

Article

A Multichannel CT and Radiomics-Guided CNN-ViT (RadCT-CNNViT) Ensemble Network for Diagnosis of Pulmonary Sarcoidosis

Jianwei Qiu ^{1,†}, Jhimli Mitra ^{1,*,†}, Soumya Ghose ¹, Camille Dumas ², Jun Yang ², Brion Sarachan ¹, Marc A. Judson ³

¹ GE HealthCare, Niskayuna, NY 12309, USA; jianwei.qiu@gehealthcare.com (J.Q.); soumya.ghose@gehealthcare.com (S.G.); brion.sarachan@gehealthcare.com (B.S.)

² Department of Medical Imaging, Albany Medical College, Albany, NY 12208, USA; dumasc@amc.edu (C.D.); yangj8@amc.edu (J.Y.)

³ Department of Medicine, Albany Medical College, Albany, NY 12208, USA; judsonm@amc.edu

* Correspondence: jhimli.mitra@gehealthcare.com

† These authors contributed equally to this work.

Abstract: Pulmonary sarcoidosis is a multisystem granulomatous interstitial lung disease (ILD) with a variable presentation and prognosis. The early accurate detection of pulmonary sarcoidosis may prevent progression to pulmonary fibrosis, a serious and potentially life-threatening form of the disease. However, the lack of a gold-standard diagnostic test and specific radiographic findings poses challenges in diagnosing pulmonary sarcoidosis. Chest computed tomography (CT) imaging is commonly used but requires expert, chest-trained radiologists to differentiate pulmonary sarcoidosis from lung malignancies, infections, and other ILDs. In this work, we develop a multichannel, CT and radiomics-guided ensemble network (RadCT-CNNViT) with visual explainability for pulmonary sarcoidosis vs. lung cancer (LCa) classification using chest CT images. We leverage CT and hand-crafted radiomics features as input channels, and a 3D convolutional neural network (CNN) and vision transformer (ViT) ensemble network for feature extraction and fusion before a classification head. The 3D CNN sub-network captures the localized spatial information of lesions, while the ViT sub-network captures long-range, global dependencies between features. Through multichannel input and feature fusion, our model achieves the highest performance with accuracy, sensitivity, specificity, precision, F1-score, and combined AUC of 0.93 ± 0.04 , 0.94 ± 0.04 , 0.93 ± 0.08 , 0.95 ± 0.05 , 0.94 ± 0.04 , and 0.97 , respectively, in a five-fold cross-validation study with pulmonary sarcoidosis ($n = 126$) and LCa ($n = 93$) cases. A detailed ablation study showing the impact of CNN + ViT compared to CNN or ViT alone, and CT + radiomics input, compared to CT or radiomics alone, is also presented in this work. Overall, the AI model developed in this work offers promising potential for triaging the pulmonary sarcoidosis patients for timely diagnosis and treatment from chest CT.

Keywords: pulmonary sarcoidosis; chest CT; radiomics; ensemble network; vision transformer; CNN



Citation: Qiu, J.; Mitra, J.; Ghose, S.; Dumas, C.; Yang, J.; Sarachan, B.; Judson, M.A. A Multichannel CT and Radiomics-Guided CNN-ViT (RadCT-CNNViT) Ensemble Network for Diagnosis of Pulmonary Sarcoidosis. *Diagnostics* **2024**, *14*, 1049. <https://doi.org/10.3390/diagnostics14101049>

Academic Editor: Qingqing Mao

Received: 1 April 2024

Revised: 10 May 2024

Accepted: 15 May 2024

Published: 18 May 2024



Copyright: © 2024 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

1.1. Background and Motivation

Pulmonary sarcoidosis is a multisystem granulomatous interstitial lung disease (ILD) with variable presentation and prognosis. Although the disease may involve any organ, the lung is most commonly involved, at a rate of 90% in most series [1,2]. On average, a diagnosis of pulmonary sarcoidosis is made after 3 months of symptoms. In 20% of cases, pulmonary sarcoidosis patients experience symptoms up to 12 months before a diagnosis is made [3]. Currently, the diagnosis of ILD relies on a multidisciplinary approach which includes three major components—clinical presentation, chest imaging, and lung histologic findings [4–7]—wherein, both clinically and radiologically, the disease may

mimic malignancies and infections [8–10]. Although chest-trained radiologists are familiar with the radiographic manifestations of pulmonary sarcoidosis, geographically remote and underserved locations may not have access to such radiologists. Therefore, increasing the speed and diagnostic accuracy of pulmonary sarcoidosis using imaging features has great potential to improve clinically important outcomes by directing these patients to expert care in a timelier fashion.

Both the chest radiograph and chest CT may be used to evaluate for pulmonary sarcoidosis. However, the chest CT scan is vastly superior to the chest radiograph in this regard. Several chest CT scan features are regarded as highly specific for pulmonary sarcoidosis [11–13], and these are often undetectable on the chest radiograph. Although such chest CT features of pulmonary sarcoidosis are regarded as highly specific for the disease and their diagnostic power was demonstrated in small cohorts [14,15], they have not been formally tested in diverse populations. Currently, there is no algorithmic diagnostic tool available that can leverage the characteristic CT findings of pulmonary sarcoidosis other than clinical diagnostic algorithms or guidelines [7,16].

1.2. Related Research and Gaps

Recent studies have shown that use of AI has significantly increased the efficiency of pulmonologists to distinguish respiratory diseases identified on chest CT or radiographs [17–19] with a limited body of research on diagnosing pulmonary sarcoidosis from chest radiographs [20,21]. There is also increasing evidence AI has the potential to democratize radiology by enabling less-experienced radiologists in underserved areas to tap into sub-specialty expertise [22]. Therefore, the development of an AI algorithm that can reliably diagnose pulmonary sarcoidosis at the level of a sub-specialty thoracic radiologist from CT would be a major advancement, an incredible asset to underserved regions, and could serve as a valued assistant for any radiologist.

Radiomics have been used extensively to build methods to automatically diagnose lung diseases and characterize lung nodules (benign vs. malignant) from CT [23–29]. Radiomics are hand-crafted features/mathematical descriptors extracted from radiology images that are relatively straightforward to define, conceptualize, and interpret, and are both standardized and reproducible. These features are used to train a machine learning classifier, and predictions are made based on the trained model. Specifically, radiomics and machine learning approaches have been used to classify or diagnose ILDs [30–34]. On the other hand, CNNs have an inherent capability to learn discriminative features within convolutional blocks for the diagnosis and classification of lung diseases [35–37]. These features are abstract, and it is often difficult to interpret multiscale features that are learned automatically. CNN features are unique to each input dataset, which allows considerable versatility but also introduces susceptibility to overfitting and lack of reproducibility [38].

The extraction of radiomic features typically involves defining a precise region of interest, which is difficult for diffuse lung diseases such as ILDs and pulmonary sarcoidosis, while CNNs operate on entire images or sub-images. Attempts to combine both radiomics and CNNs also have been made in several ways. For example, radiomic features were extracted from CT image and used in a deep learning network, which is an example of early fusion, and then further combined with clinical features at a late stage for the prediction of the EGFR gene mutation status for non-small cell lung carcinoma [39]. Similarly, radiomics features were extracted separately and then combined with features derived from CNN and fused at an intermediate stage before the classification of COPD staging [40] and lung nodule classification [28,41]. All methods that extracted radiomic features, however, depended on defining a region of interest, except in Liang et al. [42], where radiomics features were extracted from the entire lung, although the lung parenchyma was segmented. A comprehensive review of methods involving CNN and/or radiomics for ILDs is provided in Barnes et al. [43].

