RSNA 2023
LEADING THROUGH CHANGE

UCSF
University of California
San Francisco

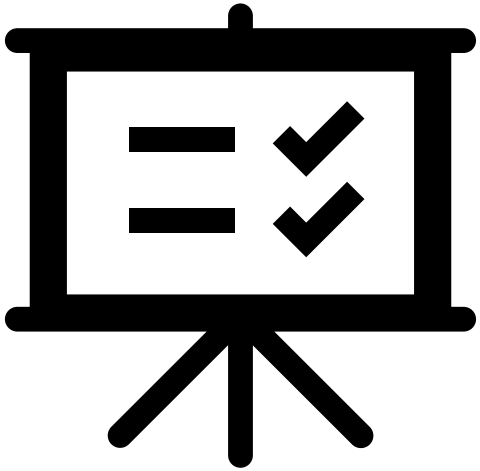
Radiomics Analysis for Predicting Progression of Part-Solid Nodules on CT

Shiny Weng[1], Masha Bondarenko[1], Gunvant Chaudhari[2], Arun Innanje[3], Terrence Chen[3], Jae Ho Sohn[2]

University of California, Berkeley [1]
University of California, San Francisco [2]
United Imaging Intelligence [3]

Abbreviations

- NLP (Natural Language Processing)
- SVM (Support Vector Machine)
- GLCM (Gray Level Co-occurrence Matrix)
- GLSZM (Gray Level Size Zone Matrix)
- GLDM (Gray Level Dependence Matrix)
- GLRLM (Gray Level Run Length Matrix)
- ROC (Receiver Operating Characteristic)
- AUC (Area Under the ROC Curve)

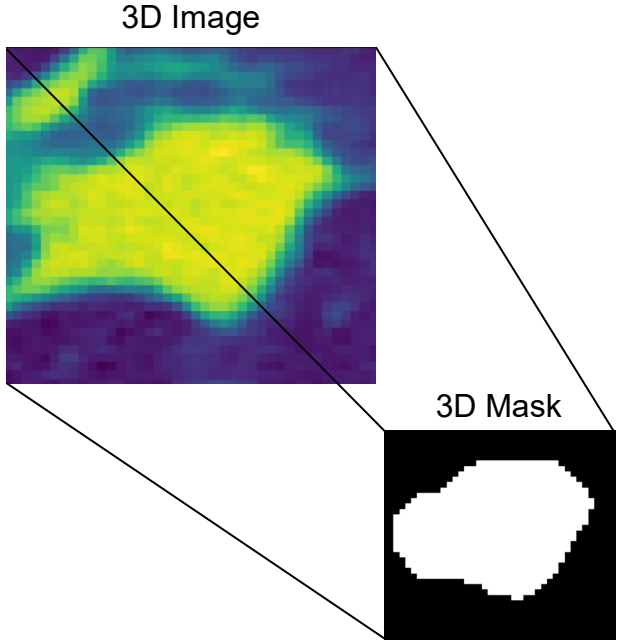


Key Terms

- **Radiomics** is a quantitative approach to medical imaging, extracting a large variety of features based on attributes such as pixel intensity, arrangement, color, and texture.
- **Support Vector Machine (SVM)** is a supervised machine learning algorithm used for classification and regression tasks.
- **Receiver Operating Characteristic (ROC)** curve is a graphical tool for illustrating the trade-off between sensitivity and specificity in the context of classification.
- **Gray Level Co-occurrence Matrix (GLCM)** is a texture analysis technique that characterizes the spatial relationships of pixel intensity values in an image.
- **Gray Level Size Zone Matrix (GLSZM)** is a texture analysis technique that provides information about how different sizes of connected regions with the same intensity values are distributed throughout the image.
- **Gray Level Dependence Matrix (GLDM)** is a texture analysis technique that characterizes the spatial dependencies or relationships between neighboring pixels with similar intensity values.
- **Gray Level Run Length Matrix (GLRLM)** is a texture analysis technique that quantifies and describes the lengths of consecutive runs of pixel intensities in an image.

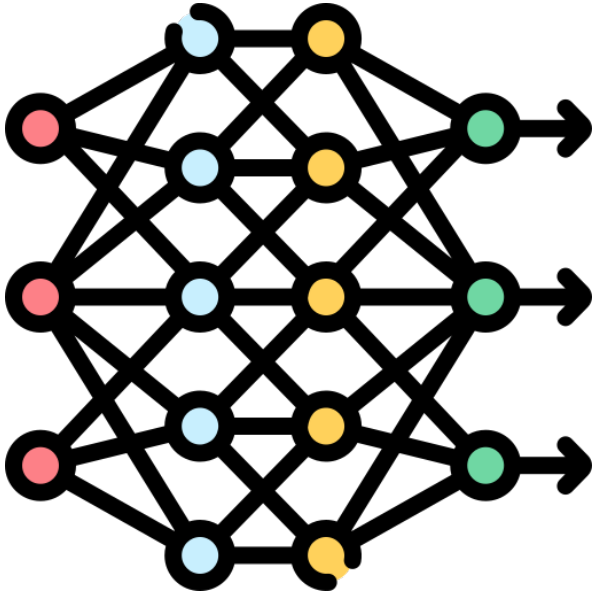
Background

- Accurate identification of growing subsolid nodules is crucial for effective risk stratification and the early detection of invasive lung cancer, allowing for timely treatment while avoiding unnecessary surgery on low-risk nodules that would otherwise remain stable.
- The traditional method of risk stratification, which relies on qualitative visual analysis of CT scans, remains challenging.
- Therefore, this study aims to leverage a longitudinal dataset of subsolid nodules on CT and develop radiomic and clinical feature-based models to identify nodules that are likely to grow over time.



Purpose

The purpose of this study is to develop a machine learning model to predict the growth of subsolid nodules using a combination of radiomic and clinical features.



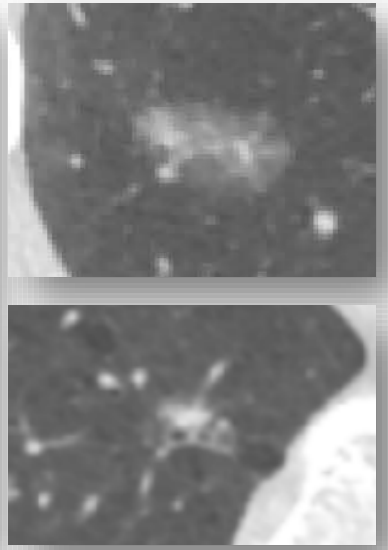
Materials and Methods

- A retrospective study was conducted on a cohort of patients who had undergone chest CT scans at a single institution between 2015 and 2019.

