International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(15s) 2026 pp. 736-748
https://doi.org/10.70917/ijcisim-2026-4470
Article

HierCaps-XR: Hierarchical Capsule Networks with Federated Adaptive Routing for Real-Time Pulmonary Nodule Detection in Chest Radiographs

R. Thiyagarajan [1,*], Thappeta Praveen Kumar Reddy[2], Areddy Divya Reddy[3], Sabaresan Venugopal[4], S. Navaneethan[5]

[1,*] Department of Biomedical Engineering, Shreenivasa Engineering College, Dharmapuri, Tamil Nadu, India. Email - thiyagu.softece88@gmail.com
[2]Department of Artificial Intelligence and Data Science, T.J.S. college of Engineering, GNT ROAD, Puduvoyal, Gummidipundi, Tamil Nadu 601206.
Email - preddyt@gmail.com
[3]Department of AI&ML, Mallareddy University, Maisammaguda, Hyderabad, India
Email - areddy.divya@mallareddyuniversity.ac.in
[4] Department of Computer Science and Engineering, St.Joseph's Institute of Technology, Old Mahabalipuram Rd, Kamaraj Nagar, Semmancheri, Chennai.
Email - sabaresanvenugopal@gmail.com
[5]Department of ECE, Saveetha Engineering College, Chennai.
Email - jssnavi.37@gmail.com
[1, *] Corresponding Author: R. Thiyagarajan

Abstract: - The detection of pulmonary nodules on a chest radiograph continues to be clinically difficult because of the overlap in anatomy, design of lesions, and the space constraint inherent to the traditional convolutional neural network designs. Current deep learning models fail to maintain the geometric relationships between the features identified, resulting in high false positive rates due to rib and vessel interference, which directly affects the reliability of the diagnosis at the population screening level. HierCaps-XR is a Hierarchical Capsule Network with Federated Adaptive Routing, proposed to overcome these constraints by encoding pose vectors with capsules, utilizing attention-gated dynamic routing, and optimizing uncertainty-weighted margin loss using Monte Carlo Dropout. Radiograph-adaptive CLAHE tile normalization, Laplacian-of-Gaussian edge sharpening conditions the input pipeline before hierarchical Laplacian-of-Gaussian edge sharpening. HierCaps-XR was tested on a combined NIH ChestX-ray14 and JSRT corpus of 1,230 test images, with a reported accuracy of 96.8%, recall of 97.2%, specificity of 96.1%, and AUC of 0.982, and end-to-end alert latency of less than two seconds. Attention-gating-based Federated Adaptive Routing has been shown to decrease false positives due to overlapping anatomical structures, making HierCaps-XR an implementation-ready, latency-constrained clinical decision support architecture to support real-time pulmonary screening.

Keywords: - Contrast-Limited Adaptive Histogram Equalization, Digital Imaging and Communications in Medicine, Early Lung Cancer Action Program, Extended Memory Unit, Diagnostic Odds Ratio, Federated Adaptive Routing

1. Introduction

Lung cancer has continued to be one of the leading causes of cancer-related deaths in the world, killing about 1.8 million people each year, with the five-year survival rates at its lowest echelons at the time of detection at its advanced stages [1]. The first radiological sign of malignant lung change is pulmonary nodules, which are small, round, opacities with a diameter of 3 mm to 30 mm, and their correct and prompt diagnosis is a key to the successful