With more recent advancements in deep learning, vision transformers (ViTs) [44] have become popular in building robust classification models, sometimes outperforming

CNNs [45,46]. The multiheaded, self-attention mechanism in ViT learns rich representations between the sequence of image patches, thus capturing global representation of an image. However, ViT requires large number of labeled images to train, limiting its application in studies with deficient data, particularly with medical imaging data. ViT also emphasizes low-resolution features because of the consecutive downsampling, and this results in the lack of detailed localized information [47]. On the other hand, due to strong inductive bias, CNNs can learn localized features such as edges, corners, and shapes, which may be common across different images, and can often achieve good performance with fewer training samples compared to ViT. To address the limitations of the CNN or ViT frameworks, a recent trend is to combine the ViT and CNN to sample both global and local information in an image for improved classification and segmentation tasks [48–52]. This combination is important for differentiating between diffuse lung diseases such as pulmonary sarcoidosis and others, as there may be some similarity in the local features of the diseases—but the relative position where the features appear within the lung becomes an important differentiating factor in classification of the disease.

1.3. Contributions and Novelty

Based on the distinct advantages of using radiomics, CNN or ViT, in this work, we present a novel approach using radiomics and a CT-guided multichannel CNNViT ensemble classification framework to classify pulmonary sarcoidosis vs. lung cancer (LCa). The novel aspects in our framework are as follows:

- Combination of 3D CNNs with 3D ViT that will allow capturing local information within convolutional blocks and the complex relationship between spatial positions of patches within a CT volume.
- Extraction of radiomic texture features from the chest CT without defining any region of interest, and introducing the multichannel CNNViT network architecture with a radiomic texture map and the CT volume as inputs, thus referring to the framework as RadCT-CNNViT.
- Our framework also provides visual explainability for the classification of pulmonary sarcoidosis vs. lung malignancies (LCa) that suggests regions of interest that are considered important by the network for making the prediction.

Finally, an ablation study is performed to show that our method can leverage the strengths of both hand-crafted radiomics, CT imaging features, and learned CNN+ViT features to provide improved prediction performance compared to a CNN or ViT alone, and radiomics or CT alone.

2. Materials and Methods

In this section, we provide overviews of the data collection process and the preprocessing steps involved in the development of our method. We subsequently explore the multichannel ensemble AI framework comprising a CNN and a ViT architecture, the extraction of radiomic texture features and combining the CNN and ViT architectures for classification. Additionally, we describe the details of the methods utilized to generate visual explanations based on the model predictions. Finally, we discuss the metrics used to evaluate the performance of the presented methods.

2.1. Data and Pre-Processing

The chest CT images for clinically confirmed pulmonary sarcoidosis (PS) (n = 126) were obtained from an IRB-approved study at Albany Medical College (AMC) (refer to the Compliance and Ethical Standards section for details). Chest CT exams for outpatients at AMC were performed using GE Revolution 256 CT scanner and GE VCT Lightspeed 16 slice scanner with a variety of protocols with an in-plane (xy) (512×512 matrix) resolution between 0.625 and 1 mm, and z-resolution of 1.25–5 mm. Patients in the pulmonary sarcoidosis database were, on average, 48.9 years of age (22–84 yrs), female (n = 77), male (n = 48), unspecified (n = 1), and white (n = 101), black (n = 17), Asian (n = 2), unspecified

race (n = 6). Images of lung cancer (LCa) cases (n = 93), comprising both primary (n = 42) and metastatic (n = 51) instances, were sourced from the TCIA (LIDC-IDRI) public archive as described in previous studies [53,54]. The 3D CT volumes were center cropped in axial view to focus on the lung region, and then resized to $256 \times 256 \times 64$. Figure 1 shows some of the chest CT patterns of pulmonary sarcoidosis.

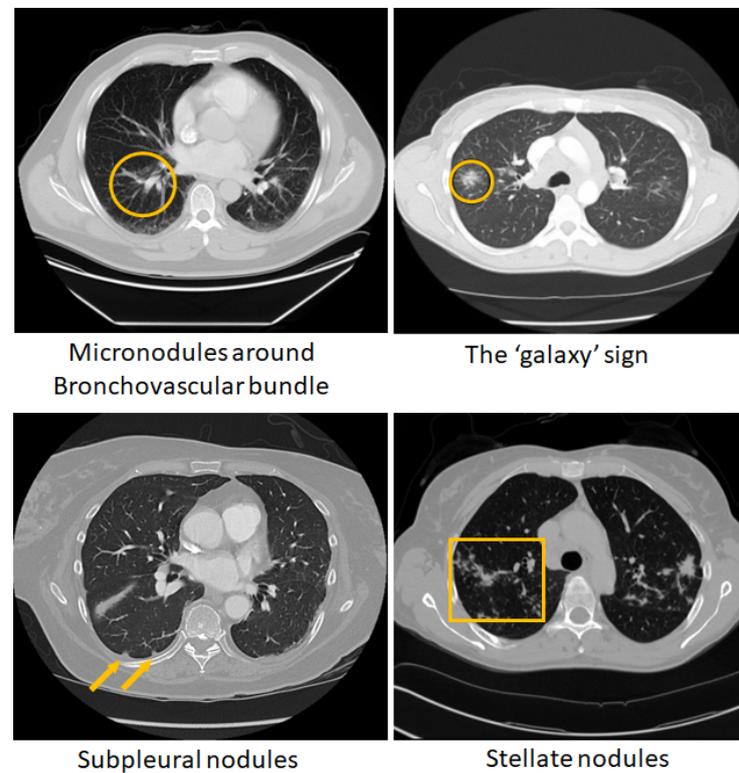


Figure 1. Visible patterns of pulmonary sarcoidosis on chest CT marked in ‘yellow’ circles, arrows and boxes.

2.2. The Multichannel Ensemble AI Framework for Classification

The standalone architectures of the CNN and ViT networks using only the CT volume as input are shown in Figures 2 and 3, respectively. Subsequently, these networks are combined in an ensemble network, incorporating both the CT volume and radiomics feature, to construct the multichannel CT and radiomics-guided CNN-ViT (RadCT-CNNViT) network. The architecture of the RadCT-CNNViT network is illustrated in Figure 4.

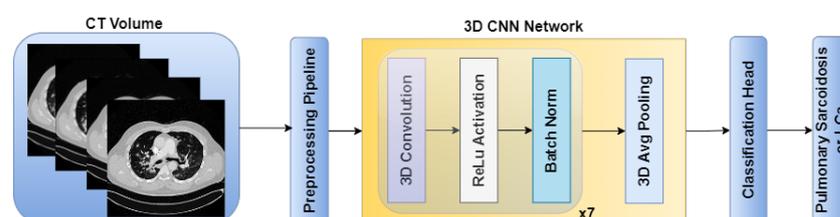


Figure 2. CNN architecture for pulmonary sarcoidosis vs. lung cancer (LCa) classification using chest CT images.

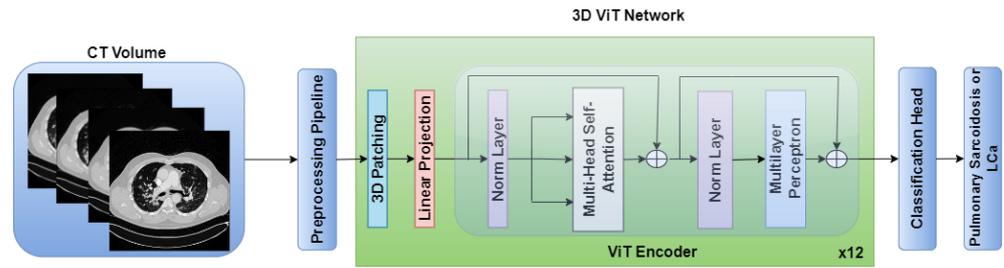


Figure 3. ViT architecture for pulmonary sarcoidosis vs. lung cancer (LCa) classification using chest CT images.

