




## Article

# Impact of Voxel Normalization on a Machine Learning-Based Method: A Study on Pulmonary Nodule Malignancy Diagnosis Using Low-Dose Computed Tomography (LDCT)

Chia-Chi Hsiao <sup>1</sup>, Chen-Hao Peng <sup>2</sup> , Fu-Zong Wu <sup>1,\*</sup>, <sup>†</sup>  and Da-Chuan Cheng <sup>2,\*</sup>, <sup>†</sup> 

<sup>1</sup> Department of Radiology, Kaohsiung Veterans General Hospital, Kaohsiung 813414, Taiwan; vghkscchsiao1@gmail.com

<sup>2</sup> Department of Biomedical Imaging and Radiological Science, China Medical University, Taichung 40400, Taiwan; 112202701@365.cmu.edu.tw

\* Correspondence: cmvwu1029@gmail.com (F.-Z.W.); dccheng@mail.cmu.edu.tw (D.-C.C.)

<sup>†</sup> These authors contributed equally to this work and should be considered co-corresponding authors.

**Abstract:** Lung cancer (LC) stands as the foremost cause of cancer-related fatality rates worldwide. Early diagnosis significantly enhances patient survival rate. Nowadays, low-dose computed tomography (LDCT) is widely employed on the chest as a tool for large-scale lung cancer screening. Nonetheless, a large amount of chest radiographs creates an onerous burden for radiologists. Some computer-aided diagnostic (CAD) tools can provide insight to the use of medical images for diagnosis and can augment diagnostic speed. However, due to the variation in the parameter settings across different patients, substantial discrepancies in image voxels persist. We found that different voxel sizes can create a compromise between model generalization and diagnostic efficacy. This study investigates the performance disparities of diagnostic models trained on original images and LDCT images reconstructed to different voxel sizes while making isotropic. We examined the ability of our method to differentiate between benign and malignant nodules. Using 11 features, a support vector machine (SVM) was trained on LDCT images using an isotropic voxel with a side length of 1.5 mm for 225 patients in-house. The result yields a favorable model performance with an accuracy of 0.9596 and an area under the receiver operating characteristic curve (ROC/AUC) of 0.9855. In addition, to furnish CAD tools for clinical application, future research including LDCT images from multi-centers is encouraged.

**Keywords:** lung cancer; isotropic voxel reconstruction; lung nodule classification; radiomics; computer-aided diagnosis



**Citation:** Hsiao, C.-C.; Peng, C.-H.; Wu, F.-Z.; Cheng, D.-C. Impact of Voxel Normalization on a Machine Learning-Based Method: A Study on Pulmonary Nodule Malignancy Diagnosis Using Low-Dose Computed Tomography (LDCT). *Diagnostics* **2023**, *13*, 3690. <https://doi.org/10.3390/diagnostics13243690>

Academic Editor: Philippe A. Grenier

Received: 20 October 2023  
Revised: 4 December 2023  
Accepted: 14 December 2023  
Published: 18 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

According to the 2023 World Cancer Statistics [1], although the overall mortality rate of lung cancer (LC) has experienced a decline in 2023 compared to previous years, it remains the second most prevalent cancer globally in terms of both overall cancer-related deaths and new cases. Approximately 230,000 individuals are newly diagnosed with LC each year globally, and among them about 120,000 suffer malignancy. Clinically, LC can be broadly classified into non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). NSCLC is the most prevalent, accounting for approximately 85% of LC cases [2], and can further be categorized into adenocarcinoma and squamous-cell carcinoma subtypes [3]. The early detection and treatment of NSCLC typically leads to better prognoses and increases survival rates by five years P [3].

Currently, low-dose computed tomography (LDCT) is commonly used in early LC screening, enabling rapid and large-scale diagnoses for individuals with high risk factors [4]. However, these screenings generate a large number of images that require interpretation by radiologists. This not only results in a substantial burden on human resources but also

introduces variations in experience and subjective awareness among different radiologists, leading to a lack of unified diagnostic criteria [5]. Within LDCT images, pulmonary nodules (PNs) represent areas of suspected lesions and can be preliminarily categorized into three types: solid nodules (SN), part-solid nodules (PSN), and ground glass nodules (GGN) [6]. Typically, radiologists rely on visible characteristics such as the PN's size, edge shape, and nodule pattern for diagnosis [7], similar to Lung-RADS™ [8]. However, this approach often yields a high rate of false positives [9], which ultimately creates the need for a pathological biopsy for a final confirmation. This process often poses unnecessary and invasive risks to patients [10].

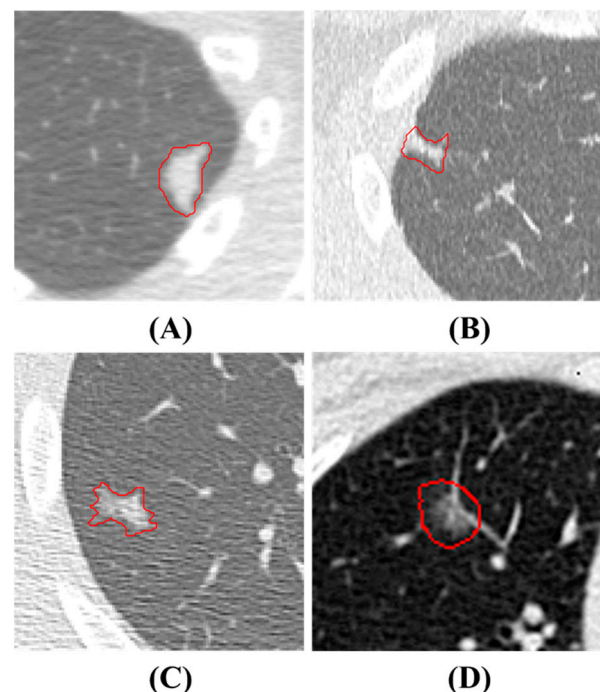
The use of machine learning (ML) and deep learning (DL) for pulmonary nodule detection and diagnosis from chest CT images has been studied for decades. In particular, radiomics [11] is a feature-extraction technique, that is usually combined with classifiers to establish models, which can provide physicians with an alternative opinion beyond visible characteristics. A previous study indicates that such models significantly enhance reading speed and diagnostic precision for physicians [12]. Numerous studies that focused on PN detection and classification found that these models can achieve promising results [13–23]. For example, Kumar et al. [16] proposed an approach utilizing autoencoders to extract features from raw images and employ a decision tree for binary classification, which achieved an accuracy of 0.7501. Shen et al. [22] introduced a method using multi-scale convolutional neural networks (MCNN) to distinguish a PN's malignancy, capturing differently sized patches of each PN and simultaneously using convolutional neural networks (CNN) for discrimination, which achieved an accuracy of 0.8612 on the LIDC-IDRI dataset [5]. Mehta et al. [19] (2021) presented a model that utilizes 3D CNNs in combination with volumetric radiomics and imaging biomarkers to diagnose a PN's malignancy. By incorporating 3D PN images, radiomics, and LIDC-IDRI biomarkers, they achieved an ROC/AUC of 0.8659. In 2021, Lu et al. [18] introduced a method that integrated the Marine Predators Algorithm with CNN for the diagnosis of PNs. This approach was impressive, achieving an accuracy of 0.934 and a sensitivity of 0.984. Halder et al. [24] attained the best PN diagnostic model performance using the LIDC-IDRI dataset in 2021. They employed morphology and two CNN networks and achieved an accuracy of 0.9610 and an AUC of 0.9936.