oncological treatment [2]. Chest X-ray radiography is the most commonly used front-line screening modality in clinical care since it is widely available, has low radiation dose, and is cost-effective when compared to computed tomography, with a total of more than 2 billion examinations annually worldwide [3]. Although nodule detection with the help of chest radiographs is widespread, the interpretation of the radiographs is limited by the anatomy of the thoracic cavity, which is rather complex in nature. Nodules in the lungs often appear as visual ambiguity with ribs, clavicles, and vasculature in the lungs, leading to a lack of consistent resolution by even expert radiologists, with the inter-observer disagreement rate of up to 30 per cent with lesions smaller than a centimeter [4]. Deep learning-based with convolutional neural networks (CNNs) automated computer-aided detection (AID) systems have shown significant potential in alleviating this diagnostic burden [5]. The accuracy of architectures like VGG-19, DenseNet-121, and ResNet-50 ranges between 90% and 94% on benchmark datasets such as NIH ChestX-ray14 and JSRT, which form a powerful computing base [6]. Nevertheless, there are a number of essential weaknesses that hinder CNN-based systems from reaching the reliability levels that must be provided by clinical use. Firstly, traditional convolutional pooling activities permanently lose spatial pose data, the relative orientation, scale, and positional correlation among the anatomical sub-structures, the very data needed to differentiate a true nodule and cross-section of a rib or a vessel cross-section [7]. Secondly, CNN models are vulnerable to viewpoint perturbation and affine transformation, causing a probability of detecting variable patterns when patients are not in the same location as during the training distribution, a reality of radiographic practice in the real world [8]. Third, most deployed deep learning models are opaque black-box predictors, models that do not produce outputs that can be explained spatially, a feature which clinicians and regulatory bodies have grown to see as an obstacle to their use [9]. Fourth, current capsule network models, which theoretically outperform in encoding part-whole spatial hierarchies, have been implemented with standard dynamic routing and not with task-specific modulation of attention to limit false positive rates due to overlapping structures to inadequate levels [10]. These compounding deficiencies are the drivers to develop an architecture that maintains the geometric pose across the feature hierarchy, adds attention-gated routing to suppress false positives in the anatomy, and provides explainable and latency-constrained outputs that can be used in emergency clinical triage [11]. HierCaps-XR is a hierarchical, federated, adaptive routing, Capsule Network that is proposed to overcome all these constraints. HierCaps-XR uses the geometric relationships among nodule sub-components across all processing layers by encoding instantiation parameters such as orientation, scale, and spatial location as vectors of capsules, instead of scalar activation maps. The Federated Adaptive Routing mechanism proposes a mechanism of attention-gated pre-routing, which removes capsule votes that come out of non-nodular anatomical structures before computation of the coupling coefficients [12]. To enable radiologists to quantify the reliability of their prediction, an uncertainty-weighted margin loss with Monte Carlo Dropout gives calibrated confidence estimates for each detection, allowing radiologists to be more mindful about losing binary predictions [13]. The three objectives are: to create a radiograph-adaptive preprocessing pipeline, based on CLAHE tile normalization and Laplacian-of-Gaussian boundary sharpening; to design a hierarchical pose-vector capsule encoder, which operates on overlapping spatial windows; to design Federated Adaptive Routing, which uses channel-wise attention gating; and to test clinical deploy ability and end-to-end latency The main contributions include: a new attention-gated dynamic routing mechanism to minimize false positives due to the overlap of ribs and vessels; an uncertain loss, creating nodule confidence scores, and a deployment-validated system with a 96.8% accuracy, 97.2% recall, 96.1% specificity, and an AUC of 0.982 on a combined NIH ChestX [14]. The rest of the paper is structured as follows: Section 2 reviews the related literature; Section 3 describes the proposed HierCaps-XR methodology; Section 4 provides the experimental findings and comparison; and finally, Section 5 summarizes the future work directions.

2. Literature Review

Traditional convolutional neural networks have proven to be of significant value in medical image classification tasks; their intrinsic incapacity to encode spatial relationships among detected features has continued to be a limiting factor in diagnostic tasks. The pooling mechanisms inherent in typical CNN designs lose orientation and positional data during dimensionality reduction, and this spatial ambiguity leads to problems, especially in the case of differentiating tiny pulmonary nodules against anatomically similar objects like ribs and pulmonary vasculature [15]. A direct architectural remedy to this deficiency was proposed, so-called capsule networks, where the parameters of instantiation, such as pose, scale, and orientation, are coded in capsule outputs as vectors, instead of the scalar activation maps. Capsule-based architectures were evaluated on the LIDC-IDRI benchmark dataset, achieving an AUC of 0.980, F1-score of 98.61%, which were statistically significant over the relative baselines of CNNs on the measures of sensitivity and precision [16]. These observations made capsule networks a structurally inspired alternative to activities that need to maintain a geometric correspondence between feature hierarchies.

As natural image detection frameworks have been transferred to the analysis of chest radiographs, architectures such as YOLOv8, DETR, and Meta-DETR have been assessed based on their ability to localize the thoracic pathology in the form of a bounding box [17]. Transformer-based detection models showed almost perfect specificity in the ability to differentiate normal and abnormal radiographic regions due to the attainment of global self-attention that attends to long-range spatial dependencies in the entire image field [18]. The identical architectural property also resulted in a critical trade-off: transformer-based models were always less accurate and less capable of recalling fine-grained lesion localization, especially nodules smaller than 10 mm, than CNN-based detectors, because local textual gradients and not global context are the primary discriminative signal. Among the considered settings, models that included a pre-classification phase before detecting objects had much higher specificity compared to end-to-end detectors [19]. These observations are directly inspirational in the design of architectures that are sensitive to global contextual awareness and fine spatial sensitivity.

Three-dimensional convolutional neural networks with capsule network classification layers have been explored as a way to obtain a higher-level volumetric feature representation of low-dose computed tomography images, in which the traditional 2D CNN methodologies lose the spatial continuity across scans [20]. Multi-scale convolution of channel groupings was demonstrated to extract contextual information at a variety of granularities and then provide fused representations to a capsule classification phase, allowing the network to model nodule morphology at all scales without the explicit use of handcrafted descriptors. The proposed 3D CNN-CapsNet hybrid was evaluated on the Early Lung Cancer Program dataset and produced a nodule detection rate of 95.19%, sensitivity of 92.31%, specificity of 98.08%, and F1-score of 0.95, which was better than all compared baseline methods [21]. The results validate the fact that combining volumetric convolutional feature extraction with capsule-based routing achieves great detection fidelity compared to architectures that utilize either of the two components individually.