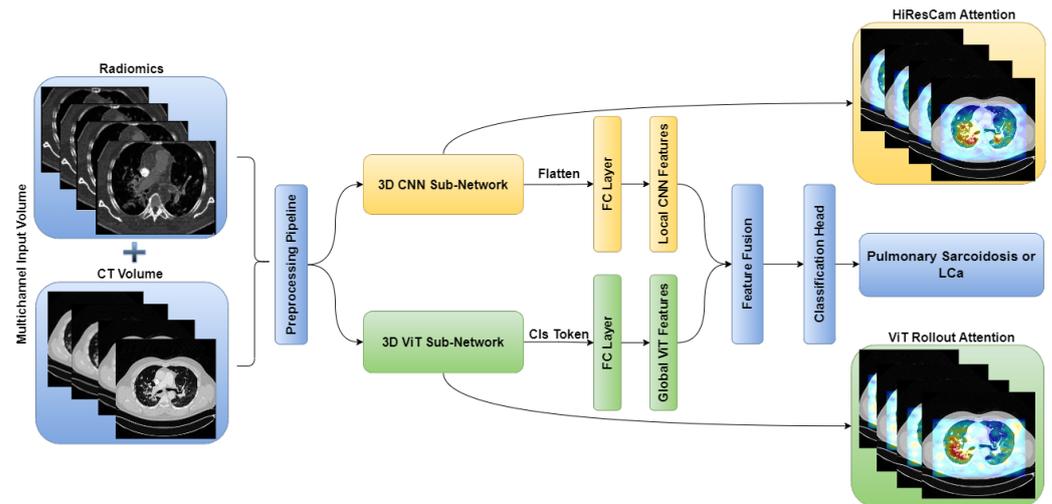


Figure 4. Multichannel RadCT-CNNViT architecture for pulmonary sarcoidosis vs. lung cancer (LCa) classification using chest CT images.

2.2.1. Extracting Radiomics Texture

The input radiomics texture map for the framework was chosen based on our previous work [55], where feature selection was performed using random forest (RF) on a subset of confirmed pulmonary sarcoidosis ($n = 61$) and the MosMed public dataset [56] of other ILDs that were not Covid-19 ($n = 154$). Haralick texture features [57] such as Cluster Prominence, Cluster Shade, Correlation, Energy, Entropy, Haralick Correlation, Inertia, and Inverse Difference Moment with an offset of 1 ($3 \times 3 \times 3$ window) were computed for each CT volume and then averaged to produce one feature map per texture feature. Each radiomic texture volume was then divided into $16 \times 16 \times 16$ patches. Patch mean and standard deviation for each of the 8 texture features were computed, resulting in a feature vector of size 16, and each patch was treated as a sample with the image label. The feature vectors from the patches were used to fit a random forest (RF) classifier [58] with 100 trees, where each patch was classified as pulmonary sarcoidosis or other ILD. The mean decrease in Gini impurity was computed as the average of feature importance scores over all trees in the RF in a 5-fold cross-validation strategy. The feature map corresponding to the highest score was chosen as input to the network architecture. Figure 5 shows all the features and their mean Gini-impurity scores after averaging across 5-folds. Figure 6 shows a case of pulmonary sarcoidosis and its corresponding Haralick correlation texture map.

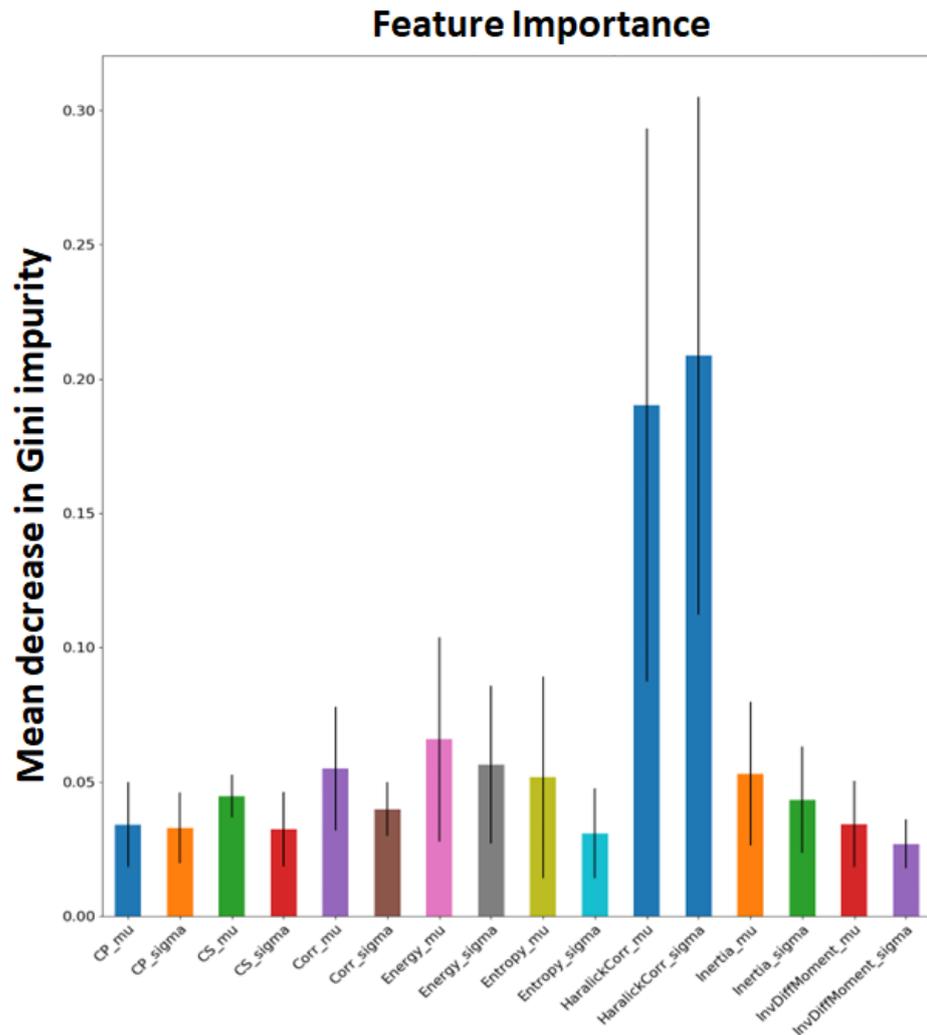


Figure 5. Feature importance was computed based on the mean decrease in Gini impurity for each of the Haralick texture features in discriminating pulmonary sarcoidosis from other ILDs. The mean and standard deviation of the Haralick correlation texture map were higher than those of other texture features.

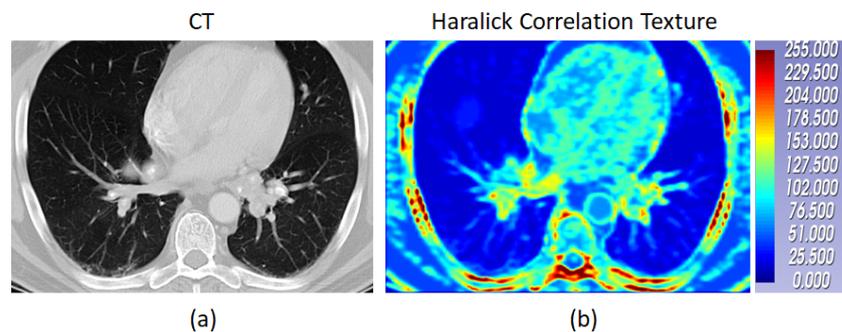


Figure 6. The CT of a case of pulmonary sarcoidosis and its corresponding Haralick correlation texture map are shown in (a) and (b) respectively. The color bar shows the radiomic values normalized between 0 to 255.