However, the above-mentioned studies face some common challenges. First, most patients in the LIDC-IDRI dataset lack pathological biopsy results on nodule malignancy diagnosis, which are considered to be the gold standard. The LIDC-IDRI data provides suspicion scores of nodule malignancy, ranging from 1 to 5, which were made by radiologists. Most previous studies classify nodules with scores of 1 and 2 as benign and those with scores of 4 and 5 as malignant, however, the actual malignancy of PNs remains uncertain. Second, previous studies typically used raw LDCT images as inputs to their models. They ignored the fact that images undergoing LDCT imaging may have had different modality settings such as pixel spacing (PS) and slice thickness (SK). This ignorance raises the question of whether the uniformity of a voxel size would affect model diagnostic performance or not. For instance, Kim et al. [25] (2019) reviewed factors that could impact the quantification of CT image features, highlighting the significant influence of PS and SK on feature quantification. Most features demonstrate different impacts before and after normalization of PS and SK. Lu et al. [26] (2017) improved model diagnostic performance by reconstructing positron emission tomography (PET) images into uniform voxel sizes. Their study offers valuable insights by investigating the impacts of different voxel sizes on PET image quantification and diagnosis. The above-mentioned studies focused on a traditional machine learning method, therefore they could examine feature impact accordingly. Recent studies often used neural networks (NN) and deep learning [27], which is outside of this study's scope. This study design was motivated by the above-mentioned problems; therefore, we did not consider using NN and DL. Furthermore, DL models demand large datasets for training, which is not suitable the small dataset used in this study.

## 2. Materials and Methods

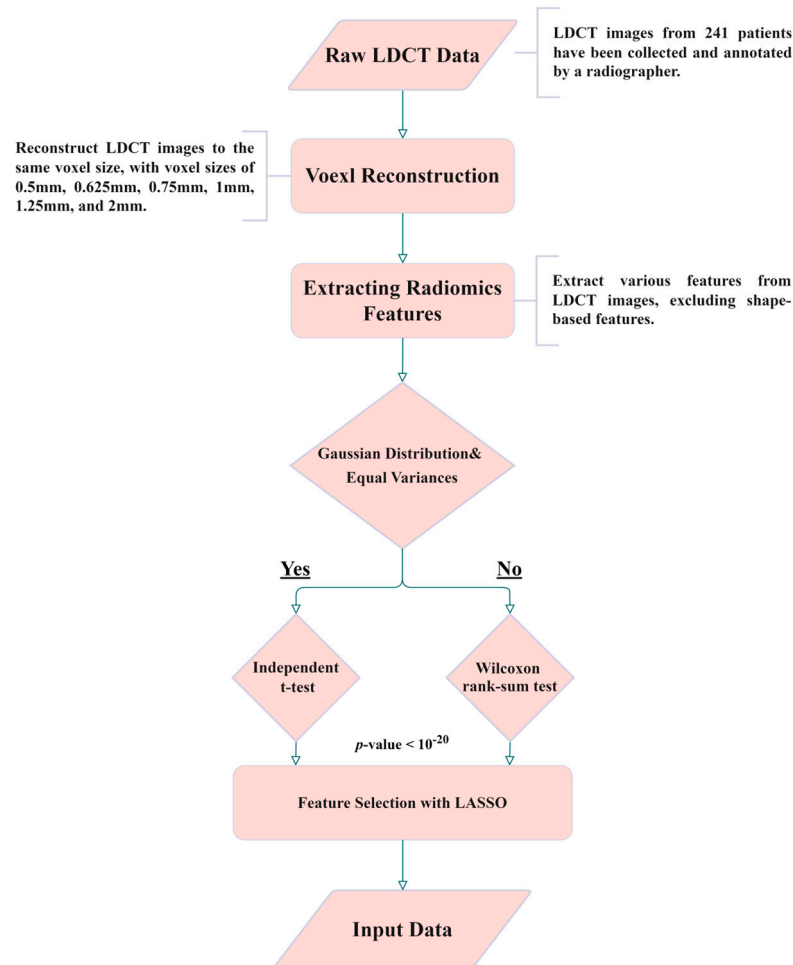
### 2.1. Dataset in House and Annotation

LDCT image data are collected from Kaohsiung Veterans General Hospital (KVGH), including 160 malignant and 81 benign PNs from a total of 241 patients. The data collection process has been approved by the Institutional Review Board (IRB) of KVGH with the IRB number VGHKS18-CT5-09. The malignancy status of each PN is defined by pathological biopsy and the boundary of the PN is annotated by an experienced radiological technician. LifeX (Version 6.2.0, C. Nioche, Inserm, Paris, France) is open-source software that was first published in 2018 [28]. Its application involves reading DICOM images and synchronizing the display and annotation of medical images in the coronal, sagittal, and axial slices. It also concurrently supports users in the extraction of radiomics features within regions of interest (ROIs). We use LifeX as an annotation tool, which allows us to output nearly raw raster data (NRRD) files for image feature extraction. However, due to technical problems with LifeX, we had ROI annotation errors in 3 benign and 13 malignant patients. Therefore, these 16 patients' data were excluded for the following process. The final images (for following training and test usage) have a size of either  $512 \times 512$  or  $768 \times 768$ , depending on the raw data. Notably, the radiomics features are extracted only in ROIs in the reconstructed images, not directly from LDCT raw images. In these LDCT images, 73 patients were scanned using equipment from TOSHIBA, 7 patients were scanned using equipment from SIEMENS (Munich, Germany), 143 patients were scanned using equipment from GE MEDICAL SYSTEMS (Barrington, IL, USA), and 2 patients were scanned using equipment from Philips (Amsterdam, The Netherlands) for contrast imaging. Figure 1 illustrates PNs, where benign and malignant nodules appear to be strikingly similar, rendering precise diagnoses challenging. In this study, all nodule contours are manually delineated by an experienced radiological technician.