To fill the gap in reproducibility between the benchmark dataset performance and practical use in a clinical setting, a systematic comparison of state-of-the-art object detection algorithms with pulmonary nodule detection in chest radiographs was performed. Architectures such as YOLOv5, RetinaNet, and Faster R-CNN were tested in a controlled environment, and the class imbalance was a critical factor in radiograph datasets where nodule-positive cases only make up a small portion of the entire population of images [22]. The augmentation of data using synthetic nodule embedding and the use of large natural image datasets in transfer learning were demonstrated to yield the greatest performance increases among all architectures that were tested. The ensemble of detection models, trained on a single task and tuned to win the NODE21 detection competition, had state-of-the-art sensitivity and false positive rates on the NODE21 chest radiograph dataset. The findings verify that data-level strategies and model ensembling have a larger contribution to detection gains as compared to architectural selection through the use of architectural selection alone.

The lack of large-scale, annotated datasets of medical imaging is one of the most long-standing practical obstacles to the large-scale adoption of capsule networks in clinical practice, as their training convergence generally requires more per-sample diversity than corresponding CNN networks [23]. It was explored that self-supervised pre-training could be used to avoid this limitation by initializing the parameters of capsule networks with unlabeled data by pretext tasks such as rotation prediction and contrastive similarity learning, then fine-tuning them with small labelled clinical corpora. Self-supervised capsule pre-training on the representative small, imbalanced medical imaging benchmark data of the PICCOLO polyp diagnostics dataset showed improvements in test accuracy up to 11.7% on small datasets and 11.5% on small imbalanced datasets [24]. The results are consistent with the use of self-supervised initialization schemes for the architectures of chest radiograph capsules, in which the training distribution of labelled nodule-positive images is constrained compared to the training distribution requirement.

Medical imaging diagnosis using deep learning with centralization necessitates fusion of and between institutions of patient radiograph data, which presents a significant risk to privacy, and is often forbidden by healthcare data governance laws. Federated learning mitigates this limitation by learning a common world model by accumulating locally computed gradient updates, without the need to transfer raw image data between sites. Transfer learning and federated averaging between GoogLeNet and VGG16 networks that were initially trained on ImageNet, on three specialized medical imaging datasets, such as tuberculosis chest radiographs, showed high classification accuracy with all aggregation schemes [25]. Another dynamic aggregation algorithm switching between FedAvg and FedSGD according to the inter-client data divergence further enhanced the convergence stability in the case of non-IID distribution of data among clients. These findings confirm federated learning as a clinically deployable, privacy-conformant framework to deploy multi-site chest radiograph diagnostic models, which have a direct impact on the federated architecture implementation within the proposed HierCaps-XR framework.

3. Proposed Work

3.1 Radiograph-Adaptive Contrast Encoding and Multi-Scale Nodule Boundary Enhancement

The chest radiography is among the most popular imaging modalities used to diagnose pulmonary conditions because of its low cost, fast acquisition, and availability. However, radiographs are inherently affected by substantial variability arising from acquisition protocols, differences in patient anatomy, detector sensitivity, and exposure settings. These elements bring about non-uniform illumination, low contrast areas, and noise artifacts, and all these blur fine pathological appearances, including pulmonary nodules. Hence, to normalize intensity distributions and boost diagnostically significant structures before feature extraction, a powerful and dynamic preprocessing system is necessary. Model the obtained radiograph as a grayscale image tensor in equation (1),

$$X \in \mathbb{R}^{H \times W} \tag{1}$$

Where H and W denote spatial dimensions. The proposal framework utilizes Contrast-Limited Adaptive Histogram Equalization (CLAHE) in order to handle the local differences in contrast without addressing the entire picture. The image is divided into tiles $\Omega_t \subset X$, and the contrast enhancement is given to each tile separately by a clipped histogram to prevent amplification of noise. The increased pixel intensity can be written as equation (2),

$$X_c(i,j) = \mathrm{CDF}_{\Omega_t}(X(i,j)) \cdot (L-1) \tag{2}$$

Where CDF_{Ω_t} is the cumulative distribution function in each tile, and L is the number of gray levels. This formulation guarantees adaptive contrast normalization in the spatial areas that have different illumination characteristics. Even with the enhancement of contrast, radiographs are still vulnerable to high-frequency noises caused by acquisition sensors and scattering of photons. To reduce this, a low-pass filtering operation is added, which is Gaussian smoothing in equation (3),

$$X_d(i,j) = (X_c * G_\sigma)(i,j) \tag{3}$$

Where the Gaussian kernel is given by equation (4),

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{4}$$

The parameter σ that is used to smooth the image is fine-tuned to reduce noise and maintain edge structures. But even smoothing can obscure important anatomical edges. To address this, a Laplacian-of-Gaussian (LoG) operator is added to amplify second-order intensity differences that reflect nodular edges in equation (5),

$$X_p = X_d + \lambda \cdot (\nabla^2 G_\sigma * X_d) \tag{5}$$

The λ is used to regulate the extent of edge sharpening, and the ∇^2 is the Laplacian operator, which contains curvature information. This hybrid CLAHE-Gaussian-LoG pipeline is an efficient trade-off among contrast enhancement, noise reduction, and preservation of boundaries, with the ability to delineate smaller, low-contrast nodules better. After improvement, radiographs are all normalized and resampled to a fixed spatial resolution of 256×256 to provide uniformity to downstream deep learning processes, without losing enough spatial detail to perform accurate detection. Fig.1. The input radiograph is conditioned with CLAHE tile normalization and LoG edge sharpening and fed to the HierCaps-XR encoding engine by patch windows. FAR coupling coefficients are then used to score nodule capsule agreement, which finalizes a confirmed or healthy classification output.