Dataset and Patient Characteristics

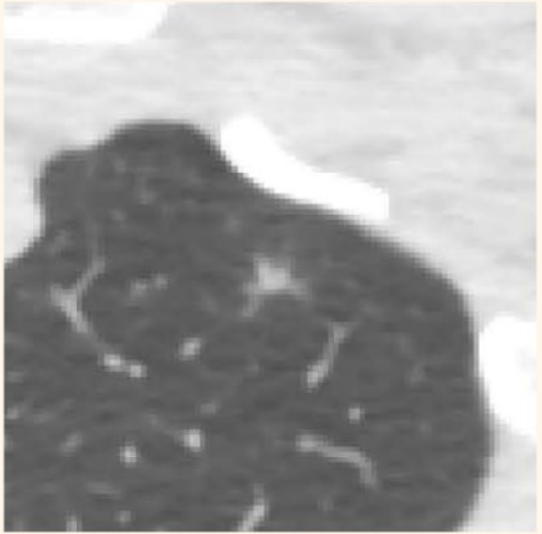
Characteristics	Unchanging Group (N = 799, 85.2%)	Increasing Group (N = 139, 14.8%)	P-Value §
Age — year (% > 60) *†	70.3 ± 10.6 (84.3%)	73.1 ± 11.6 (86.3%)	<0.001
Male Sex — no. (%)	239	47	0.411
Nodule Size — mm. ‡	6.9 ± 4.0	10.4 ± 4.6	<0.001
Nodule Volume — mm^3	440.3 ± 1085.3	1072.0 ± 1353.0	<0.001

*** Plus–minus values are means ± SD**
† The percentage is the proportion of patients with age > 60 out of both unchanging and increasing groups, respectively
‡ Nodule Size is the average diameter of the nodule
§ P-Value is computed from Mann-Whitney U Tests (increasing vs. unchanging) and Chi-Squared Tests (male vs. female)



Materials and Methods

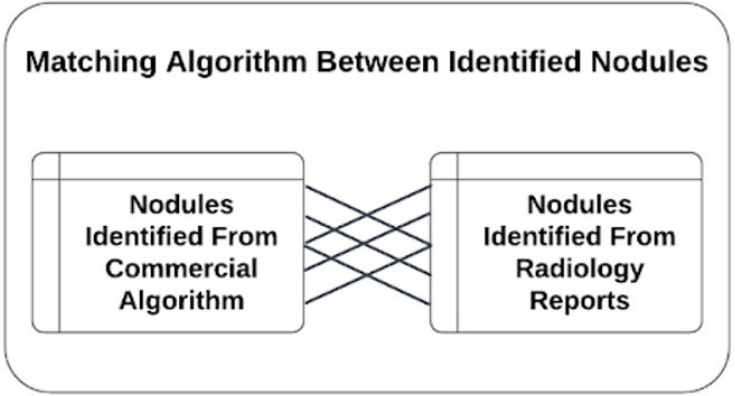
- Corresponding radiology reports were used to extract ground truth labels for the nodules' growth status (i.e., increased, decreased, unchanged), as well as other related information such as nodule size, nodule type, location, slice number, component sizes, and compared studies.
- An automated NLP pipeline was employed for the extraction of labels.
- Additionally, associated CT scans were processed through a commercial nodule characterization algorithm, which generated parameters such as nodule size, location, and segmentation coordinates.



Average (Report): 8.0 mm
Previous Average (Report): 1.5 mm
Nodule Type: partial solid
Component Sizes: 3-4 mm
Component Types: solid
Study Date: 4/9/2019
Comparison Date: ['2018-10-10']
Elapsed Days: 181
Doubling Time (Report): 74.9

Materials and Methods

- To link the nodules reported in the radiology reports with their specific coordinates and segmentations, we processed the CT scans using a commercial algorithm (United Imaging Intelligence, Boston, MA, USA), which identified the location, coordinates, and nodule segmentations.
- We then applied a matching algorithm using the parameters generated by the commercial algorithm and those extracted from the radiology reports to identify the intersection of the two groups.



Criteria for matching include:

1) Accession Number
2) Image/Slice
3) Nodule Location
4) Nodule Size

Materials and Methods

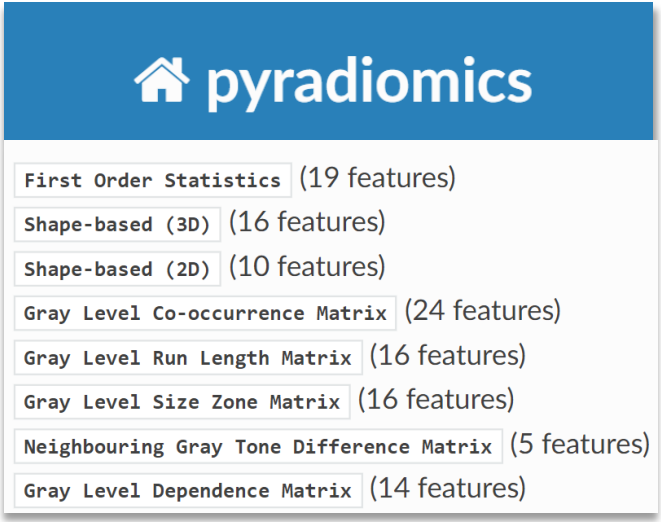
Utilizing this dataset, along with radiomic features (from pyRadiomics) and clinical data (e.g., patient age and sex), models for predicting the progression of subsolid nodules were developed.

The primary metric for evaluating model performance was the area under the curve (AUC) of the receiver operating characteristic (ROC), assessed on both the independent validation set and averaged across 5-fold cross-validation.

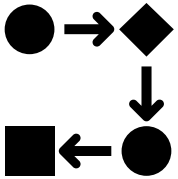
A total of 15 features were used for the model, selected through a combination of recursive feature elimination, random forest importance analysis, and univariate selection during cross-validation.