2.2.2. The RadCT-CNNViT Architecture

Based on Figure 5, the Haralick correlation maps were computed for pulmonary sarcoidosis and LCa cases and used as input to the RadCT-CNNViT framework with min–max intensity normalization for each 3D texture volume along with the CT volume

clipped to a lung window of the $(-1000, 400)$ intensity range. The RadCT-CNNViT is a 3D multichannel ensemble network, which consists of two input channels feeding into two subnetworks: a 3D CNN feature extractor and a 3D ViT encoder. The 3D CNN feature extractor is responsible for learning local features from the volumetric radiomic and CT feature inputs. It consists of 7 convolution blocks, where each block comprises a 3D convolution layer, ReLU activation, and batch normalization. These convolution blocks employ 3D convolutional filters to capture spatial patterns and extract relevant features from the input data. The numbers of filters in each of the convolution blocks are 16, 32, 64, 128, 256, and 512, respectively. The first convolution block utilizes a kernel size of $3 \times 3 \times 3$ and a stride number of 1. For downsampling, the subsequent convolution blocks use a kernel size of $4 \times 4 \times 4$ and a stride number of 2. The last layer of the 3D CNN is followed by a 3D average pooling operation and fully connected layer, which help to reduce the spatial dimensions of the CNN features to 768.

On the contrary, the 3D ViT encoder focuses on capturing global features by treating the input as a sequence of 3D patches, each with a size of $16 \times 16 \times 16$. The 3D ViT encoder consists of 12 transformer blocks with a hidden layer dimension of 768, and each block utilizes multihead self-attention with 6 heads. The outputs from the 3D CNN feature extractor and the 3D ViT encoder are finally concatenated and fed into a fully connected (FC) layer with sigmoid activation for the classification of pulmonary sarcoidosis vs. LCa. Binary cross-entropy was used as a loss function with AdamW optimization; a learning rate of 1×10^{-5} and 50 epochs were used to train the network. Additionally, we utilized random flip, random noise, and random affine transformations from TorchIO [59], a Python library designed for medical imaging augmentation, to augment the 3D data during training. An overview of the combined CNN-ViT network architecture is shown in Figure 4.

2.3. Generating Visual Explanations for Predictions

To generate visual explanations and enhance the interpretability of our model's predictions, we applied two techniques: HiResCAM [60] and Attention Rollout [61]. These methods offer crucial insights by generating visual attention maps for both CNN and ViT sub-networks, particularly beneficial for understanding complex deep learning models applied to medical imaging, such as chest CT scans. The overarching goal is to localize relevant disease features within the chest CT volume.

HiResCAM utilizes attention mechanisms to selectively weigh the contributions of different features within the CNN subnetwork. The computation of HiResCAM is described by Equation (1). The process begins by computing the gradient of the raw score s_m corresponding to class m with respect to a specific CNN feature map A . This gradient, represented as $\frac{\partial s_m}{\partial A}$, highlights the significance of various features in influencing the prediction. Subsequently, an attention map is generated by element-wise multiplication between the computed gradient and the CNN feature map, followed by summation over the feature dimension F . This attention map \tilde{A} provides visual cues, aiding in the localization of relevant disease features within the chest CT volume:

$$\tilde{A}_m^{HiResCAM} = \sum_{f=1}^F \frac{\partial s_m}{\partial A} \odot A^f \quad (1)$$

In contrast, Attention Rollout offers a distinct approach by tracing the path of attention from an initial region of interest to all other patches in the image. This recursive method dynamically visualizes how the ViT sub-network distributes its attention across different parts of the image. By quantifying the attention flow, Attention Rollout provides profound insights into how the ViT sub-network distributes its attention across various parts of the image, facilitating a deeper understanding of the underlying mechanisms driving predictions. The computation of Attention Rollout at layer L is described by Equation (2), where A_L represents the average of the multihead self-attention matrix at layer L , and I denotes the identity matrix:

$$AttentionRollout_L = (A_L + I)AttentionRollout_{L-1} \quad (2)$$

2.4. Performance Metrics

The performance metrics for evaluation of all methods in this ablation study included sensitivity, specificity, precision, accuracy, F1-score, and combined AUC, computed across 5-folds of cross-validation. These metrics were computed based on a confusion matrix which contains four parameters: *TP* (true positive), *TN* (true negative), *FP* (false positive), and *FN* (false negative). *TP* indicates correctly predicted pulmonary sarcoidosis, *TN* denotes correctly predicted LCa, *FP* represents incorrectly predicted pulmonary sarcoidosis, and *FN* indicates incorrectly predicted LCa. Sensitivity, specificity, precision, accuracy, and F1-score values were derived from these parameters using Equations (3)–(7):

$$Sensitivity = Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$F1-Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \quad (7)$$

3. Experiments and Results

We conducted a comprehensive ablation study to evaluate the performance of different network architectures (CNN, ViT, and CNNViT) using CT, radiomics, and multichannel CT-radiomics data. In this study, we performed a five-fold cross-validation with a dataset of clinically confirmed cases of pulmonary sarcoidosis ($n = 126$) and lung cancer ($n = 93$). Figure 7 illustrates the training and validation loss curves of a single fold over 50 epochs for all different methods compared. It demonstrates that 3D ViT failed to converge due to the limited training dataset, while 3D CNN showed slower convergence with unstable loss. Conversely, 3D CNN-ViT ensemble network demonstrated improved convergence due to the combination of global and local features. Moreover, RadCT-CNNViT achieved the lowest loss and best converged training and validation losses in differentiating the diseases, further demonstrating the effectiveness of leveraging radiomics texture maps as input along with CT. The normalized confusion matrices for all experiments in this ablation study are shown in Figure 8. The confusion matrices show that the true prediction rate for pulmonary sarcoidosis increased higher than LCa when the CNN and ViT networks were combined, suggesting the value of combining global and local features for pulmonary sarcoidosis. Performance metrics for these experiments were derived from the confusion matrices and are summarized in Table 1. Additionally, the corresponding ROC curves are depicted in Figure 9. The RadCT-CNNViT model demonstrated the best performance, with accuracy, sensitivity, specificity, precision, F1-score, and combined AUC of 0.93 ± 0.04 , 0.94 ± 0.04 , 0.93 ± 0.08 , 0.95 ± 0.05 , 0.94 ± 0.04 and 0.97 , respectively, compared to other variations in the ablation study, with statistical significance of $p < 0.0001$.

Figure 10 shows detailed visual explanations utilizing HiResCAM and ViT Attention Rollout techniques for both pulmonary sarcoidosis and LCa. The computed visual attention maps are overlaid onto the CT images to emphasize the regions of interest. We observed that features from the CNN subnetworks had denser visual representations (color maps) within local regions, while ViT showed overall global representations as expected. These

visual cues highlight features in specific regions of interest contributing to pulmonary sarcoidosis and lung cancer diagnoses.

Table 1. Performance statistics (Sensitivity, Specificity, Accuracy and combined area under curve (AUC)) for CT-ViT, CT-CNN, CT-CNNViT, Rad-CNNViT, and RadCT-CNNViT.

Network	Sensitivity	Specificity	Precision	Accuracy	F1-Score	AUC
CT-ViT	0.68 ± 0.09	0.66 ± 0.02	0.72 ± 0.08	0.67 ± 0.05	0.70 ± 0.08	0.67
CT-CNN	0.83 ± 0.04	0.88 ± 0.05	0.89 ± 0.06	0.85 ± 0.04	0.86 ± 0.05	0.84
CT-CNNViT	0.87 ± 0.05	0.89 ± 0.06	0.92 ± 0.05	0.88 ± 0.04	0.89 ± 0.05	0.92
Rad-CNNViT	0.88 ± 0.06	0.77 ± 0.09	0.84 ± 0.06	0.84 ± 0.05	0.86 ± 0.06	0.86
RadCT-CNNViT	0.94 ± 0.04	0.93 ± 0.08	0.95 ± 0.05	0.93 ± 0.04	0.94 ± 0.04	0.97

Bold values indicate best performance.