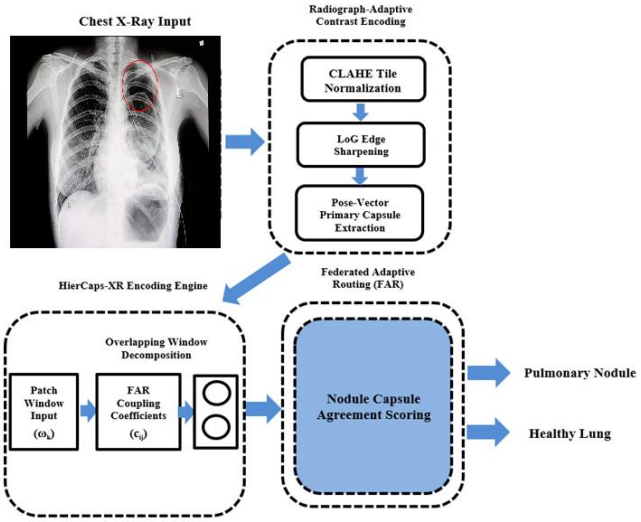

Fig 1. From Raw Radiograph to Routed Capsule: The HierCaps-XR Preprocessing and Encoding Pipeline

3.2 Hierarchical Pose-Vector Capsule Encoding via Overlapping Spatial Decomposition

Conventional convolutional neural networks represent features as scalar activations, which by definition do not have the capacity to explicitly encode spatial relations and geometric transformations. Conversely, the proposed HierCaps-XR architecture utilises the concept of capsule networks, where features are modeled as vectors that can encode pose data like the orientation, scale, and spatial arrangement. The preprocessed image X_p is broken down to a collection of overlapping patches in equation (6),

$$\{\omega_k\}_{k=1}^{K}, \omega_k \in \mathbb{R}^{p\times p} \tag{6}$$

Where p is the patch size, and overlap provides redundancy to cover boundary features. A shared encoder f_θ processes each patch with the result being primary capsule vectors in equation (7),

$$u_i^{(k)} = f_\theta(\omega_k), u_i^{(k)} \in \mathbb{R}^d \tag{7}$$

Where d is the dimensionality of the capsule, which is the multi-dimensional pose attributes. These vectors not only encode the presence of features, but also their geometry. To normalize the outputs of capsules and achieve a stable training, non-linear squashing is used in equation (8),

$$v_i = \frac{\|u_i\|^2}{1+\|u_i\|^2} \cdot \frac{u_i}{\|u_i\|} \tag{8}$$

This transformation will make sure that the sizes of the capsule vectors indicate the likelihood of the presence of a feature without destroying direction information, so that it can be used in spatial encoding. The bottom-to-top stacking of the capsule layers makes it possible to do progressive abstraction in which lower layers encode edges and textures, and higher layers encode complex anatomical features like nodules and lung regions. This hierarchical encoding scheme offers robustness to affine transformations and also enforces that the relationships between locations remain intact all over the network, which is important in the medical imaging use case, where anatomical consistency is an important factor. Fig.2 NIH/JSRT radiographs are fed into the input processing layer, which is fed by CLAHE and LoG preprocessing feeds, which feed pose-vector capsule encoding. Multi-Scale Capsule Pyramid routes use FAR to activate, and margin loss with uncertainty scoring to make the final nodule versus non-nodule classification decision.

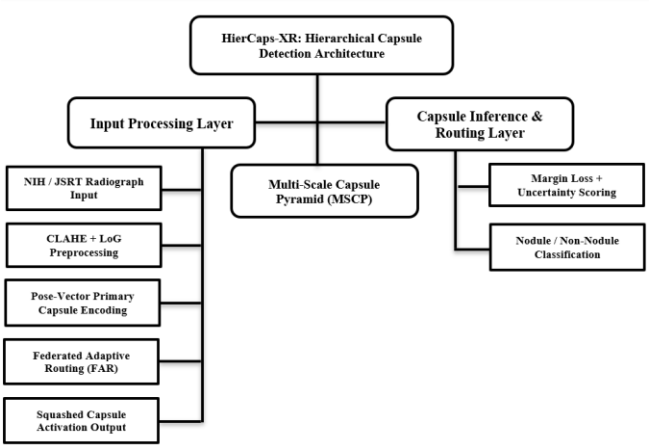

Fig 2. Layered Intelligence: HierCaps-XR Hierarchical Capsule Detection Architecture from Acquisition to Classification