**Figure 1.** PNs Presentation. The red contours are manual delineations on nodules. (A) and (B) depict benign PNs, while (C) and (D) show malignant PNs.

For ease of understanding, Figure 2 shows the flowchart of this study. The details of every block in the flowchart are described in the following paragraphs.



**Figure 2.** Experimental flowchart of this study.

### 2.2. Isotropic Voxel Normalization and Image Reconstruction

To explore the reproducibility and model performance of PN malignancy diagnoses using the ML method, we normalized the PS and ST of all of the LDCT images using bicubic interpolation. This reconstruction aims to achieve consistent spatial resolution across all LDCT images, making them isotropic in three axes. Furthermore, to evaluate the influence of different voxel sizes and spatial resolutions on the model and features, we reconstructed all images to various voxel sizes, including images with side lengths of 0.5, 0.625, 0.75, 1, 1.25, 1.5, 1.75, and 2 mm. We utilized bicubic interpolation to achieve isotropic voxels.

### 2.3. Radiomics and Feature Selection

We employed radiomics to extract 2112 features from the PN regions of the reconstructed images. Radiomics encompass various quantitative image feature extraction methods, including first-order statistics, shape-based, and texture-based methods [11]. Not all extracted features contribute significantly to PN malignancy diagnosis; therefore, feature selection is necessary and crucial. Given that PN malignancy is often related to its size [10], we intentionally excluded features with size information. Before feature dimension reduction, 14 features related to shape and size were manually excluded (see Table A1 in Appendix A). Moreover, 123 features with identical feature values and no discriminability were also manually excluded. The remaining 1975 features underwent further feature dimension reduction; this process it outlined in the following paragraph.

The image analysis problem could induce a feature selection problem. In order to select significant features for classification in our model, we employed various methods. Firstly, to identify features with significant differences in mean and mode among benign and

malignant PNs, we initially utilized the following two statistical methods: the independent *t*-test [29] and the Wilcoxon rank-sum [30] test. Features without significant differences were then excluded, with the goal of reducing the number of input features for the model. To perform an independent *t*-test, the data must be normally distributed and have equal variance. Meeting these criteria is crucial for using the independent *t*-test. Therefore, we first examine normality tests [31] and Levene tests [32] to assess the equality of variances and the type of distribution for both benign and malignant pulmonary nodule groups. If the data met the criteria of a normal distribution and equal variance, we proceeded to use an independent *t*-test for analysis, otherwise, we employed the Wilcoxon rank-sum test. Features between two classes with *p*-values less than  $10^{-20}$  are considered to be significantly different, which means they have good and distinguishable features. Second, we employed the well-known LASSO [33] algorithm. The LASSO algorithm effectively reduces coefficients to zero for those features with less contribution to classification, thus achieving feature selection. Third, we applied t-distributed stochastic neighbor embedding (t-SNE) [34] for dimension reduction of selected feature combinations to visualize the distribution of patients in a two-dimensional space. t-SNE, belonging to manifold learning, achieves dimension reduction and meanwhile preserves the local structure of data distribution. This t-SNE can be found for data visualization and dimension reduction in recent publications. The reason that LASSO can be used in feature selection is explained in Appendix B.

#### 2.4. Support Vector Machine (SVM) and Hyperparameter Optimization

SVM is a classic classifier [35] known for its strong performance in various classification tasks, particularly in dealing with small datasets. It exhibits a better generalization ability compared to DL models when dealing with small dataset [36]. In our study, we compared four combinations of feature selection: (1) all features without exclusion, (2) features with a *p*-value less than  $10^{-20}$ , (3) features selected by LASSO, and (4) features selected by the t-SNE algorithm. Before inputting all features into the SVM model, we performed a Min-Max Normalization on all features to ensure that different features had comparable numeric ranges. This is a basic feature-normalization process. In SVM, the choice of kernel function, which determines the decision boundary, is crucial. We employed the Gaussian Radial Basis Function (RBF) kernel function for our SVM model because most feature distributions have a normal distribution after feature selection. In ML models, hyperparameter optimization is highly correlated with final model performance. We used Gaussian Bayesian optimization [37] to tune the hyperparameters of the SVM model to achieve a better performance.

#### 2.5. K-Fold Cross-Validation and Model Performance Evaluation

To measure the model performance, we manipulated data for training and tested it as follows: We randomly sampled 80% of the benign and malignant data as training data, with the remaining 20% serving as test data. This sampling and test process was repeated 4000 times. We then evaluated the model's performance on different voxel sizes and feature selection methods using various metrics including Balanced Accuracy, Weighted Sensitivity, Weighted Precision, Weighted F-score, and Weighted AUC. Equations (1)–(5) represent the formulas for these metrics. Finally, we conducted a 10-fold cross-validation to assess the model performance and plot ROC curves.

$$\text{Balanced Accuracy} = \frac{1}{k} \sum_{i=0}^k \frac{TP_i}{(TP_i + FN_i)} \quad (1)$$

$$\text{Weighted Sensitivity} = \frac{\sum_{i=0}^k w_i \times TP_i}{\sum_{i=0}^k w_i * (TP_i + FN_i)} \quad (2)$$



$$\text{Weighted Precision} = \frac{\sum_{i=0}^k w_i \times TP_i}{\sum_{i=0}^k w_i \times (TP_i + FP_i)} \quad (3)$$

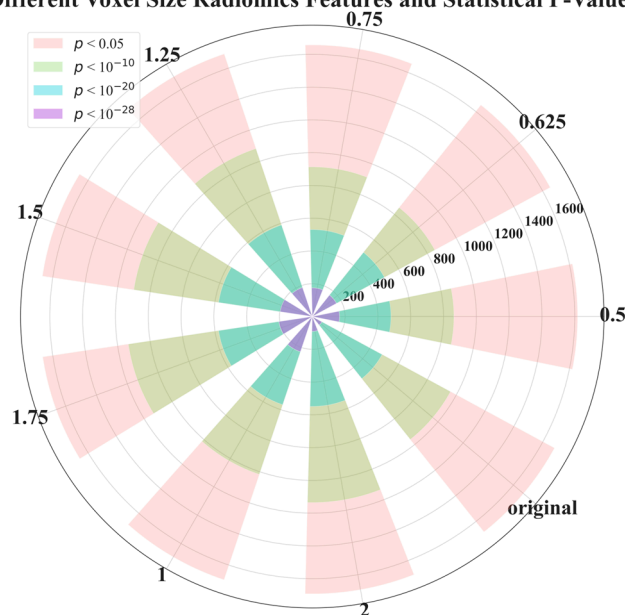
$$\text{Weighted F1 score} = \frac{\sum_{i=0}^k w_i \times \left( \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i} \right)}{\sum_{i=0}^k w_i} \quad (4)$$

$$\text{One-vs-All AUC} = \frac{\sum_{i=0}^k w_i \times AUC_i}{\sum_{i=0}^k w_i} \quad (5)$$