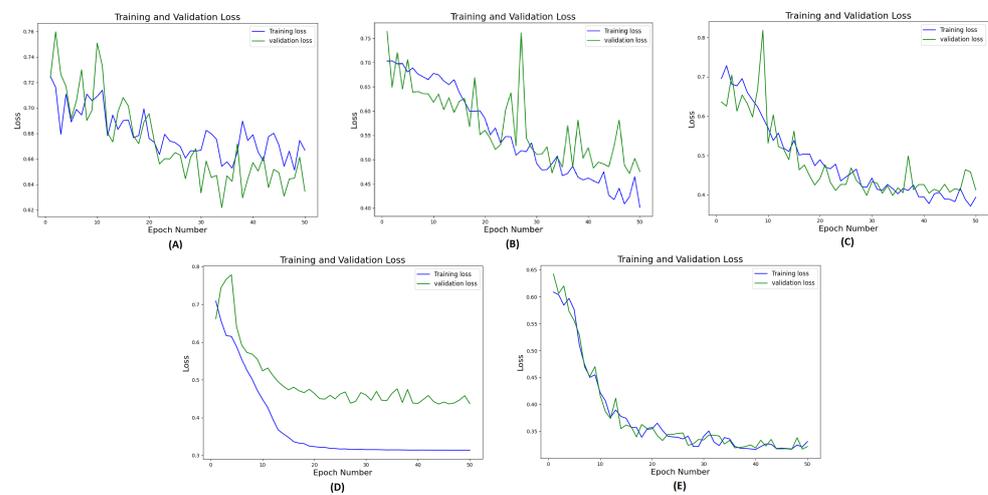


Figure 7. Training and validation loss curves (one-fold) for 50 epochs for the methods in ablation study: (A) CT-ViT, (B) CT-CNN, (C) CT-CNNViT, (D) Rad-CNNViT, (E) RadCT-CNNViT.

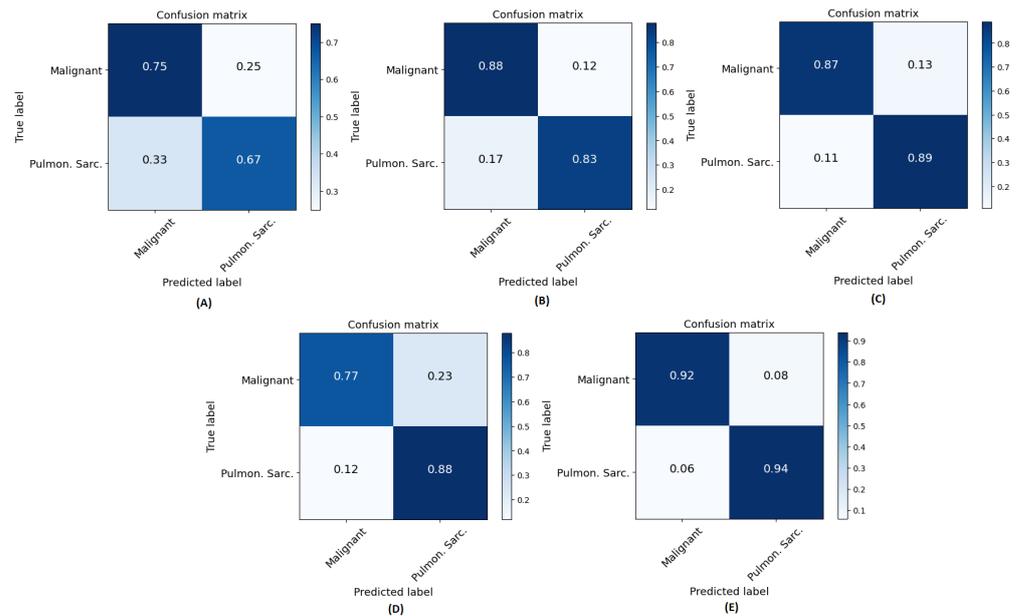


Figure 8. Normalized confusion matrices for all methods across all folds: (A) CT-ViT, (B) CT-CNN, (C) CT-CNNViT, (D) Rad-CNNViT, and (E) RadCT-CNNViT. ‘Pulmon. Sarc.’ in axes labels is the abbreviation for pulmonary sarcoidosis and ‘malignant’ relates to LCa.

classification layer, which demonstrated improved performance compared to CNN features or radiomics features based classification only [39,40,62]. In our previous work [55], we showed how radiomics texture features used as input to a CNN-ViT framework had improved performance over using radiomics features with a traditional machine learning classifier to classify pulmonary sarcoidosis from other ILDs.

In this work, we show that compared to a CNN or ViT alone, or using CT or radiomics only in a CNN-ViT ensemble network, a CT and radiomics-guided deep learning approach provides improved feature representation. Specifically, it highlights the effectiveness of feature fusion, in both early and intermediate stages, i.e., the proper utilization of radiomics texture maps, which are also 3D volumes extracted from CT as input features along with CT imaging features, and combining features extracted from CNN and ViT sub-networks before classification. The strong inductive bias of CNNs is necessary to reach the desired classification accuracy with less data. However, for diffuse lung diseases with no specific location within the lung, the global, long-range context offered by ViT is more adept at identifying/embedding interactions between image patches. Unfortunately, ViT does not provide as much local context compared to CNNs. Nevertheless, the problem of precisely embedding the local and global representations into one another remains. Hence, in this work, a dual structure of CNN-ViT is created to capture the respective feature representations for enhanced representation learning.

Radiomics texture features are computationally well defined compared to abstract, hierarchical, and difficult-to-interpret CNN features. Haralick texture (correlation map), used in this work, captures features from the CT images that are not perceptible for the human eye [63]. In essence, it describes how often one gray tone will appear in a specified spatial relationship to another gray tone on the image [64]. As a result, subtle differentiation between different granulomatous disorders such as between the 'galaxy sign' of pulmonary sarcoidosis and that mimicking metastatic lung cancer [65,66] is possible using such radiomics texture features. Our experiments showed that although the inclusion of radiomics feature as multichannel input with CT did improve all performance metrics in differentiating pulmonary sarcoidosis from LCa, neither CT nor radiomics alone could provide similar accuracies, leading to the confirmation of the hypothesis that the radiomics texture indeed was complementary to CT imaging features. However, we acknowledge that an extensive set of radiomics texture maps was not computed in our experiments, and only the Haralick textures were computed, which is a limitation of this work. The types of radiomics texture features are myriad, and the computation of all features, down-selecting the best features to remove noisy representations, is an intensive process. In future, we plan to include transform-based texture features in our experiments. The radiomics texture map extraction for pulmonary sarcoidosis or LCa did not involve the annotation of regions of interest in the CT volume, which makes our AI framework further suitable for differentiating between other types of diffuse lung diseases.

The major limitations of this work include training and validation using a cross-validation approach due to the limited sample size, and unavailability of a separate validation set from a multicenter study, which may affect the generalizability of the method. However, this being a pilot study to choose the best performing method between the radiomics, CNN, and ViT combinations, the improvement of one method over the other is observed in the results without testing on a separate cohort. Additionally, our method only addresses a two-class problem, i.e., diagnosing pulmonary sarcoidosis vs. LCa; however, in clinical settings, differentiating between pulmonary sarcoidosis and other forms of ILDs would be necessary, which is part of our future work.