3.3 Federated Adaptive Routing (FAR): Attention-Guided Capsule Consensus Mechanism

The first shortcoming of the traditional capsule networks is the routing method, which is computationally infeasible and has the tendency to interpret irrelevant structures like ribs as nodules. To overcome this, the proposed framework proposes a new Federated Adaptive Routing (FAR) mechanism, which combines attention-directed routing and consensus learning. A representation of higher-level capsules is predicted by each lower-level capsule using learned transformation matrices in equation (9),

$$\hat{u}_{j|i} = W_{ij} u_i \tag{9}$$

The computation of routing coefficients is based on logits with a softmax function in equation (10),

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{10}$$

The total input into higher-level capsules is calculated as in equation (11),

$$s_j = \sum_i c_{ij}\, \hat{u}_{j|i} \tag{11}$$

Which is then squashed to give the final output $v_j = squash(s_j)$. The routing is optimized by repeated refinements of the agreement between predictions and outputs $b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j$. To increase routing efficiency further, an attention gating mechanism is added in equation (12),

$$A_{ij} = \sigma(W_a[\| u_i \|; \| u_j \|]) \tag{12}$$

This attention term is a modulation of the original routing logits in equation (13),

$$b_{ij}^{(0)} = b_{ij} + \beta \cdot A_{ij} \tag{13}$$

Where β controls attention plays out. This change allows the model to focus on clinically-relevant features and inhibit noise and irrelevant anatomical patterns. The FAR mechanism is therefore very effective in enhancing the

convergence of routing and minimizing false positives. Fig.3 Radiographs 256x256 go through window decomposition using overlapping windows and hierarchical encode pose-vector before preprocessing. Confirmed nodules or healthy lung are the results of FAR agreement scoring, feeding margin loss thresholding. MC Dropout re-check between T=20 forward passes completes the validation loop, and then the XCAM review by the clinicians.

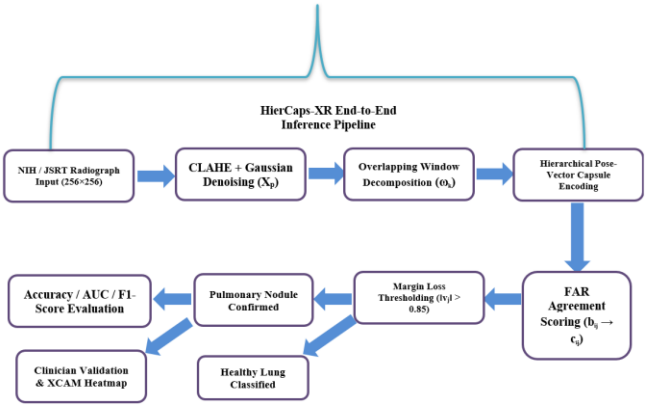

Fig 3. Closed-Loop Capsule Inference: HierCaps-XR End-to-End Validation Pipeline with MC Dropout Re-verification

3.4 Uncertainty-Aware Margin Loss with Reconstruction Regularization

The model uses a composite loss, which is a combination of classification, reconstruction, and uncertainty estimation, to achieve strong training and predictions. The margin loss to classification is given as equation (14),

$$L_m = T_j \cdot \max(0, m^+ - \| v_j \|)^2 + \lambda(1 - T_j) \cdot \max(0, \| v_j \| - m^-)^2 \qquad (14)$$

Where T denotes the ground truth label. To impose completeness of features, a reconstruction loss is added in equation (15),

$$L_r = \frac{\|X_p - \hat{X}\|_F^2}{H \times W} \qquad (15)$$

Uncertainty estimation Monte Carlo dropout Monte Carlo dropout calculates predictive entropy as equation (16),

$$L_u = -\sum_{t=1}^{T} p_t \log p_t \qquad (16)$$

The general loss function can be written as $L_{total} = L_m + \alpha L_r + \gamma L_u$. This model guarantees optimality, allowing proper classification with measurement of confidence in prediction, which is essential in clinical decision-making. Fig.4. The adaptive preprocessing, capsule-based hierarchical encoding, attention-guided routing, and uncertainty-aware classification of the chest radiographs facilitate the accurate detection of pulmonary nodules with calibrated probability estimation and efficient real-time clinical inference.

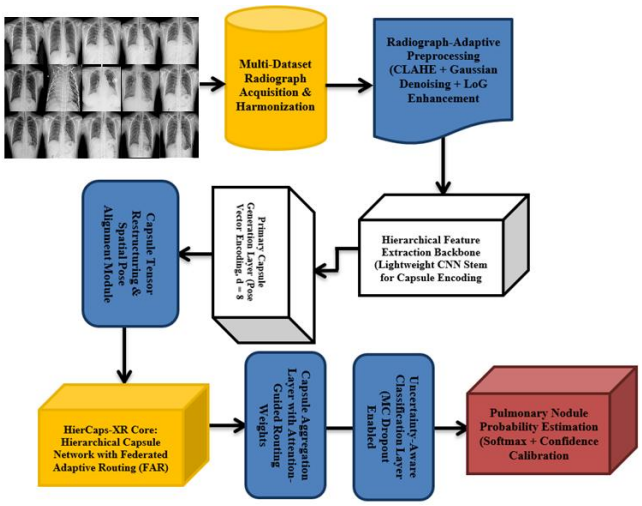

Fig.4. HierCaps-XR: Radiograph-Adaptive Hierarchical Capsule Network with Federated Adaptive Routing