### 3. Results

In Figure 3, we demonstrate the distribution of  $p$ -values for features extracted at different voxel sizes reconstructed from the raw image data (LDCT). The value started from 0.5, and 0.625 to 2 is the side length of the isotropic voxel in mm. The word ‘original’ means there is no reconstruction, it is raw LDCT data. The value started from 200, and 400 to 1600 is the number of features. We categorized  $p$ -values into four groups: less than 0.05,  $1 \times 10^{-10}$ ,  $1 \times 10^{-20}$ , and  $1 \times 10^{-28}$ . Interestingly, Figure 3 shows that the majority of features have  $p$ -values in the range:  $[0.05, 1 \times 10^{-10}]$ . However, a  $p$ -value in this range is not capable of distinguishing or classifying. Therefore, for the most part, features are useless. We also found that features extracted from the original LDCT images without uniform voxel sizes had the least useful features. Here the useful features are indicated by the purple part of the circle in Figure 3 ( $p$ -value  $< 1 \times 10^{-28}$ ). This indicates that voxel normalization indeed affects feature extraction. Through this voxel normalization process using image reconstruction, the extracted image features exhibit significant statistical differences between benign and malignant nodules. This result confirms our hypothesis. Notably, features extracted from reconstructed data normalized to a side length of 2 mm performed poorly compared to other side lengths. The exact quantity of features plotted in Figure 3 can be seen in Table 1.

Different Voxel Size Radiomics Features and Statistical P-Values



**Figure 3.**  $p$ -value distribution of radiomics features extracted from voxels with different side lengths reconstructed from raw LDCT data. This experiment aimed to examine the feature efficacy of benign and malignant nodule classification, based on different voxel sizes.

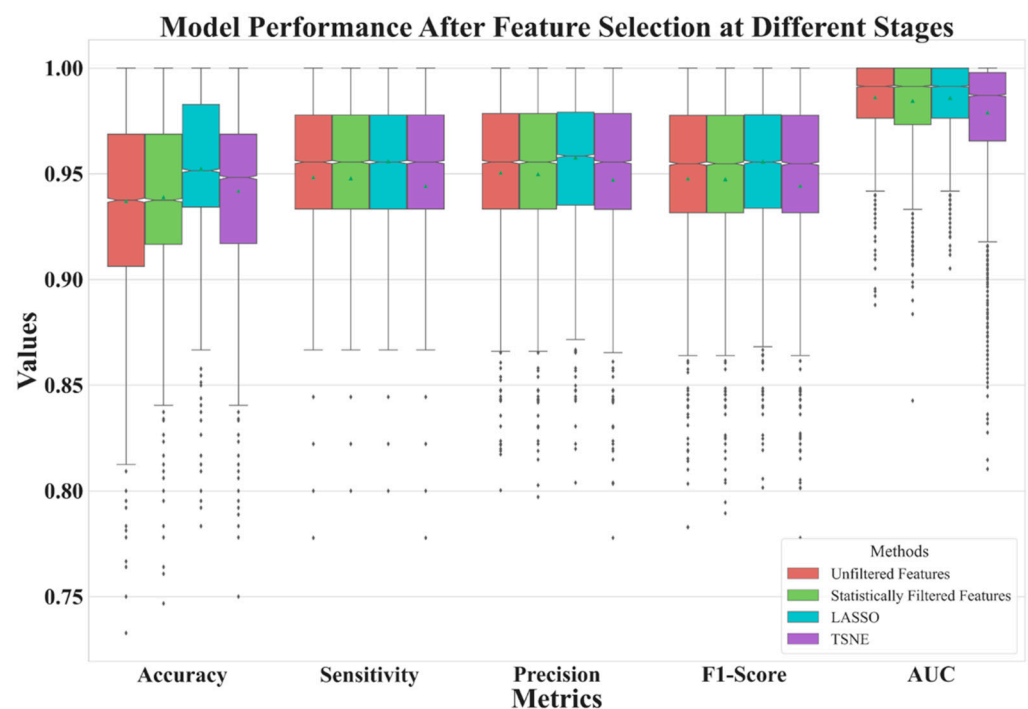
**Table 1.** The quantity of  $p$ -value of radiomics features distribution plotted in Figure 3.

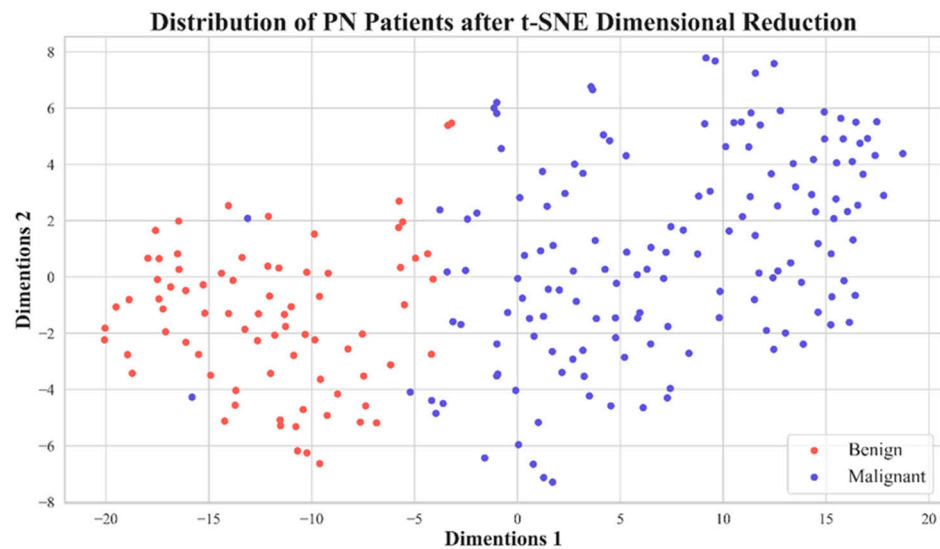
Voxel Size	0.5	0.625	0.75	1	1.25	1.5	1.75	2	Original
$p < 0.05$	1617	1650	1657	1694	1690	1663	1661	1692	1680
$p < 1 \times 10^{-10}$	863	850	913	1016	1081	1100	1134	1135	959
$p < 1 \times 10^{-20}$	480	501	531	568	590	578	578	549	485
$p < 1 \times 10^{-28}$	166	168	175	227	187	198	206	91	67

In Table 2, we illustrate the number of features before (the baseline) and after feature selection. In Figure 4, we demonstrate the model performance for comparison using 5 metrics illustrated in Equations (1)–(5). The results shows that the best feature selection method is LASSO, which used only 11 features. Surprisingly, the second-best performance was achieved by reducing the dimensionality of 11 LASSO-selected features to 2 ‘features’ (i.e., 2 directions in the feature space) using t-SNE, resulting in an outstanding performance. The feature distribution of these 2 directions are shown in Figure 5.