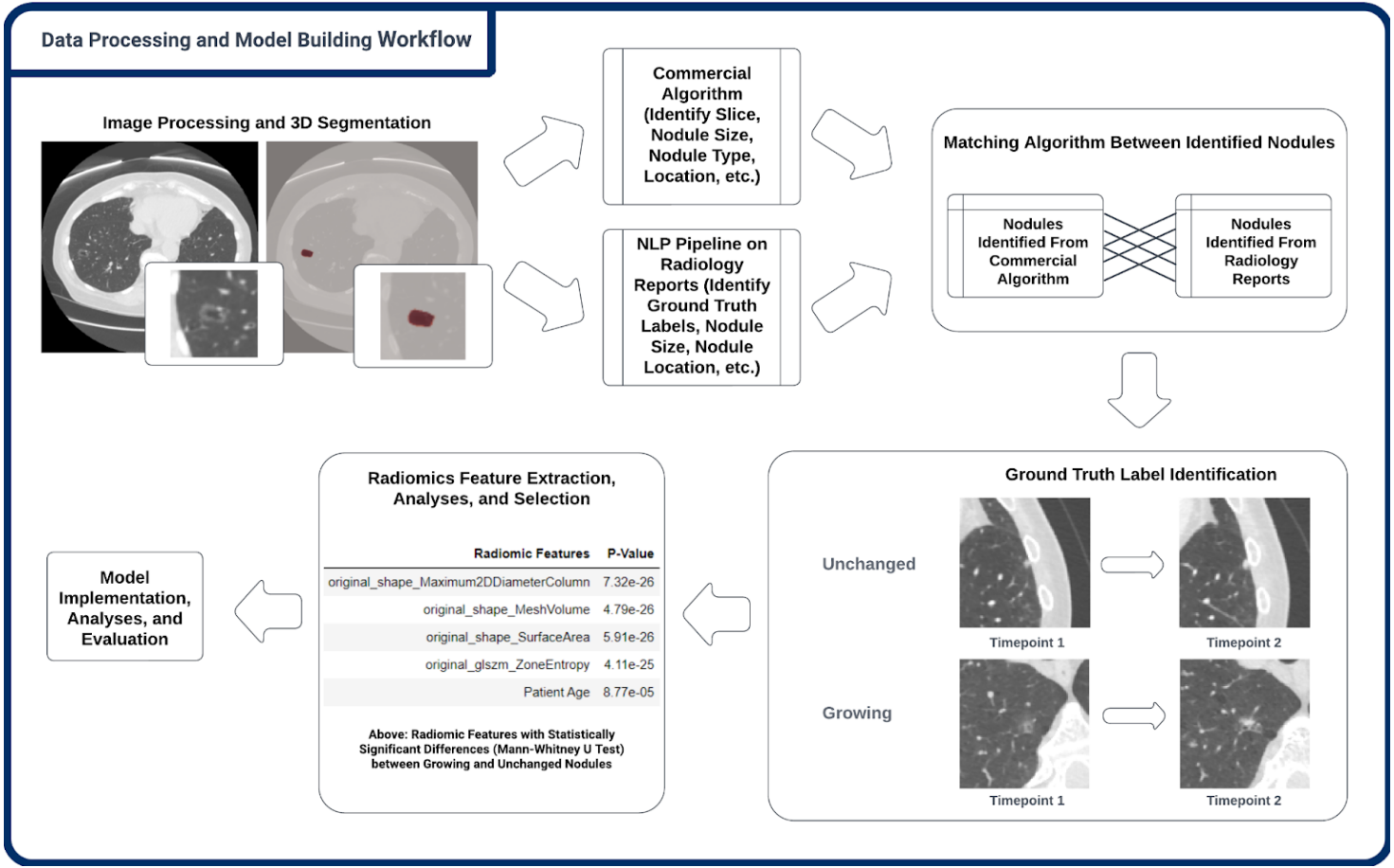
Materials and Methods

- Using the results of the commercial algorithm and the intersection with the radiology report dataset, an image processing algorithm was applied to locate the nodule within the CT scan.
- A 3D bounding box with dimensions (50 px, 50 px, nodule thickness) was generated, centered around the nodule and volume of interest.
- From these bounding boxes, the 3D bounded image and its segmentation mask was extracted. Following this approach, pyRadiomics (v3.1.0), an open-source package for extracting features from 2D and 3D medical images, was employed to generate radiomic features for each nodule.



Pipeline




Data Processing and Model Building Workflow
Image Processing and 3D Segmentation
Commercial Algorithm (Identify Slice, Nodule Size, Nodule Type, Location, etc.)
NLP Pipeline on Radiology Reports (Identify Ground Truth Labels, Nodule Size, Nodule Location, etc.)
Matching Algorithm Between Identified Nodules
Nodules Identified From Commercial Algorithm
Nodules Identified From Radiology Reports
Ground Truth Label Identification
Unchanged
Growing
Timepoint 1
Timepoint 2
Timepoint 1
Timepoint 2
Radiomics Feature Extraction, Analyses, and Selection
Radiomic Features
P-Value
original_shape_Maximum2DDiameterColumn 7.32e-26
original_shape_MeshVolume 4.79e-26
original_shape_SurfaceArea 5.91e-26
original_glszm_ZoneEntropy 4.11e-25
Patient Age 8.77e-05
Above: Radiomic Features with Statistically Significant Differences (Mann-Whitney U Test) between Growing and Unchanged Nodules
Model Implementation, Analyses, and Evaluation

Materials and Methods

Additional Data Processing:

- Due to the incomplete and inconsistent nature of free-text radiology reports and the false-positive identification of acute ground glass nodular opacities by the commercial algorithm, a graduated intersection algorithm was applied to identify those that were most confidently matched on a scale from 1 to 12.
- To ensure greater accuracy and confidence in a nodule's reported growth status, we applied a 180-day rule: nodules reported as unchanged from a previous scan that occurred less than or equal to 180 days ago were removed from the dataset before training.
- Further analysis and review were conducted to ensure that the dataset consisted of unique nodules at any given time point.