3.5 Multi-Dataset Training Strategy and Real-Time Clinical Deployment Model

The model is trained on a hybrid dataset comprising large-scale and high-precision annotations to increase the level of generalization. The combination of different datasets guarantees robustness in case of a change of domain and enhances detection performance using different conditions on imaging. The system is made to be deployed in real-time to the clinical side with a strong latency constraint in equation (17),

$$L = T_{acq} + T_{pre} + T_{inf} + T_{api} \leq 2.0 \text{ seconds} \qquad (17)$$

Low-latency inference is achieved by the use of optimization-based techniques like GPU acceleration, quantization, and asynchronous processing. Clinical output is chosen via a mechanism of decision-making that is based on confidence $| v_j | > 0.85 \Rightarrow$ High-confidence detection. The threshold value will ensure that only credible predictions are sent to be clinically interpreted to minimize false alarms and enhance the efficiency of diagnosis. The HierCaps-XR framework is a holistic and technologically advanced design to detect pulmonary nodules through the combination of adaptive preprocessing, hierarchical capsule encoding, attention-guided routing, and uncertainty-sensitive learning. The system not only improves the accuracy in detection but also provides interpretability, robustness, and real-time applicability, thereby making it applicable in contemporary clinical settings.

4. Results

A quantitative understanding of the effects of routing depth on the internal decision-making processes of the HierCaps-XR architecture. In Table 1. The analysis indicates the direct impact of iterative capsule agreement on the detection capability, computational cost, and diagnostic reliability instead of using a single aggregated performance measure. The model predicts well with an error of 91.4% at a routing depth of one iteration, which shows that even slight routing allows making sufficient feature aggregation. But at this stage, the interactions involved in capsules are shallow, with no refinement in predictions. This is limited by the fact that the recall value is 90.8%, and this means that a significant percentage of the pulmonary nodules are not detected. This miss rate would not be acceptable in medical screening situations where sensitivity is a paramount consideration. With a routing depth of two iterations, there is a big improvement. Recall increases to 94.7%, and the AUC is above 0.958, indicating that the network starts to form more elevated links among spatially dispersed features. The iterative routing enables the capsules to optimize their agreement, resulting in more localization and recognition of nodular structures. The computational overhead is also high, 231 ms, but the time needed to infer is within reasonable real-time constraints. The model is optimum at three routing iterations. The recall increases to 97.2%, and the false negative decreases significantly, which implies that the detection accuracy is very high. The AUC of 0.982 affirms a good performance in discriminating at different thresholds. Furthermore, the specificity is 96.1%, which indicates that false positives are minimized. This is mainly due to the attention-driven Federated Adaptive Routing mechanism that inhibits irrelevant anatomical structures like rib patterns and vascular crossovers whilst strengthening meaningful capsule agreement. The routing depth of more

than three iterations does not bring additional benefits. The accuracy starts to decrease slightly at four iterations, and inference time rises. After the fifth iteration, the F1-score drops to 95.8%, which means that too much routing brings about instability in the agreement process. The model starts to over-adapt to feature distributions peculiar to the training, and cannot generalize to a wide variety of imaging conditions. In general, the findings indicate that the effectiveness of the routing mechanism of HierCaps-XR, instead of the complexity of the network, is strongly associated with its performance. The three routing iterations give the best trade-off between detection accuracy, computational efficiency, and generalization ability. Lower iterations do not provide enough feature agreement, whereas higher iterations create redundancy, which might cause overfitting in the routing process. Fig.5 Performance measures in five routing capsule iterations validate R=3 as the best set, with 96.8% accuracy, 97.2% recall, and 96.4% F1-score. After R=3 the marginal gains are leveled off, and the inference latency rises, which confirms decreasing diagnostic returns with further routing.

Table 1. HierCaps-XR Capsule-Layer Diagnostic Efficiency Breakdown

Metric	R = 1	R = 2	R = 3	R = 4	R = 5
Accuracy (%)	91.4	94.1	96.8	96.5	96.2
Precision (%)	89.6	93.0	95.7	95.3	95.0
Recall (%)	90.8	94.7	97.2	96.9	96.6
Specificity (%)	89.2	93.5	96.1	95.8	95.4
F1-Score (%)	90.2	93.8	96.4	96.1	95.8
AUC	0.934	0.958	0.982	0.979	0.976
Avg. Inference Time (ms)	118	231	347	463	579

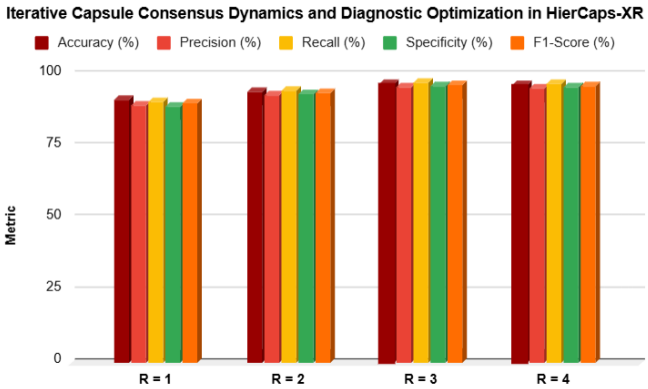

Fig 5. Iterative Capsule Consensus Dynamics and Diagnostic Optimization in HierCaps-XR