**Table 2.** Number of features after feature selection and raw LDCT images.

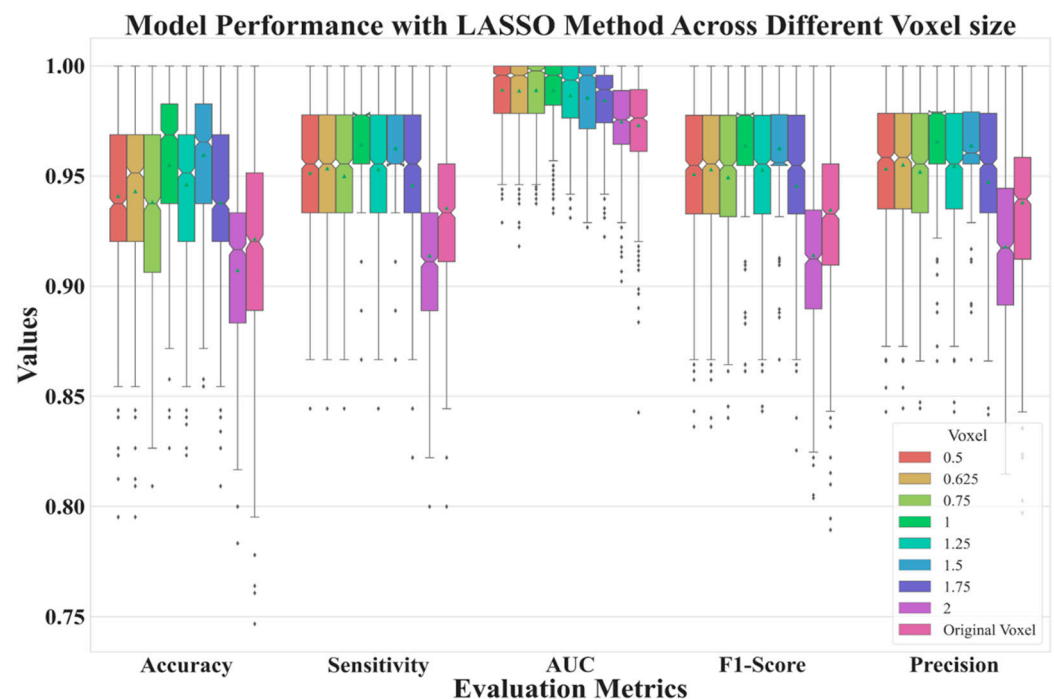
	Unfiltered Features	Statistically Filtered Features	LASSO	t-SNE
Feature Number	2061	480	11	2

**Figure 4.** Model performance comparison of three feature selection methods and a baseline.



**Figure 5.** Distribution plot of 11 features extracted from PNs with a voxel with a side length of 1.5 mm after dimension reduction using t-SNE.

Figure 6 depicts the performance of features extracted from different voxel sizes' reconstructions. Here, we used LASSO to select features. From the figure, it is clear that the model has better performances with voxel normalization with side lengths less than 2 mm. The best model was one with features extracted from LDCT images with a side length of 1.5 mm.

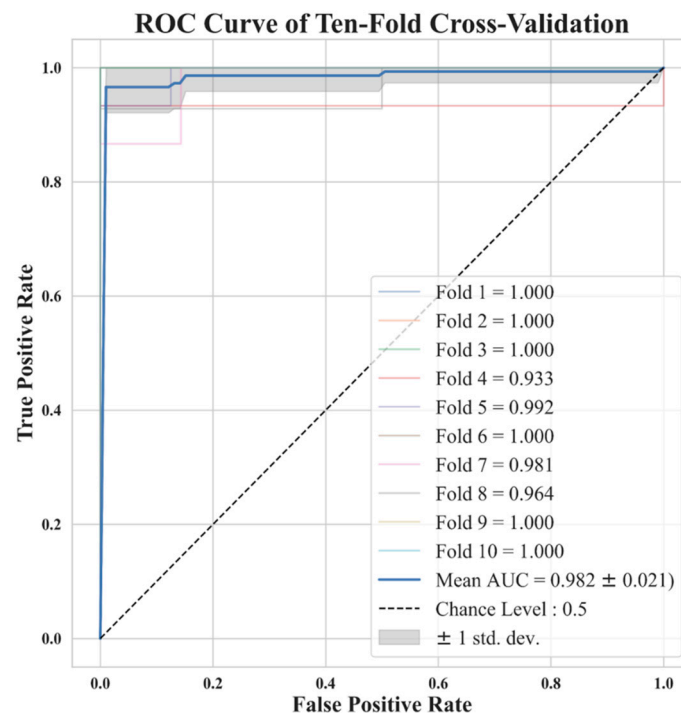


**Figure 6.** Model performance for reconstructed data with various side lengths.

In Figure 7, we show ROCs using ten-fold cross-validations with radiomic features selected by LASSO from image reconstruction with a side-length of 1.5 mm. The model average AUC reached 0.982, indicating a stable diagnostic performance. This is clinically acceptable for PN diagnoses. In Figure 5 we demonstrate the distribution of 2D feature space, with two dimension directions. After dimensionality reduction, only 2 features were



selected by t-SNE. From this distribution, it is clear that these 11 features can effectively classify benign and malignant nodules.



**Figure 7.** Ten-fold cross-validation ROC curve for the model trained with 11 features selected by LASSO with a voxel with a side length of 1.5 mm. Note that some lines are overlapping.

Finally, in Table 3 we compare different results with different voxel sizes. These results are the averaged values of various metrics in different voxel sizes of reconstruction repeated over 4000 times. We also compare our results to previous state-of-the-art models. Our best result outperforms most previous methods and is comparable to the best previous study [24]. Indeed, deep learning approaches need more time in training than the traditional method.