We also acknowledge that the presented RadCT-CNNViT is complex in terms of training the network, as it requires a GPU compute; we do not, however, think this limits the adoption of the method in a clinical setting, as, based on our experience, 3D network inferences can be often performed on CPUs with advanced Intel optimization techniques. Additionally, we used a vanilla CNN and a standard ViT network in our implementation without trying different CNN or Vision Transformer versions such as ResNets or Swin

Transformers because of the limited training sample size, as complex deep learning networks need a lot more data for model convergence during training. Applying individually pretrained CNNs or ViTs to mitigate the limited training data issue was not in the scope of this work, as pretrained networks are 2D and a variety of approaches for transfer learning may be taken into consideration for 3D medical images, and choosing the right strategy for utilizing such models for disease diagnosis depends on network design choices, and on the similarity and the amount of the dataset [67]. One of the hypotheses of this work was to show that while ViTs provide rich global feature representations, they do not outperform CNNs in a low-data setting, and by combining CNN and ViT, a reasonably acceptable classification performance can be achieved. In addition, the combination of local and global features is clinically relevant in the diagnosis of pulmonary sarcoidosis. Although, there is no prior literature on the sensitivity and specificity of diagnosing pulmonary sarcoidosis from chest CT, prior work [15] suggests that the performance of our method is similar to that of expert radiologists in diagnosing cardiac sarcoidosis with pulmonary and mediastinal involvement. The performance of our method is also higher than previous works that used chest X-ray to diagnose pulmonary sarcoidosis from healthy patients [21] or patients with pneumonia [20] involving deep learning or radiomics respectively with much smaller cohorts.

5. Conclusions

Pulmonary sarcoidosis is a diffuse lung disease, which is difficult to diagnose from CT imaging without a multidisciplinary clinical team, specifically in geographically underserved locations. The visual attention maps and intelligent network architecture from CNN and ViT used in our method are likely to reduce the burden of radiologists and provide a timely and reliable probability of pulmonary sarcoidosis diagnoses. Clinicians may also use this information directly to adjust their diagnostic probabilities in patients with diffuse lung disease. As our AI method to diagnose pulmonary sarcoidosis does not depend on input from radiologists, it may truly augment the radiologist's impression, as the approaches of the radiologist and our method most probably will be different. This suggests that our method may not only increase the speed of the radiographic assessment of diffuse lung disease but may surpass current chest imaging diagnostic standards. Finally, although our method was applied to pulmonary sarcoidosis in this instance, it could be adapted to any interstitial lung disease. We therefore believe that our method ultimately has the capability to be used as a general diagnostic tool for all interstitial lung diseases as well as localized lung diseases.

Author Contributions: Conceptualization, J.M., J.Q. and S.G.; methodology, J.M. and J.Q.; software, J.Q., J.M. and S.G.; validation, J.Q., C.D. and M.A.J.; formal analysis, J.M.; investigation, M.A.J.; data curation, J.Y., C.D. and B.S.; writing—original draft preparation, J.Q. and J.M.; writing—review and editing, M.A.J., C.D., S.G., J.Y. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This research study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Institutional Review Board of Albany Medical Center, study number 6039, approved 7 December 2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Retrospective data for pulmonary sarcoidosis were collected at Albany Medical College and cannot be shared publicly.

Acknowledgments: The authors acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health, and their critical role in the creation of the free publicly available LIDC/IDRI Database used in this study.

Conflicts of Interest: Mitra, J., Qiu, J., Ghose, S. and Sarachan, B are employees of GE HealthCare. Authors declare that research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

References

1. Judson, M.A.; Boan, A.D.; Lackland, D.T. The clinical course of sarcoidosis: Presentation, diagnosis, and treatment in a large white and black cohort in the United States. *Sarcoidosis Vasc. Diffus. Lung Dis. Off. J. Wasog* **2012**, *29*, 119–127.
2. Baughman, R.P.; Teirstein, A.S.; Judson, M.A.; Rossman, M.D.; Yeager, H.J.; Bresnitz, E.A.; DePalo, L.; Hunninghake, G.; Iannuzzi, M.C.; Johns, C.J.; et al. Clinical characteristics of patients in a case control study of sarcoidosis. *Am. J. Respir. Crit. Care Med.* **2001**, *164*, 1885–1889. [[CrossRef](#)] [[PubMed](#)]
3. Judson, M.A.; Thompson, B.W.; Rabin, D.L.; Steimel, J.; Knattreud, G.L.; Lackland, D.T.; Rose, C.; Rand, C.S.; Baughman, R.P.; Teirstein, A.S. The diagnostic pathway to sarcoidosis. *Chest* **2003**, *123*, 406–412. [[CrossRef](#)] [[PubMed](#)]
4. Crouser, E.D.; Maier, L.A.; Wilson, K.C.; Bonham, C.A.; Morgenthau, A.S.; Patterson, K.C.; Abston, E.; Bernstein, R.C.; Blankstein, R.; Chen, E.S.; et al. Diagnosis and detection of sarcoidosis. An official American Thoracic Society clinical practice guideline. *Am. J. Respir. Crit. Care Med.* **2020**, *201*, e26–e51. [[CrossRef](#)] [[PubMed](#)]
5. Teoh, A.K.Y.; Holland, A.E.; Morisset, J.; Flaherty, K.R.; Wells, A.U.; Walsh, S.L.F.; Glaspole, I.; Wuyts, W.A.; Corte, T.J.; Collaborators, I.M.D. Essential Features of an Interstitial Lung Disease Multidisciplinary Meeting: An International Delphi Survey. *Ann. Am. Thorac. Soc.* **2022**, *19*, 66–73. [[CrossRef](#)] [[PubMed](#)]
6. Lee, C.T. Multidisciplinary Meetings in Interstitial Lung Disease: Polishing the Gold Standard. *Ann. Am. Thorac. Soc.* **2022**, *19*, 7–9. [[CrossRef](#)] [[PubMed](#)]
7. Grutters, J.C. Establishing a Diagnosis of Pulmonary Sarcoidosis. *J. Clin. Med.* **2023**, *12*, 6898. [[CrossRef](#)] [[PubMed](#)]
8. van't Hoog, A.H.; Meme, H.K.; Van Deutekom, H.; Mithika, A.M.; Olunga, C.; Onyino, F.; Borgdorff, M.W. High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. *Int. J. Tuberc. Lung Dis.* **2011**, *15*, 1308–1314. [[CrossRef](#)] [[PubMed](#)]
9. Mortaz, E.; Adcock, I.M.; Barnes, P.J. Sarcoidosis: Role of non-tuberculosis mycobacteria and Mycobacterium tuberculosis. *Int. J. Mycobacteriol.* **2014**, *3*, 225–229. [[CrossRef](#)]
10. El Jammal, T.; Pavic, M.; Gerfaud-Valenti, M.; Jamilloux, Y.; Sève, P. Sarcoidosis and Cancer: A Complex Relationship. *Front. Med.* **2020**, *24*, 594118. [[CrossRef](#)]
11. Abehsera, M.; Valeyre, D.; Grenier, P.; Jaillet, H.; Battesti, J.P.; Brauner, M.W. Sarcoidosis with pulmonary fibrosis: CT patterns and correlation with pulmonary function. *AJR Am. J. Roentgenol.* **2000**, *174*, 1751–1757. [[CrossRef](#)] [[PubMed](#)]
12. Tana, C.; Donatiello, I.; Coppola, M.G.; Ricci, F.; Maccarone, M.T.; Ciarambino, T.; Cipollone, F.; Giamberardino, M.A. CT Findings in Pulmonary and Abdominal Sarcoidosis. Implications for Diagnosis and Classification. *J. Clin. Med.* **2020**, *9*, 3028. [[CrossRef](#)] [[PubMed](#)]
13. Nakatsu, M.; Hatabu, H.; Morikawa, K.; Uematsu, H.; Ohno, Y.; Nishimura, K.; Nagai, S.; Izumi, T.; Konishi, J.; Itoh, H. Large coalescent parenchymal nodules in pulmonary sarcoidosis: “sarcoid galaxy” sign. *AJR Am. J. Roentgenol.* **2002**, *178*, 1389–1393. [[CrossRef](#)] [[PubMed](#)]
14. Koide, T.; Saraya, T.; Tsukahara, Y.; Bonella, F.; Börner, E.; Ishida, M.; Ogawa, Y.; Hirukawa, I.; Oda, M.; Shimoda, M.; et al. Clinical significance of the “galaxy sign” in patients with pulmonary sarcoidosis in a Japanese single-center cohort. *Sarcoidosis Vasc. Diffus. Lung Dis.* **2016**, *33*, 247–252.
15. Russo, J.J.; Nery, P.B.; Ha, A.C.; Healey, J.S.; Juneau, D.; Rivard, L.; Friedrich, M.G.; Gula, L.; Wisenberg, G.; deKemp, R.; et al. Sensitivity and specificity of chest imaging for sarcoidosis screening in patients with cardiac presentations. *Sarcoidosis Vasc. Diffus. Lung Dis.* **2019**, *36*, 18–24.
16. Judson, M.A.; Costabel, U.; Drent, M.; Wells, A.; Maier, L.; Koth, L.; Shigemitsu, H.; Culver, D.A.; Gelfand, J.; Valeyre, D.; et al. The WASOG Sarcoidosis Organ Assessment Instrument: An update of a previous clinical tool. *Sarcoidosis Vasc. Diffus. Lung Dis. Off. J. Wasog* **2014**, *31*, 19–27.
17. İn, E.; Geçkil, A.A.; Kavuran, G.; Şahin, M.; Berber, N.K.; Kuluöztürk, M. Using artificial intelligence to improve the diagnostic efficiency of pulmonologists in differentiating COVID-19 pneumonia from community-acquired pneumonia. *J. Med. Virol.* **2022**, *94*, 3698–3705. [[CrossRef](#)] [[PubMed](#)]
18. Kaplan, A.; Cao, H.; FitzGerald, J.M.; Iannotti, N.; Yang, E.; Kocks, J.W.H.; Kostikas, K.; Price, D.; Reddel, H.K.; Tsiligianni, I.; et al. Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis. *J. Allergy Clin. Immunol. Pract.* **2021**, *9*, 2255–2261. [[CrossRef](#)] [[PubMed](#)]
19. Chan, J.; Auffermann, W.F. Artificial Intelligence in the Imaging of Diffuse Lung Disease. *Radiol. Clin.* **2022**, *60*, 1033–1040. [[CrossRef](#)]
20. Baghdadi, N.; Maklad, A.S.; Malki, A.; Deif, M.A. Reliable Sarcoidosis Detection Using Chest X-rays with EfficientNets and Stain-Normalization Techniques. *Sensors* **2022**, *22*, 3846. [[CrossRef](#)]
21. Prokop, P. Computer-aided Diagnosis of Sarcoidosis Based on X-Ray Images. *Procedia Comput. Sci.* **2023**, *225*, 4611–4620. [[CrossRef](#)]
22. Langlotz, C.P. Will Artificial Intelligence Replace Radiologists? *Radiol. Artif. Intell.* **2019**, *1*, e190058. [[CrossRef](#)] [[PubMed](#)]