Materials and Methods

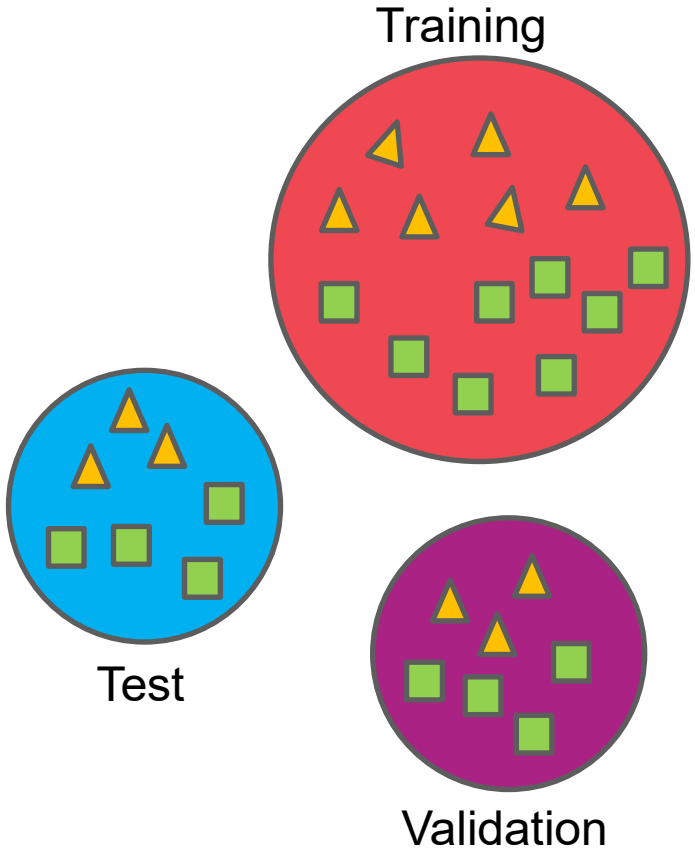
Training, Testing, and Validation

- As observed in many medical datasets, a significant class imbalance dilemma was noted within the dataset.
- Specifically, in the nodule dataset assembled for this study, only approximately 15% (139) of nodules were classified as increasing.

Materials and Methods

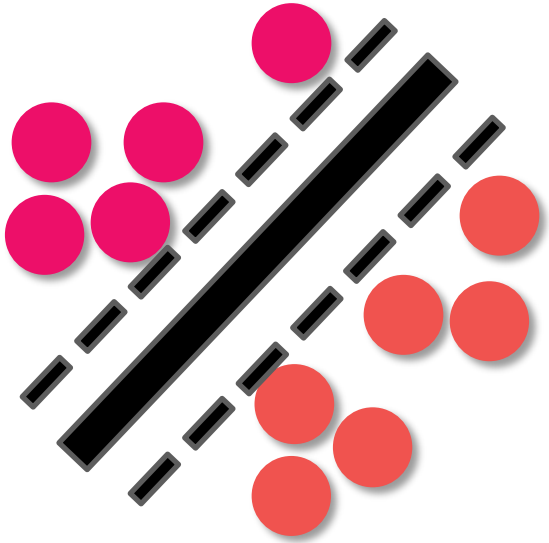
Training, Testing, and Validation

- Therefore, to split the dataset into training, testing, and validation sets, the 'train_test_split' function from scikit-learn was used, which has the ability to stratify based on the output label.
- Furthermore, in order to address the significant class imbalance, 5-fold cross-validation was employed to verify and analyze the model’s results.

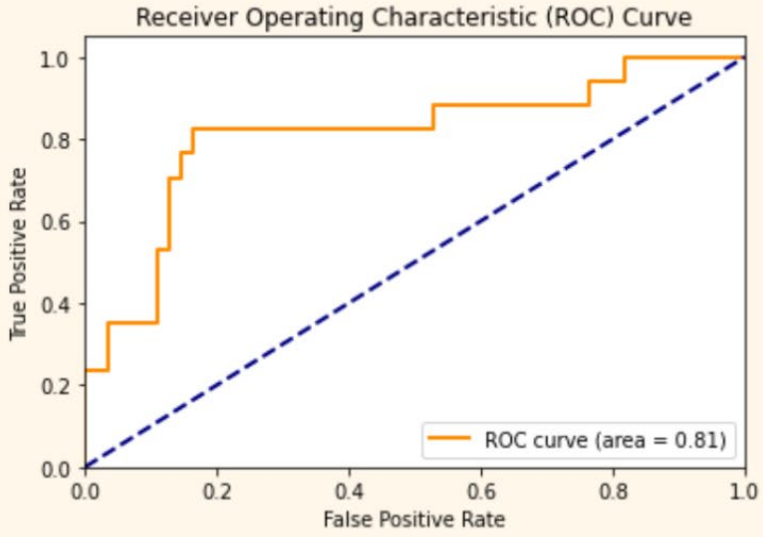


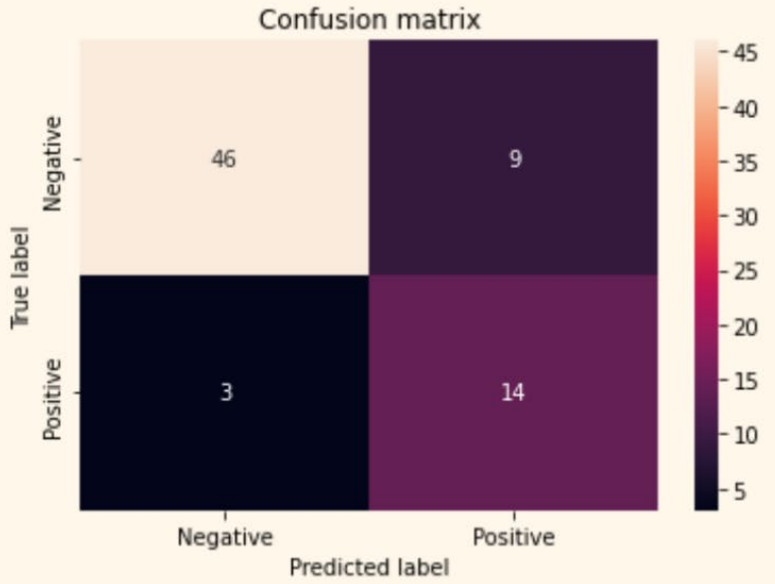
Results

- The intersection dataset of radiology report and commercial algorithm consisted of 1234 subsolid nodules, with ground truth labels (i.e., increase, unchanged, decrease) available for 950 of them.
- There are 799 (~84%) unchanged nodules, 139 (~15%) growing nodules, and 12 (~1%) decreasing nodules.
- We implemented various machine learning models, with SVM (utilizing a radial basis function kernel) emerging as the top-performing model, achieving an AUC of 0.81 on both the 5-fold cross-validation and independent validation sets.



SVM Model Results





*** An SVM model with a radial basis kernel achieved a notable average AUC of 0.81 across both 5-fold cross-validation and an independent validation dataset (left). The confusion matrix (right) was computed using the top-performing cross-validated model on the independent validation set and tuned with a False Positive Rate (FPR) threshold of 0.23, identifying nearly all progressive nodules in the independent validation set.**

Results

In the statistical analysis of the association between radiomic and clinical features and increasing versus unchanged nodules, we determined that 88 radiomic features were statistically significant ($p < 0.05$) based on a Mann-Whitney U Test, with 82 of these radiomic features being highly statistically significant ($p < 0.01$).