**Table 3.** Comparison of test results averaged from 4000 times; 80% training and 20% test in partition. The values in the first column denote different side lengths of voxels after image reconstruction.

	Accuracy	AUC	Sensitivity	Precision	F1 Score
0.5	0.9409	0.9891	0.9514	0.9533	0.9509
0.625	0.9431	0.9887	0.9535	0.9551	0.9530
0.75	0.9350	0.9890	0.9481	0.9501	0.9473
1	0.9531	0.9890	0.9624	0.9640	0.9620
1.25	0.9467	0.9866	0.9532	0.9548	0.9530
1.5	<b>0.9596</b>	<b>0.9855</b>	<b>0.9619</b>	0.9633	0.9619
1.75	0.9371	0.9844	0.9452	0.9468	0.9449
2	0.9073	0.9747	0.9156	0.9197	0.9159
Original	0.9223	0.9731	0.9357	0.9381	0.9349
Halder et al., 2021 [24]	<b>0.9610</b>	<b>0.9936</b>	<b>0.9685</b>	-	-
Mehta et al., 2021 [19]	-	0.8659	-	-	-
Shen et al., 2017 [22]	0.8612	-	-	-	-
Lu et al., 2021 [18]	0.934	0.984	-	-	-

Table 4 presents the 11 radiomic features selected by the LASSO in this study. Notably, 7 of the 11 selected features are texture-related, demonstrating their important role in

discriminating between benign and malignant PNs. Therefore, voxel normalization is essential and has a profound impact on reproducibility.

**Table 4.** Description of the 11 Features Utilized by the Model.

Feature Description	Types
original_gldm_SmallDependenceLowGrayLevelEmphasis	Texture
log-sigma-2-0-mm-3D_gldm_DifferenceEntropy	Texture
log-sigma-2-0-mm-3D_gldm_SmallDependenceEmphasis	Texture
log-sigma-3-0-mm-3D_glszm_ZonePercentage	Texture
lbp-2D_gldm_DependenceNonUniformityNormalized	Texture
lbp-3D-m1_gldm_DependenceNonUniformityNormalized	Texture
lbp-3D-m2_gldm_DependenceNonUniformityNormalized	Texture
log-sigma-2-0-mm-3D_firstorder_Mean	First order
lbp-3D-m1_firstorder_Skewness lbp-3D-wavelet-LLH_firstorder_Mean	First order
wavelet-LHL_firstorder_Mean	First order

#### 4. Discussion

Due to the diversity of patients in clinics and orders from physicians, radiographers use various parameter settings to acquire LDCT images. If these parameter settings are not consistent, it is difficult to compare them due to vacancy on baseline. Particularly, we are referring to texture features; different pixel-size on images can cause different results, such as features extracted from Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and other texture features. Many previous studies have indicated that texture features are crucial for distinguishing the benign and malignant nature of PNs.

In this research, normalizing the voxel size from LDCT images to 1–1.5 mm yields good model performance. Typically, CT has a spatial resolution ranging from 0.5 mm to 0.625 mm in the  $x$ - $y$  axis; our collected data fall in the range of 0.6 mm to 0.8 mm. The resolution on the  $z$ -axis depends on the temporal resolution. Most images on the  $z$ -axis resolution fall in the range of 1 mm to 5 mm in our in-house dataset. We do not have data in the  $z$ -axis measuring less than 1 mm. The image reconstruction must consider the spatial and temporal resolution of the raw data; otherwise, it can lead to a partial volume effect and subsequently impacts the model performance. Further, reconstructing images with a voxel larger than 1.5 mm of side length will significantly decrease spatial resolution, making it impossible to capture fine texture of PNs. Therefore, we recommend that future studies consider voxel reconstruction within the range of 1–1.5 mm, based on a prerequisite that data has higher spatial and temporal resolution than 1 mm in the  $x$ - $y$ - $z$  axis, i.e.,  $\leq 1$  mm.

In 2023, Fischbach et al. [38] conducted a study on the diagnosis of nodules with respect to slice thickness (SK). In [38], they mention that as the SK decreases, the image noise decreases, and contrast increases. The authors tested SKs of 0.625 mm, 1.25 mm, 2.5 mm, 3.75 mm, 5 mm, 7.5 mm, and 10 mm. They found that a slice thickness of 0.625 mm yielded the best image quality. However, the best diagnostic quality was achieved with a slice thickness of 1.25 mm. Their finding aligns with our research findings, once again demonstrating that voxel normalization to an appropriate range (1.25–1.5 mm) can yield the best diagnostic benefits, while excessively small or large SK can lead to a decline in diagnostic performance.

One advantage of our study compared to previous studies is that all of our patients had pathological diagnoses to confirm the PN's nature, i.e., benign or malignant. This advantage is not present in large open datasets such as LIDC-IDRI. Research limitations regarding this study are discussed below. This study mainly analyzed Asian ethnic groups, and most malignant lesions are adenocarcinoma spectrum. Therefore, the predictive ability for other pathological types of lung cancer may be lower. Particularly, this study mainly distinguishes lung adenocarcinoma spectrum lesions from other benign lesions. Therefore, we did not conduct any further analysis on the clinical management of these nodules.

In this study, PN contours are manually drawn by experts. However, if the PN regions could be automatically segmented on the LDCT images, together with the proposed system, this could create a potential for in-clinic utilization. This will contribute to the futural development of a comprehensive end-to-end diagnostic system. Many previous studies are devoted to pulmonary nodule detection such as competitions using the LUNA2016 dataset [39]. LUNA2016 is a subset of LIDC-IDRI [5]. We also performed a study [40] on nodule detection using the complete LIDC-IDRI dataset. In the future, we will combine these two systems to form one comprehensive system.

In our study, a significant limitation arises from the relatively small dataset obtained from a single center. The exclusive collection of patient imaging data with pathology verification posed challenges in the data acquisition process. Furthermore, the need for experienced personnel to manually delineate Regions of Interest (ROIs) for each case added complexity to dataset acquisition. To overcome this limitation, our future efforts will focus on the continuous collection of imaging data with pathology verification and uniform voxel sizes from various centers. Our aim is to establish a comprehensive database tailored to the Asian population, providing a substantial resource for subsequent researchers conducting studies on PNs recognition.