23. Frix, A.N.; Cousin, F.; Refaee, T.; Bottari, F.; Vaidyanathan, A.; Desir, C.; Vos, W.; Walsh, S.; Occhipinti, M.; Lovinfosse, P.; et al. Radiomics in Lung Diseases Imaging: State-of-the-Art for Clinicians. *J. Pers. Med.* **2021**, *11*, 602–621. [[CrossRef](#)]
24. Padmakumari, L.T.; Guido, G.; Caruso, D.; Nacci, I.; Gaudio, A.D.; Zerunian, M.; Polici, M.; Gopalakrishnan, R.; Mohamed, A.K.S.; De Santis, D.; et al. The Role of Chest CT Radiomics in Diagnosis of Lung Cancer or Tuberculosis: A Pilot Study. *Diagnostics* **2022**, *12*, 739. [[CrossRef](#)]
25. Hunter, B.; Chen, M.; Ratnakumar, P.; Alemu, E.; Logan, A.; Linton-Reid, K.; Tong, D.; Senthivel, N.; Bhamani, A.; Bloch, S.; et al. A radiomics-based decision support tool improves lung cancer diagnosis in combination with the Herder score in large lung nodules. *eBioMedicine* **2022**, *86*, 104344. [[CrossRef](#)]
26. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **2020**, *295*, 328–338. [[CrossRef](#)] [[PubMed](#)]
27. Wu, Y.J.; Wu, F.Z.; Yang, S.C.; Tang, E.K.; Liang, C.H. Radiomics in Early Lung Cancer Diagnosis: From Diagnosis to Clinical Decision Support and Education. *Diagnostics* **2022**, *12*, 1064. [[CrossRef](#)] [[PubMed](#)]
28. Astaraki, M.; Yang, G.; Zakko, Y.; Toma-Dasu, I.; Smedby, O.; Wang, C. A Comparative Study of Radiomics and Deep-Learning Based Methods for Pulmonary Nodule Malignancy Prediction in Low Dose CT Images. *Front. Oncol.* **2021**, *11*, 737368. [[CrossRef](#)]
29. Jing, R.; Wang, J.; Li, J.; Wang, X.; Li, B.; Xue, F.; Shao, G.; Xue, H. A wavelet features derived radiomics nomogram for prediction of malignant and benign early stage lung nodules. *Sci. Rep.* **2021**, *11*, 22330. [[CrossRef](#)]
30. Rosas, I.O.; Yao, J.; Avila, N.A.; Chow, C.K.; Gahl, W.A.; Gochuico, B.R. Automated quantification of high-resolution CT scan findings in individuals at risk for pulmonary fibrosis. *Chest* **2011**, *140*, 1590–1597. [[CrossRef](#)]
31. Chang, Y.; Lim, J.; Kim, N.; Seo, J.B.; Lynch, D.A. A support vector machine classifier reduces interscanner variation in the HRCT classification of regional disease pattern in diffuse lung disease: Comparison to a Bayesian classifier. *Med. Phys.* **2013**, *40*, 051912. [[CrossRef](#)] [[PubMed](#)]
32. Depeursinge, A.; Chin, A.S.; Leung, A.N.; Terrone, D.; Bristow, M.; Rosen, G.; Rubin, D.L. Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution computed tomography. *Investig. Radiol.* **2015**, *50*, 261–267. [[CrossRef](#)] [[PubMed](#)]
33. Chong, D.Y.; Kim, H.J.; Lo, P.; Young, S.; McNitt-Gray, M.F.; Abtin, F.; Goldin, J.G.; Brown, M.S. Robustness-Driven Feature Selection in Classification of Fibrotic Interstitial Lung Disease Patterns in Computed Tomography Using 3D Texture Features. *IEEE Trans. Med. Imaging* **2016**, *35*, 144–157. [[CrossRef](#)] [[PubMed](#)]
34. Budzikowski, J.D.; Foy, J.J.; Rashid, A.A.; Chung, J.H.; Noth, I.; Armato, S.G. Radiomics-based assessment of idiopathic pulmonary fibrosis is associated with genetic mutations and patient survival. *J. Med. Imaging* **2021**, *8*, 031903. [[CrossRef](#)]
35. Kim, G.B.; Jung, K.H.; Lee, Y.; Kim, H.J.; Kim, N.; Jun, S.; Seo, J.B.; Lynch, D.A. Comparison of Shallow and Deep Learning Methods on Classifying the Regional Pattern of Diffuse Lung Disease. *J. Digit. Imaging* **2018**, *31*, 415–424. [[CrossRef](#)] [[PubMed](#)]
36. Walsh, S.L.F.; Calandriello, L.; Silva, M.; Sverzellati, N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: A case-cohort study. *Lancet Respir. Med.* **2018**, *6*, 837–845. [[CrossRef](#)]
37. Furukawa, T.; Oyama, S.; Yokota, H.; Kondoh, Y.; Kataoka, K.; Johkoh, T.; Fukuoka, J.; Hashimoto, N.; Sakamoto, K.; Shiratori, Y.; et al. A comprehensible machine learning tool to differentially diagnose idiopathic pulmonary fibrosis from other chronic interstitial lung diseases. *Respirology* **2022**, *27*, 73–74. [[CrossRef](#)] [[PubMed](#)]
38. Rawat, W.; Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.* **2017**, *29*, 2352–2449. [[CrossRef](#)] [[PubMed](#)]
39. Zhang, B.; Qi, S.; Pan, X.; Li, C.; Yao, Y.; Qian, W.; Guan, Y. Deep CNN Model Using CT Radiomics Feature Mapping Recognizes EGFR Gene Mutation Status of Lung Adenocarcinoma. *Front. Oncol.* **2021**, *10*, 598721. [[CrossRef](#)]
40. Yang, Y.; Zeng, N.; Chen, Z.; Li, W.; Guo, Y.; Wang, S.; Duan, W.; Liu, Y.; Chen, R.; Kang, Y. Multi-Layer Perceptron Classifier with the Proposed Combined Feature Vector of 3D CNN Features and Lung Radiomics Features for COPD Stage Classification. *J. Healthc. Eng.* **2023**, *2023*, 3715603. [[CrossRef](#)]
41. Lin, C.Y.; Guo, S.M.; Lien, J.J.; Lin, W.T.; Liu, Y.S.; Lai, C.H.; Hsu, I.L.; Chang, C.C.; Tseng, Y.L. Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT. *Radiol. Med.* **2024**, *129*, 56–69. [[CrossRef](#)]
42. Liang, C.H.; Liu, Y.C.; Wan, Y.L.; Yun, C.H.; Wu, W.J.; López-González, R.; Huang, W.M. Quantification of Cancer-Developing Idiopathic Pulmonary Fibrosis Using Whole-Lung Texture Analysis of HRCT Images. *Cancers* **2021**, *13*, 5600. [[CrossRef](#)] [[PubMed](#)]
43. Barnes, H.; Humphries, S.M.; George, P.M.; Assayag, D.; Glaspole, I.; Mackintosh, J.A.; Corte, T.J.; Glassberg, M.; Johannson, K.A.; Calandriello, L.; et al. Machine learning in radiology: The new frontier in interstitial lung diseases. *Lancet Digit. Health* **2023**, *5*, e41–e50. [[CrossRef](#)]
44. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gellyand, S.o. An image is worth 16×16 words: Transformers for image recognition at scale. In Proceedings of the 9th International Conference on Learning Representations, ICLR, Virtual, 3–7 May 2021.
45. Umejiaku, A.P.; Dhakal, P.; Sheng, V.S. Detecting COVID-19 Effectively with Transformers and CNN-Based Deep Learning Mechanisms. *Appl. Sci.* **2023**, *13*, 4050. [[CrossRef](#)]
46. Okolo, G.I.; Katsigiannis, S.; Ramzan, N. IEViT: An enhanced vision transformer architecture for chest X-ray image classification. *Comput. Methods Programs Biomed.* **2022**, *226*, 107141. [[CrossRef](#)] [[PubMed](#)]