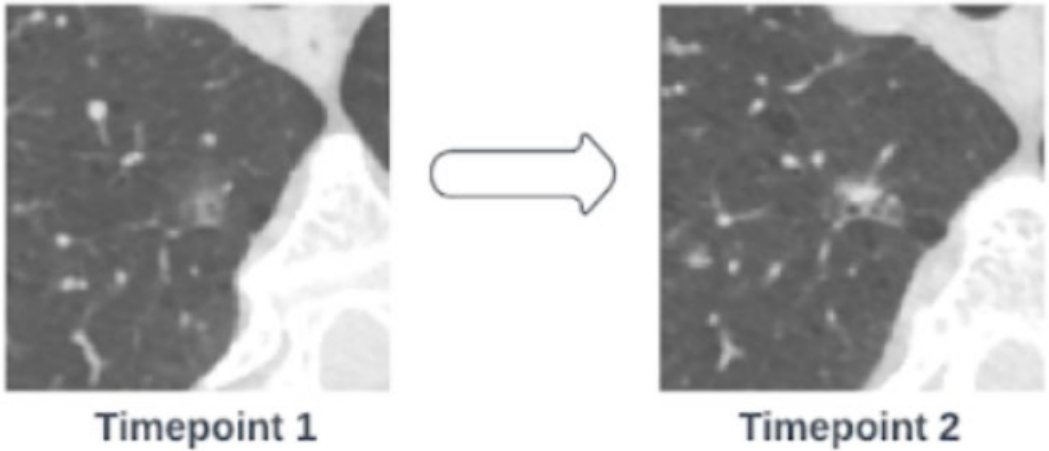
SVM Features

#	Feature Name	Feature Type	Radiomic Family	P-Value †
1	Patient Age	Clinical	N/A	<0.001
2	Maximum 2D Diameter Row	Radiomic	Shape	<0.001
3	Maximum 3D Diameter	Radiomic	Shape	<0.001
4	Maximum 2D Diameter Slice	Radiomic	Shape	<0.001
5	Maximum 2D Diameter Column	Radiomic	Shape	<0.001
6	Surface Volume Ratio	Radiomic	Shape	<0.001
7	Surface Area	Radiomic	Shape	<0.001
8	Voxel Volume	Radiomic	Shape	<0.001
9	Zone Entropy	Radiomic	GLSZM	<0.001
10	Gray Level Non-Uniformity	Radiomic	GLSZM	<0.001
11	Correlation	Radiomic	GLCM	<0.001
12	Inverse Difference Normalized	Radiomic	GLCM	<0.001
13	Run Length Non-Uniformity	Radiomic	GLRLM	<0.001
14	Dependence Entropy	Radiomic	GLDM	<0.001
15	Dependence Non-Uniformity	Radiomic	GLDM	<0.001

*** The above features were used in the SVM model, resulting in an AUC of 0.81 on both independent validation and cross-validation.**
†The P-value is calculated by performing a Mann-Whitney U Test between two groups: increasing nodules and unchanging nodules.

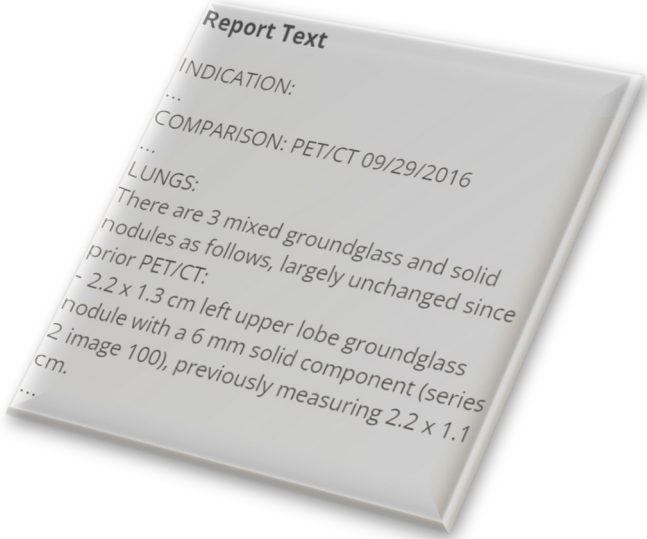
Discussion

- This study underscores the practical importance of integrating radiomic and clinical attributes to predict nodule progression in CT scans — a crucial step in the risk stratification of early neoplasia in the lung.



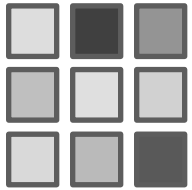
Discussion

- When managing pulmonary nodules, especially in the context of deciding whether surgical intervention is necessary, radiologists frequently refer to established guidelines, such as those provided by the Fleischner Society and Lung-RADS.
- While these guidelines provide a foundational framework for pulmonary nodule management, they also leave room for interpretation, resulting in varying recommendations among radiologists assessing the same nodule.

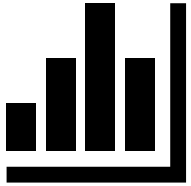


Discussion

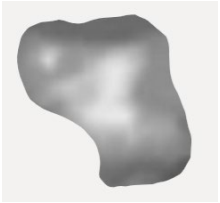
- The interplay between radiomic features, which delve into the intricate characteristics of nodules using advanced imaging analysis, and clinical attributes like patient history and risk factors, can offer a more comprehensive and nuanced understanding of nodule behavior and progression.



Texture



First Order

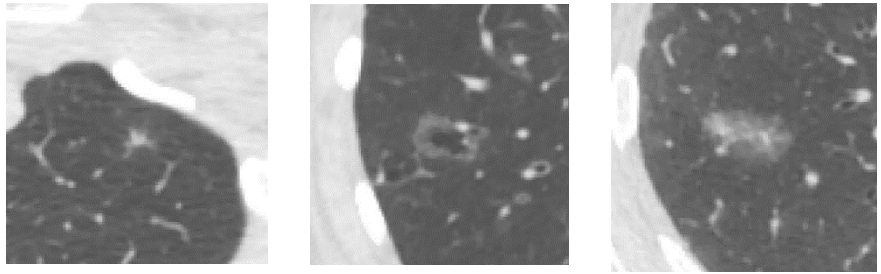


Shape

Discussion

- Our most effective model, employing an SVM with a radial basis kernel, achieved an impressive average AUC of 0.81 in both 5-fold cross-validation and independent validation.
- To underscore the model's effectiveness, it correctly identified 14 out of 17 growing nodules in the independent validation dataset.