## 5. Conclusions

This study used a ML method combined with radiomic features extracted from reconstructed images with voxel normalization from LDCT. We mainly explored the impact of voxel normalization to predict performance between benign and malignant pulmonary nodules. Our study offers a recommendation: before using radiomics, the voxel normalization is important and crucial to texture-related studies. The reconstruction must consider the limitation on raw data temporal resolution. According to our finding, the best prediction of nodule classification, benign or malignant, is achieved using an isotropic voxel with a side length of 1.5 mm.

**Author Contributions:** Conceptualization, D.-C.C.; methodology, D.-C.C. and C.-H.P.; software, C.-H.P.; validation, C.-H.P.; formal analysis, D.-C.C. and C.-H.P.; investigation, D.-C.C. and C.-H.P.; resources, F.-Z.W.; data curation, C.-C.H.; writing—original draft preparation, D.-C.C. and C.-H.P.; writing—review and editing, D.-C.C. and F.-Z.W.; visualization, C.-H.P.; supervision, D.-C.C.; project administration, D.-C.C.; funding acquisition, C.-C.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Kaohsiung Veterans General Hospital, grant number KSVGH-112-110 and KSVGH-113-075.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (IRB) of KVGH with the IRB number VGHKS18-CT5-09 (protocol code:18-CT5-09, 3 April 2018).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing is not applicable to this article.

**Conflicts of Interest:** Not applicable.

## Appendix A

**Table A1.** Description of the 14 manually excluded features.

Feature Description	Types
original_shape_Elongation	Shape-Based
original_shape_Flatness	Shape-Based
original_shape_LeastAxisLength	Shape-Based
original_shape_MajorAxisLength	Shape-Based
original_shape_Maximum2DDiameterColumn	Shape-Based
original_shape_Maximum2DDiameterRow	Shape-Based
original_shape_Maximum2DDiameterSlice	Shape-Based
original_shape_Maximum3DDiameter	Shape-Based
original_shape_MeshVolume	Shape-Based
original_shape_MinorAxisLength	Shape-Based
original_shape_Sphericity	Shape-Based
original_shape_SurfaceArea	Shape-Based
original_shape_SurfaceVolumeRatio	Shape-Based
original_shape_VoxelVolume	Shape-Based

## Appendix B

$X$  and  $Y$  are known, where  $X$  is an  $n$  by  $p$  matrix;  $X \in R^{n \times p}$ , and  $Y$  is a vector;  $Y \in R^{n \times 1}$ . To solve a linear system problem defined by  $Y = X\beta$ , where unknown  $\beta \in R^{p \times 1}$ , using the least square error one obtains the solution  $\hat{\beta} = (X^T X)^{-1} X^T Y$ . Here  $\hat{\beta}$  is the least square (LS) estimator. The problem of this regression is singular or near singular, since the computation of LS estimator depends on  $(X^T X)^{-1}$ . To solve this problem, a penalty term can be added. We rewrite the estimator to be  $\hat{\beta}_{ridge} = (X^T X + \lambda I_p)^{-1} X^T Y$ , where  $I_p$  is an identity matrix with a rank  $p$  and  $\lambda > 0$ . Since  $\lambda I_p$  is non-singular, the singular problem is solved. This is so-called ridge regression.  $\hat{\beta}_{ridge}$  is chosen to minimize the LS error  $(Y - X\beta)^T (Y - X\beta) + \lambda \beta^T \beta$ . The penalty coefficient  $\lambda$  forces the estimator  $\hat{\beta}_{ridge}$  to be small in order to get an optimal solution, which is a similar form to minimize the LS error  $(Y - X\beta)^T (Y - X\beta)$ . However, the over-fitting problem is relieved but not solved, because the ridge regression is still using an  $L_2$  norm, i.e.,  $\beta^T \beta$ .

The over-fitting causes the  $L_2$  norm regression to fit well to the training data but not to the test data. LASSO is a regression using  $L_1$  norm. It is necessary to minimize the error  $(Y - X\beta)^T (Y - X\beta) + \lambda |\beta|$ . This mechanism forces some elements in estimator  $\hat{\beta}_{LASSO}$  to be zero, which are corresponding to large values. Therefore, LASSO is a method of feature selection.

## References

1. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [[CrossRef](#)] [[PubMed](#)]
2. Herbst, R.S.; Morgensztern, D.; Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **2018**, *553*, 446–454. [[CrossRef](#)] [[PubMed](#)]
3. Knight, S.B.; Crosbie, P.; Balata, H.; Chudziak, J.; Hussell, T.; Dive, C. Progress and prospects of early detection in lung cancer. *Open Biol.* **2017**, *7*, 170070.
4. Jemal, A.; Fedewa, S.A. Lung cancer screening with low-dose computed tomography in the United States—2010 to 2015. *JAMA Oncol.* **2017**, *3*, 1278–1281. [[CrossRef](#)] [[PubMed](#)]
5. Armato, S.G., III; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B.; Aberle, D.R.; Henschke, C.I.; Hoffman, E.A. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med. Phys.* **2011**, *38*, 915–931. [[CrossRef](#)] [[PubMed](#)]
6. Cha, M.J.; Lee, K.S.; Kim, H.S.; Lee, S.W.; Jeong, C.J.; Kim, E.Y.; Lee, H.Y. Improvement in imaging diagnosis technique and modalities for solitary pulmonary nodules: From ground-glass opacity nodules to part-solid and solid nodules. *Expert Rev. Respir. Med.* **2016**, *10*, 261–278. [[CrossRef](#)] [[PubMed](#)]
7. Swensen, S.J.; Viggiano, R.W.; Midthun, D.E.; Müller, N.L.; Sherrick, A.; Yamashita, K.; Naidich, D.P.; Patz, E.F.; Hartman, T.E.; Muhm, J.R. Lung nodule enhancement at CT: Multicenter study. *Radiology* **2000**, *214*, 73–80. [[CrossRef](#)]