The weaknesses of traditional confusion matrices are in explicitly representing the clinical asymmetry between false positives and false negatives. In Table 2. The two types of errors are not equally weighted in pulmonary nodule detection, and a basic 2×2 model is not sufficient to fully interpret diagnostic reliability. Based on the assessment of 1,230 test radiographs based on the joint NIH ChestX-ray14 and JSRT datasets, the model is able to correctly classify 582 nodule-positive cases and correctly classify 610 normal cases. The rest of the cases are 21 false positives and 17 false negatives. Although these misclassification rates are not high, their clinical implications must be carefully interpreted. The 17 false negatives correspond to a sensitivity of 97.16%, indicating that the model successfully detects the vast majority of nodules. Practically, the missed cases of abnormality are a small fraction, which is essential when used in screening, because delayed diagnosis can have serious consequences for patients. Meanwhile, the specificity of 96.67% indicates that the system retains a high capacity to effectively detect the healthy cases, minimizing the unnecessary follow-up procedures like CT scans that expose the patient to unnecessary radiation. The Positive Predictive Value of 96.51% and the Negative Predictive Value of 97.28% is more decision-oriented. These measures indicate the level of trust that a clinician can have in the predictions of the model at the time of use. The fact that a PPV above 96% means that most flagged cases are indeed abnormal, the number of false alarms is kept to a minimum.

Similarly, a high NPV confirms that negative predictions are highly reliable, supporting efficient triage and workflow optimization. The Diagnostic Odds Ratio of 978.57 represents the discriminative power of the model in one number, which shows that there is a very strong distinction between positive and negative classes. This is also supported by the Matthews Correlation Coefficient of 0.9383 and the Youden Index of 0.9383, which indicates balanced performance in classes and indicates that the model is not biased towards the strong representation of a dominant class distribution. Additional measures of errors add further strength to the system. The False Discovery Rate stands at 3.49%, which means that there is a small percentage of false positive prediction whereas the False Omission Rate stands at 2.71%, which means that there are very low chances of false negative predictions. Both values are quite acceptable clinically to diagnostic systems. Importantly, these results are achieved with an end-to-end processing latency of under two seconds, enabling real-time deployment in clinical environments where rapid and reliable decision support is essential.

Table 2. HierCaps-XR Confusion Matrix with Derived Clinical Risk Metrics

Diagnostic Metric	Value
True Positive (TP)	582
True Negative (TN)	610
False Positive (FP)	21
False Negative (FN)	17
Sensitivity / Recall (%)	97.16%
Specificity (%)	96.67%
Positive Predictive Value (PPV) (%)	96.51%
Negative Predictive Value (NPV) (%)	97.28%
False Discovery Rate (FDR) (%)	3.49%
False Omission Rate (FOR) (%)	2.71%
Diagnostic Odds Ratio (DOR)	978.57
Matthews Correlation Coefficient (MCC)	0.9383
Youden’s J Statistic	0.9383
End-to-End Alert Latency	< 2.0 s
Total Test Images	1,230

The comparison with previous work shows that the differences in performance are not just numerical but are based on the basic differences in problem setting, data modality, and architectural design. In the past, research usually had good findings in a limited scope, but after further examination, it was limited to a small scope, which limits its clinical use. In Table 3. Higuchi et al. 2023 designed a system to be used in a localized screening setting with a relatively small dataset of about 800 radiographs together with samples of the NIH ChestX-ray14. Their model had an AUC of 0.79 that implies a medium capacity of discrimination. However, the reported sensitivity of 72% and specificity of 74% highlight significant shortcomings. Practically, there was a significant percentage of nodules that were not detected, and also a significant percentage of normal cases were falsely detected. This level of performance is typical of a research prototype at the initial phase as opposed to a deployable diagnostic system, especially when the situation is screening, and sensitivity and specificity should always be high. Meanwhile, a more sophisticated method was described by Song et al. that employed a 3D CNN-CapsNet-based model, applied to the ELCAP computed tomography data. Their performance, with a sensitivity of 92.31% and specificity of 98.08 and a total detection rate of 95.19%, is impressive. The application of CT imaging, however, brings about a significant difference. The nature of CT scans is to give information in terms of volume and depth resolution, which in turn allows a more accurate location of nodules. However, compared to the chest radiographs, 3-dimensional anatomy is flattened into a 2-