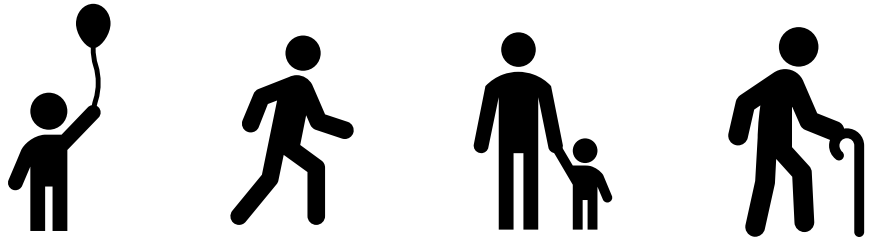
Increasing Nodules →



Discussion

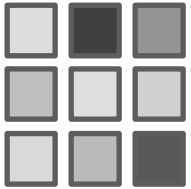
Feature Analysis:

- Clinical Features: We observed that age could serve as a predictor of nodule growth, but whether sex could predict nodule growth is unclear.



- Patient Age was statistically significant ($\alpha = 0.001$) in a Mann-Whitney U Test between increasing and unchanging groups. Sex was not found to be statistically significant in a Chi-Squared Test ($p = 0.411$).

Discussion



GLSZM

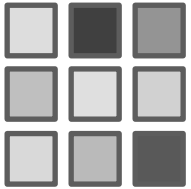
Feature Analysis:

- Radiomic Features: Predictive radiomic features suggested that progressive nodules often appear less homogeneous and more heterogeneous, as well as exhibiting larger sizes. For example:
  - Zone Entropy (ZE): a feature that quantifies the uncertainty and randomness in the distribution of zone sizes and gray levels.

$$ZE = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} p(i,j)\log_2(p(i,j)+\epsilon)$$

  - A higher value indicates more heterogeneity in the texture patterns.
  - For growing nodules, ZE was approximately 528% higher than for stable nodules on average.

Discussion



GLRLM

Following the previous slide:

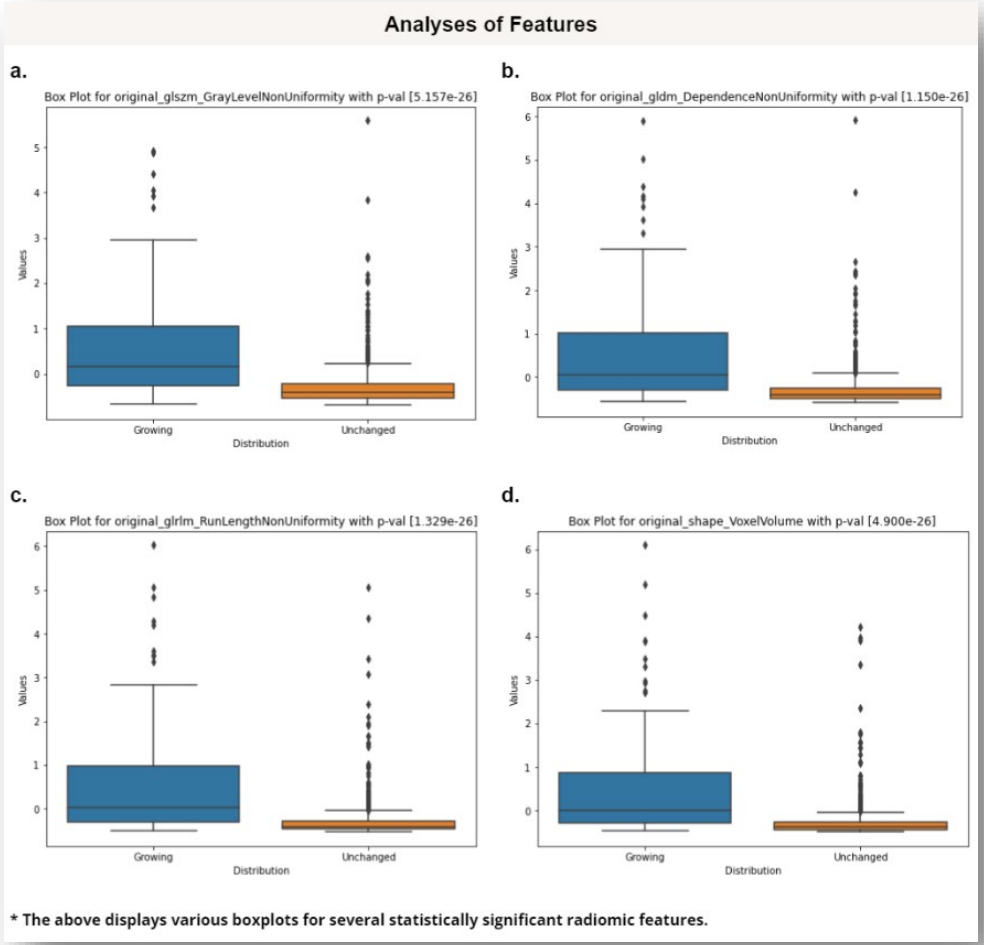
- Run Length Non-Uniformity (RLN): measures the dissimilarity in intensity among consecutive pixels or elements in an image, with lower RLN values indicating that the same color or intensity persists over a relatively long distance without interruption.

$$RLN = \frac{\sum_{j=1}^{N_r} \left(\sum_{i=1}^{N_g} \mathbf{P}(i,j|\theta) \right)^2}{N_r(\theta)}$$

- RLN exhibited an approximately 336% increase for growing nodules compared to stable nodules on average.

Discussion

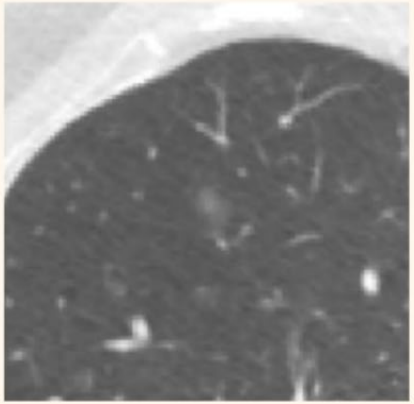
a. GLN: measures the similarity of gray-level intensity values. Lower GLN correlates with greater similarity in intensity values.
b. DNN: measures the similarity of dependence. A lower value indicates more homogeneity among dependencies.
c. RLN: measures the dissimilarity in intensity among consecutive pixels or elements in an image. Lower RLN values indicating that the same color or intensity persists over a relatively long distance without interruption.
d. Voxel Volume: an approximation of nodule volume.