47. Chen, J.; He, Y.; Frey, E.C.; Li, Y.; Du, Y. Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. *arXiv* **2021**, arXiv:2104.06468.
48. Wang, T.; Lan, J.; Han, Z.; Hu, Z.; Huang, Y.; Deng, Y.; Zhang, H.; Wang, J.; Chen, M.; Jiangand, H.o. O-Net: A novel framework with deep fusion of CNN and transformer for simultaneous segmentation and classification. *Front. Neurosci.* **2022**, *16*, 876065. [[CrossRef](#)] [[PubMed](#)]
49. Islam, M.D.; Rahman, M.M.; Ali, M.S.; Mahim, S.M.; Miah, M.S. Enhancing lung abnormalities diagnosis using hybrid DCNN-ViT-GRU model with explainable AI: A deep learning approach. *Image Vis. Comput.* **2024**, *142*, 104918. [[CrossRef](#)]
50. Cao, K.; Deng, T.; Zhang, C.; Lu, L.; Li, L. A CNN-transformer fusion network for COVID-19 CXR image classification. *PLoS ONE* **2022**, *17*, e0276758. [[CrossRef](#)]
51. Mabrouk, A.; Díaz Redondo, R.P.; Dahou, A.; Abd Elaziz, M.; Kayed, M. Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks. *Appl. Sci.* **2022**, *12*, 6448. [[CrossRef](#)]
52. Ukwuoma, C.C.; Qin, Z.; Belal Bin Heyat, M.; Akhtar, F.; Bamisile, O.; Muaad, A.Y.; Addo, D.; Al-antari, M.A. A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images. *J. Adv. Res.* **2023**, *48*, 191–211. [[CrossRef](#)] [[PubMed](#)]
53. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [[CrossRef](#)] [[PubMed](#)]
54. Armato III, S.G.; 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.; et al. 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](#)]
55. Qiu, J.; Mitra, J.; Dumas, C.; Sarachan, B.; Ghose, S.; Judson, M. Radiomics-guided 3D CNN-Vision Transformer (Rad-CNNViT) ensemble to diagnose pulmonary sarcoidosis from CT. In Proceedings of the SPIE Medical Imaging: Image Processing, San Diego, CA, USA, 18–22 February 2024; 12926; pp. 144–149.
56. Morozov, S.; Chernina, V.; Blokhin, A.I.; Gombolevskiy, V.A. Chest computed tomography for outcome prediction in laboratory-confirmed COVID-19: A retrospective analysis of 38,051 cases. *Digit. Diagn.* **2020**, *1*, 27–36. [[CrossRef](#)]
57. Conners, R.W.; Trivedi, M.M.; Harlow, C.A. Segmentation of a High-Resolution Urban Scene using Texture Operators. *Comput. Vision Graph. Image Process.* **1984**, *25*, 273–310. [[CrossRef](#)]
58. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
59. Pérez-García, F.; Sparks, R.; Ourselin, S. TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput. Methods Programs Biomed.* **2021**, *208*, 106236. [[CrossRef](#)] [[PubMed](#)]
60. Draelos, R.L.; Carin, L. Hirescam: Faithful location representation in visual attention for explainable 3d medical image classification. *arXiv* **2020**, arXiv:2011.08891.
61. Abnar, S.; Zuidema, W. Quantifying attention flow in transformers. *arXiv* **2020**, arXiv:2005.00928.
62. Cho, H.; Lee, H.Y.; Kim, E.; Lee, G.; Kim, J.; Kwon, J.; Park, H. Radiomics-guided deep neural networks stratify lung adenocarcinoma prognosis from CT scans. *Commun. Biol.* **2021**, *4*, 1286. [[CrossRef](#)]
63. Haralick, R.M. Statistical and Structural Approaches to Texture. *Proc. IEEE* **1979**, *67*, 786–804. [[CrossRef](#)]
64. Haralick, R.M.; Shanmuga, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621.
65. Kuhlman, J.E.; Fishman, E.K.; Hamper, U.M.; Knowles, M.; Siegelman, S.S. The computed tomographic spectrum of thoracic sarcoidosis. *Radio Graph.* **1989**, *9*, 449–466. [[CrossRef](#)] [[PubMed](#)]
66. Kumazoe, H.; Matsunaga, K.; Nagata, N.; Komori, M.; Wakamatsu, K.; Kajiki, A.; Nakazono, T.; Kudo, S. “Reversed halo sign” of high-resolution computed tomography in pulmonary sarcoidosis. *J. Thorac. Imaging* **2009**, *24*, 66–68. [[CrossRef](#)]
67. Salehi, A.W.; Khan, S.; Gupta, G.; Alabdullah, B.I.; Almjally, A.; Alsolai, H.; Siddiqui, T.; Mellit, A. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. *Sustainability* **2023**, *15*, 5930. [[CrossRef](#)]

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.