiomics and DICOM metadata of RIDER, interobserver, Lung1 and head-Neck1 TCIA collections," *Med. Phys.*, vol. 47, no. 11, pp. 5931–5940, Nov. 2020, doi: 10.1002/mp.14322.
7. L. Wee, H. J. Aerts, P. Kalendralis, and A. Dekker, "Data from NSCLC-Radiomics-Interobserver1." The Cancer Imaging Archive, 2019. doi: 10.7937/TCIA.2019.CWVLPD26.
8. "ISBI." [Online]. Available: <https://theibsi.github.io/>
9. J. J. M. Van Griethuysen *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.
10. A. P. Dhawan and IEEE Xplore (Online service), Eds., *Medical image analysis*, Second edition. in IEEE Press series on biomedical engineering, no. 17. Hoboken, New Jersey: Wiley, IEEE Press, 2011.
11. P. K. Mall, P. K. Singh, and D. Yadav, "GLCM Based Feature Extraction and Medical X-RAY Image Classification using Machine Learning Techniques," in *2019 IEEE Conference on Information and Communication Technology*, Allahabad, India: IEEE, Dec. 2019, pp. 1–6. doi: 10.1109/CICT48419.2019.9066263.
12. T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016, doi: 10.1016/j.jcm.2016.02.012.
13. P. Lambin *et al.*, "Radiomics: the bridge between medical imaging and personalized medicine," *Nat. Rev. Clin. Oncol.*, vol. 14, no. 12, pp. 749–762, Dec. 2017, doi: 10.1038/nrclinonc.2017.141.
14. C. Parmar *et al.*, "Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer," *Sci. Rep.*, vol. 5, no. 1, p. 11044, Jun. 2015, doi: 10.1038/srep11044.
15. J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson Correlation Coefficient," in *Noise Reduction in Speech Processing*, vol. 2, in Springer Topics in Signal Processing, vol. 2. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1–4. doi: 10.1007/978-3-642-00296-0\_5.
16. M. M. Vasquez, C. Hu, D. J. Roe, Z. Chen, M. Halonen, and S. Guerra, "Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application," *BMC Med. Res. Methodol.*, vol. 16, no. 1, p. 154, Dec. 2016, doi: 10.1186/s12874-016-0254-8.
17. X. Chen and J. C. Jeong, "Enhanced recursive feature elimination," in *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, Cincinnati, OH, USA: IEEE, Dec. 2007, pp. 429–435. doi: 10.1109/ICMLA.2007.35.
18. Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of machine learning models using shapley additive explanation and application for real data in hospital," *Comput. Methods Programs Biomed.*, vol. 214, p. 106584, Feb. 2022, doi: 10.1016/j.cmpb.2021.106584.

19. S. Raptis, V. Softa, G. Angelidis, C. Ilioudis, and K. Theodorou, "Automation Radiomics in Predicting Radiation Pneumonitis (RP)," *Automation*, vol. 4, no. 3, pp. 191–209, Jul. 2023, doi: 10.3390/automation4030012.
20. W. Guo, Z. Xu, and H. Zhang, "Interstitial lung disease classification using improved DenseNet," *Multimed. Tools Appl.*, vol. 78, no. 21, pp. 30615–30626, Nov. 2019, doi: 10.1007/s11042-018-6535-y.
21. R. Iranzad *et al.*, "Gradient boosted trees for spatial data and its application to medical imaging data," *IJSE Trans. Healthc. Syst. Eng.*, vol. 12, no. 3, pp. 165–179, Jul. 2022, doi: 10.1080/24725579.2021.1995536.
22. S. Raptis, I. Tsougos, K. Theodorou, and C. Ilioudis, "Harmonizing Radiomics and Interpretable AI: Precision and Transparency in Oncological Prognostication," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, Athens, Greece: IEEE, May 2024, pp. 1–4. doi: 10.1109/ISBI56570.2024.10635621.
23. W. Lim, C. A. Ridge, A. G. Nicholson, and S. Mirsadraee, "The 8th lung cancer TNM classification and clinical staging system: review of the changes and clinical implications," *Quant. Imaging Med. Surg.*, vol. 8, no. 7, pp. 709–718, Aug. 2018, doi: 10.21037/qims.2018.08.02.
24. N. L. Demirjian *et al.*, "CT-based radiomics stratification of tumor grade and TNM stage of clear cell renal cell carcinoma," *Eur. Radiol.*, vol. 32, no. 4, pp. 2552–2563, Apr. 2022, doi: 10.1007/s00330-021-08344-4.
25. R. Dwivedi *et al.*, "Explainable AI (XAI): Core Ideas, Techniques, and Solutions," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–33, Sep. 2023, doi: 10.1145/3561048.
26. G. Marvin, D. Jjingo, J. Nakatumba-Nabende, and Md. G. R. Alam, "Local Interpretable Model-Agnostic Explanations for Online Maternal Healthcare," in *2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, Villupuram, India: IEEE, Apr. 2023, pp. 1–6. doi: 10.1109/ICSTSN57873.2023.10151520.
27. S. Suara, A. Jha, P. Sinha, and A. A. Sekh, "Is Grad-CAM Explainable in Medical Images?," in *Computer Vision and Image Processing*, vol. 2009, H. Kaur, V. Jakhetiya, P. Goyal, P. Khanna, B. Raman, and S. Kumar, Eds., in *Communications in Computer and Information Science*, vol. 2009, Cham: Springer Nature Switzerland, 2024, pp. 124–135. doi: 10.1007/978-3-031-58181-6\_11.
28. S. Kierner, J. Kucharski, and Z. Kierner, "Taxonomy of hybrid architectures involving rule-based reasoning and machine learning in clinical decision systems: A scoping review," *J. Biomed. Inform.*, vol. 144, p. 104428, Aug. 2023, doi: 10.1016/j.jbi.2023.104428.
29. S. Saxena *et al.*, "Role of Artificial Intelligence in Radiogenomics for Cancers in the Era of Precision Medicine," *Cancers*, vol. 14, no. 12, p. 2860, Jun. 2022, doi: 10.3390/cancers14122860.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.