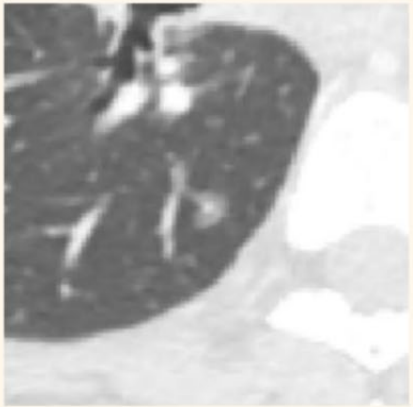
Discussion

Error Analysis

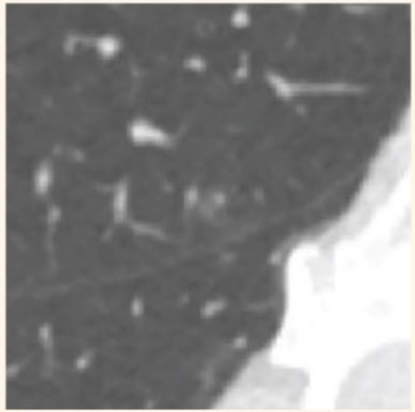
Error Analysis



Average Size: 7.0 mm
Previous Size: 4.0 mm
Location: Right Upper Lobe
Study Date: 2/16/2018
Compared Date: 2/24/2017
Doubling Time: ~442 Days



Average Size: 6.0 mm
Previous Size: N/A
Location: Right Lower Lobe
Study Date: 5/31/2018
Compared Date: 3/19/2007
Doubling Time: N/A



Average Size: 5.0 mm
Previous Size: 3.0 mm
Location: Right Upper Lobe
Study Date: 02/05/2022
Compared Date: 2019
Doubling Time: ~1151 Days

*** Of the 17 increasing cases, the SVM model misclassified 3 growing nodules as stable (false negatives). A Mann-Whitney U Test was conducted to assess the differences between the true positives (14) and the false negatives (3).**
*** Features with P-Values < 0.01: Surface Volume Ratio, Surface Area, Correlation, Dependence Non-Uniformity, Run Length Non-Uniformity, and Voxel Volume.**
*** Features with P-Values < 0.05: Patient Age, Maximum 2d Diameter Row, Maximum 3D Diameter, Maximum 2D Diameter Slice, Maximum 2D Diameter Volume, Zone Entropy, Inverse Difference Normalized, Gray Level Non-Uniformity, Dependence Entropy.**

Discussion

Limitations

- While our study achieved significant success, it also unveiled several limitations inherent in the nature of radiology reports.
- The ground truth label generation process was based solely on the reporting radiologists' assessment within the report text. Thus, these factors have the potential to influence the consistency of the ground truth labels assigned to nodule growth statuses.
- High interreader variability poses significant challenges in achieving consistent and accurate diagnoses of pulmonary nodules.

Discussion

The level of detail provided within radiology reports vary, as shown to the right.

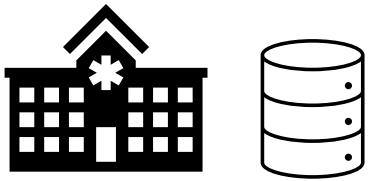
Report Text Feature Extraction		
Detail Level	**Report Text**	**Extracted Features**
High	INDICATION: ... COMPARISON: PET/CT 09/29/2016 ... LUNGS: There are 3 mixed groundglass and solid nodules as follows, largely unchanged since prior PET/CT: - 2.2 x 1.3 cm left upper lobe groundglass nodule with a 6 mm solid component (series 2 image 100), previously measuring 2.2 x 1.1 cm. ...	Accession Number: ######## Overall Status: unchanged Nodule Status: unchanged Component Status: NaN Nodule Type: partial solid Nodule Size: 2.2 x 1.3 cm Previous Nodule Size: 2.2 x 1.1 cm Component Types: solid Component Sizes: 6 mm General Location: left upper lobe Specific Location: NaN Study Date: 3/5/2018 Comparison Date: 2016-09-29 Elapsed Days: 522 Image Number: 100 Average (Report): 17.5 Short (Report): 13.0 Long (Report): 22.0

Discussion

- Compared to existing studies in the literature, our study utilized a large curated dataset of 950 subsolid nodules, all of which were longitudinally confirmed, 3D segmented, and manually verified, although not all pathology confirmed.
- However, arguably, this dataset is subject to much less selection bias than the cases that are surgically resected and pathology confirmed.
- This is because most surgically resected datasets would include nodules that were radiologically and clinically deemed highly suspicious, with patients healthy enough to undergo surgery.
- Moreover, nodules pathologically compatible with adenocarcinoma spectrum lesions do not necessarily indicate growth and invasion in patients, whereas longitudinal observation would arguably represent another potentially more informative way of assessing the risk status of a nodule.