8. Chelala, L.; Hossain, R.; Kazerooni, E.A.; Christensen, J.D.; Dyer, D.S.; White, C.S. Lung-RADS version 1.1: Challenges and a look ahead, from the AJR special series on radiology reporting and data systems. *Am. J. Roentgenol.* **2021**, *216*, 1411–1422. [\[CrossRef\]](#)
9. Swensen, S.J.; Silverstein, M.D.; Edell, E.S.; Trastek, V.F.; Aughenbaugh, G.L.; Ilstrup, D.M.; Schleck, C.D. Solitary pulmonary nodules: Clinical prediction model versus physicians. *Mayo Clin. Proc.* **1999**, *74*, 319–329. [\[CrossRef\]](#)
10. Khan, T.; Usman, Y.; Abdo, T.; Chaudry, F.; Keddissi, J.I.; Youness, H.A. Diagnosis and management of peripheral lung nodule. *Ann. Transl. Med.* **2019**, *7*, 348. [\[CrossRef\]](#)
11. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **2016**, *278*, 563–577. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Doi, K. Current status and future potential of computer-aided diagnosis in medical imaging. *Br. J. Radiol.* **2005**, *78* (Suppl. S1), s3–s19. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Dey, R.; Lu, Z.; Hong, Y. Diagnostic classification of lung nodules using 3D neural networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; IEEE: New York, NY, USA, 2018; pp. 774–778.
14. Huang, H.; Wu, R.; Li, Y.; Peng, C. Self-supervised transfer learning based on domain adaptation for benign-malignant lung nodule classification on thoracic CT. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 3860–3871. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Kang, G.; Liu, K.; Hou, B.; Zhang, N. 3D multi-view convolutional neural networks for lung nodule classification. *PLoS ONE* **2017**, *12*, e0188290. [\[CrossRef\]](#)
16. Kumar, D.; Wong, A.; Clausi, D.A. Lung nodule classification using deep features in CT images. In Proceedings of the 2015 12th Conference on Computer and Robot Vision, Halifax, NS, Canada, 3–5 June 2015; IEEE: New York, NY, USA, 2015; pp. 133–138.
17. Liu, K.; Kang, G. Multiview convolutional neural networks for lung nodule classification. *Int. J. Imaging Syst. Technol.* **2017**, *27*, 12–22. [\[CrossRef\]](#)
18. Lu, X.; Nanehkaran, Y.A.; Karimi Fard, M. A method for optimal detection of lung cancer based on deep learning optimized by marine predators algorithm. *Comput. Intell. Neurosci.* **2021**, *2021*, 3694723. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Mehta, K.; Jain, A.; Mangalagiri, J.; Menon, S.; Nguyen, P.; Chapman, D.R. Lung nodule classification using biomarkers, volumetric radiomics, and 3D CNNs. *J. Digit. Imaging* **2021**, *34*, 647–666. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Saihood, A.; Karshenas, H.; Nilchi, A.R.N. Deep fusion of gray level co-occurrence matrices for lung nodule classification. *PLoS ONE* **2022**, *17*, e0274516. [\[CrossRef\]](#)
21. Shen, W.; Zhou, M.; Yang, F.; Yang, C.; Tian, J. Multi-scale convolutional neural networks for lung nodule classification. In Proceedings of the 24th International Conference, Information Processing in Medical Imaging IPMI 2015, Isle of Skye, UK, 28 June–3 July 2015; Springer: Cham, Switzerland, 2015; pp. 588–599.
22. Shen, W.; Zhou, M.; Yang, F.; Yu, D.; Dong, D.; Yang, C.; Zang, Y.; Tian, J. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognit.* **2017**, *61*, 663–673. [\[CrossRef\]](#)
23. Tomassini, S.; Falcionelli, N.; Sernani, P.; Burattini, L.; Dragoni, A.F. Lung nodule diagnosis and cancer histology classification from computed tomography data by convolutional neural networks: A survey. *Comput. Biol. Med.* **2022**, *146*, 105691. [\[CrossRef\]](#)
24. Halder, A.; Chatterjee, S.; Dey, D. Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification. *Biomed. Signal Process. Control* **2022**, *72*, 103347. [\[CrossRef\]](#)
25. Kim, H.; Goo, J.M.; Ohno, Y.; Kauczor, H.-U.; Hoffman, E.A.; Gee, J.C.; Van Beek, E.J. Effect of reconstruction parameters on the quantitative analysis of chest computed tomography. *J. Thorac. Imaging* **2019**, *34*, 92–102. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Lu, Y.; Fontaine, K.; Germino, M.; Mulnix, T.; Casey, M.E.; Carson, R.E.; Liu, C. Investigation of sub-centimeter lung nodule quantification for low-dose PET. *IEEE Trans. Radiat. Plasma Med. Sci.* **2017**, *2*, 41–50. [\[CrossRef\]](#)
27. Tonekaboni, S.; Joshi, S.; McCradden, M.D.; Goldenberg, A. What clinicians want: Contextualizing explainable machine learning for clinical end use. In Proceedings of the Machine Learning for Healthcare Conference (PMLR), Ann Arbor, MI, USA, 8–10 August 2019; pp. 359–380.
28. Nioche, C.; Orlhac, F.; Boughdad, S.; Reuzé, S.; Goya-Outi, J.; Robert, C.; Pellot-Barakat, C.; Soussan, M.; Frouin, F.; Buvat, I. LIFEx: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* **2018**, *78*, 4786–4789. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Heeren, T.; D’Agostino, R. Robustness of the two independent samples *t*-test when applied to ordinal scaled data. *Stat. Med.* **1987**, *6*, 79–90. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Rosner, B.; Glynn, R.J.; Ting Lee, M.L. Incorporation of clustering effects for the Wilcoxon rank sum test: A large-sample approach. *Biometrics* **2003**, *59*, 1089–1098. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Das, K.R.; Imon, A. A brief review of tests for normality. *Am. J. Theor. Appl. Stat.* **2016**, *5*, 5–12.
32. Schultz, B.B. Levene’s test for relative variation. *Syst. Zool.* **1985**, *34*, 449–456. [\[CrossRef\]](#)
33. Fonti, V.; Belitser, E. Feature selection using lasso. *VU Amst. Res. Pap. Bus. Anal.* **2017**, *30*, 1–25.
34. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
35. Suthaharan, S.; Suthaharan, S. Support vector machine. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Springer: Boston, MA, USA, 2016; pp. 207–235.
36. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **2020**, *408*, 189–215. [\[CrossRef\]](#)
37. Frazier, P.I. A tutorial on Bayesian optimization. *arXiv* **2018**, arXiv:1807.02811.

38. Abdulkareem, N.K.; Hajee, S.I.; Hassan, F.F.; Ibrahim, I.K.; Al-Khalidi, R.E.H.; Abdulqader, N.A. Investigating the slice thickness effect on noise and diagnostic content of single-source multi-slice computerized axial tomography. *J. Med. Life* **2023**, *16*, 862. [[CrossRef](#)] [[PubMed](#)]
39. LUNG Nodule Analysis (LUNA) 2016, Grand Challenges. Available online: <https://luna16.grand-challenge.org/Data/> (accessed on 20 November 2023).
40. Huang, C.H.; Cheng, D.C. Lung Nodule Detection Method on Low-Dose Chest Computer Tomography Images Using Deep Learning and Its Computer Program Product. Invention Patent I733627, 11 July 2021.

**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.