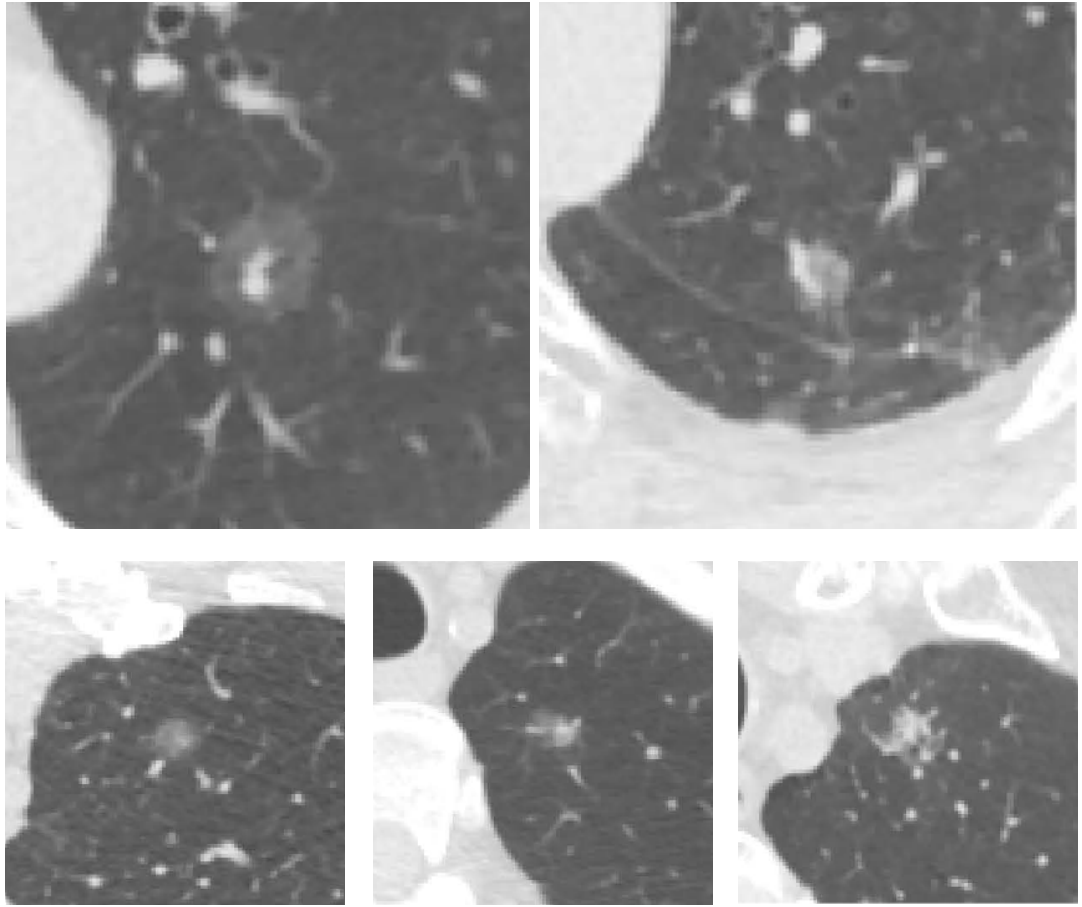
Discussion

- It is also worth noting the challenging aspects of curating large datasets for machine learning and deep-learning-based algorithms, especially in medical contexts.
- Challenges may include the following: limited availability of radiologists, laborious annotation processes, privacy and ethical concerns, heterogeneity of data from various sources, class imbalance issues, and the need to adhere to strict regulatory compliance.



Discussion

- However, by leveraging AI-powered tools, such as automated report generation, image interpretation, and workflow optimization, we may mitigate such issues in data collection.
- In this study, we leveraged the use of computer vision and natural language processing to curate a dataset of 1234 nodules, with 950 labeled nodules for growth status.



Conclusion

- In conclusion, this study developed a predictive model for identifying the overall nodule growth status, leveraging a combination of radiomic and clinical features.
- The model achieved an AUC of 0.81 in both cross-validation and an independent validation set.
- Additionally, we curated a valuable dataset comprising 950 subsolid nodules, complete with segmentations and various attributes, including nodule size, type, and location.
- By employing radiomics, statistical analysis, and machine learning, we may produce accurate early characterization of growing adenocarcinoma spectrum nodules and optimize management and outcomes.

References

[1] “Lung cancer.” Accessed: Nov. 02, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/lung-cancer

[2] J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, “Radiomics in medical imaging—‘how-to’ guide and critical reflection,” Insights into Imaging, vol. 11, no. 1, p. 91, Aug. 2020, doi: 10.1186/s13244-020-00887-2.

[3] J. P. Ko et al., “Pulmonary Nodules: Growth Rate Assessment in Patients by Using Serial CT and Three-dimensional Volumetry,” Radiology, vol. 262, no. 2, pp. 662–671, Feb. 2012, doi: 10.1148/radiol.11100878.

[4] “Evaluation of Individuals With Pulmonary Nodules: When Is It Lung Cancer? - PMC.” Accessed: Nov. 02, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3749714/

[5] “pyradiomics · PyPI.” Accessed: Nov. 02, 2023. [Online]. Available: https://pypi.org/project/pyradiomics/

[6] “Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017 | Radiology.” Accessed: Nov. 02, 2023. [Online]. Available: https://pubs.rsna.org/doi/full/10.1148/radiol.2017161659

[7] A. R. Larici et al., “Lung nodules: size still matters,” European Respiratory Review, vol. 26, no. 146, Dec. 2017, doi: 10.1183/16000617.0025-2017.

[8] E. J. Limkin et al., “The complexity of tumor shape, spiculatedness, correlates with tumor radiomic shape features,” Sci Rep, vol. 9, no. 1, Art. no. 1, Mar. 2019, doi: 10.1038/s41598-019-40437-5.

[9] R. Yang, D. Hui, X. Li, K. Wang, C. Li, and Z. Li, “Prediction of single pulmonary nodule growth by CT radiomics and clinical features — a one-year follow-up study,” Frontiers in Oncology, vol. 12, 2022, Accessed: Nov. 02, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fonc.2022.1034817

References

[10] C.-H. Chen et al., “Radiomic features analysis in computed tomography images of lung nodule classification,” PLoS One, vol. 13, no. 2, p. e0192002, Feb. 2018, doi: 10.1371/journal.pone.0192002.
[11] Y. Sun et al., “Computed tomography radiomics in growth prediction of pulmonary ground-glass nodules,” European Journal of Radiology, vol. 159, p. 110684, Feb. 2023, doi: 10.1016/j.ejrad.2022.110684.
[12] S. R. Digumarthy et al., “Predicting malignant potential of subsolid nodules: can radiomics preempt longitudinal follow up CT?,” Cancer Imaging, vol. 19, no. 1, p. 36, Jun. 2019, doi: 10.1186/s40644-019-0223-7.
[13] M. A. Bruno, E. A. Walker, and H. H. Abujudeh, “Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction,” RadioGraphics, vol. 35, no. 6, pp. 1668–1676, Oct. 2015, doi: 10.1148/rg.2015150023.
[14] S. Bakr et al., “Interreader Variability in Semantic Annotation of Microvascular Invasion in Hepatocellular Carcinoma on Contrast-enhanced Triphasic CT Images,” Radiology: Imaging Cancer, vol. 2, no. 3, p. e190062, May 2020, doi: 10.1148/rycan.2020190062.
[15] S. S. Hsieh et al., “Understanding Reader Variability: A 25-Radiologist Study on Liver Metastasis Detection at CT,” Radiology, vol. 306, no. 2, p. e220266, Feb. 2023, doi: 10.1148/radiol.220266.
[16] L. M. Prevedello et al., “Challenges Related to Artificial Intelligence Research in Medical Imaging and the Importance of Image Analysis Competitions,” Radiology: Artificial Intelligence, vol. 1, no. 1, p. e180031, Jan. 2019, doi: 10.1148/ryai.2019180031.

RSNA® 2023

LEADING THROUGH CHANGE

THANK YOU