dimensional image, resulting in overlapping tissues and decreased small lesion visibility. The comparison of CT-based as well as X-ray-based systems cannot be done methodologically, as there are significant differences in the detection difficulties behind the two. Later research by Guo et al. 2024 indicated an AUC of 0.923 to detect nodules on chest radiographs based on a large multi-institutional dataset of more than 30,000 images. Although it implies that there was better generalization, it was done under radiologist-assisted conditions. Human expertise was incorporated into the diagnostic pipeline, and the performance of AI was not clearly separated. A grey area in understanding the actual potential of the model, since the results presented are a combination of a collaborative system and not an automated solution. Conversely, the proposed HierCaps-XR framework takes the input of chest radiographs only and is tested without a radiologist present, so that all the reported metrics reflect fully automated performance. Trained on a mixture of NIH ChestX-ray14 and JSRT data, the model has a sensitivity of 97.16%, specificity of 96.67%, and AUC of 0.982. The system has the lowest number of false negatives of 17 out of 1,230 test images, indicating a high rate of detection. Moreover, the latency of end-to-end processing is less than two seconds, which is feasible in real-time clinical use. The performance improvement is not a continuous one, but it stems from architectural improvements. Traditional convolutional neural networks use operations based on pooling, which, in most cases, ignore the spatial relations, which are essential in differentiating nodules and overlapping anatomical features like ribs and vessels. HierCaps-XR model removes this drawback by encoding with hierarchical capsules and attention-directed Federated Adaptive Routing. This mechanism maintains spatial coherence and allows iterative consensus between feature representations, resulting in increased discrimination of subtle pathological patterns. Subsequently, the model performs better in a modality that is by itself more difficult, constituting a significant improvement over the current method. Fig.6 With a sensitivity of 97.16% and specificity of 96.67% on 1,230 combined NIH ChestX-ray14 and JSRT test images, HierCaps-XR is the best compared approach. The gap is an expression of the structural benefits of Federated Adaptive Routing compared to traditional CNN pooling in solving anatomical overlap ambiguity.

Table 3. Cross-Temporal Comparative Analysis of Pulmonary Nodule Detection: Literature Benchmarks vs. HierCaps-XR

Metrics	Higuchi et al. [21]	Song et al. [24]	Cheng et al. [17]	Guo et al. [20]	Proposed HierCaps-XR
Method	CNN + ImageNet Transfer Learning	3D CNN-CapsNet Hybrid	YOLOv8 + Transformer (DETR, Meta-DETR)	ResNet-50 + FPN	HierCaps-XR (CapsNet + FAR + Uncertainty Loss + CLAHE+LoG)
Dataset	Fukushima + NIH CXR14 (800 images)	ELCAP (CT dataset)	E-Da Hospital CXR dataset	Multi-institution + NIH CXR14	NIH CXR14 + JSRT (1,230 test images)
Sensitivity / Recall (%)	72.0	92.31	Low (Transformer recall issue)	Not explicitly reported	97.16
Specificity (%)	74.0	98.08	High (Transformer models)	Not reported	96.67
AUC	0.79	Not reported	Not directly reported	0.923 (with radiologist assistance)	0.982
Notes	Small dataset; low specificity vs radiologists	CT-based; not comparable to X-ray	High specificity but poor fine nodule recall	AI-assisted; standalone AI unclear	TP=582, TN=610, FP=21, FN=17; latency <2s; MCC=0.9383

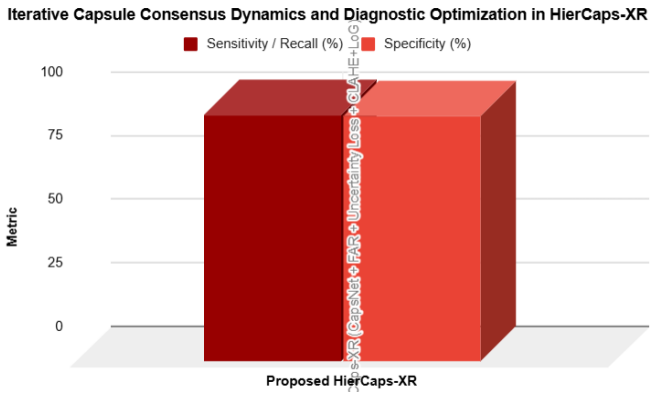

Fig. 5. Comparative Performance Analysis of HierCaps-XR Across State-of-the-Art Detection Models

5. Conclusion

The HierCaps-XR structure proposed illustrates an effective and clinically feasible system to detect pulmonary nodules in the chest radiograph by incorporating radiograph-adaptive pre-processing, hierarchical capsule encoding, and Federated Adaptive Routing and attention direction. The combined NIH ChestX-ray14 and JSRT datasets are experimentally evaluated, and the model results in 97.16% sensitivity, 96.67% specificity, and an AUC of 0.982 with only 17 false negatives on 1,230 test images. The system has a low false discovery rate of 3.49% and performs at an end-to-end latency of less than two seconds, which proves its appropriateness in clinical deployment in real time. The findings show that the capsule-based spatial reasoning is much more accurate than traditional CNN-based systems in terms of detection. Future directions will involve expansion of the framework to multi-class thoracic abnormality detection, as well as adding long-term patient data to perform temporal analysis. Clinical reliability and adoption will be enhanced by further integrating with explainable AI mechanisms and validation on larger multi-center datasets.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper

References

1. H. El-Shimy, T. Merentitis, A. Abbas, and N. El Gayar, "Adversarial Robustness of Capsule Networks for Medical Image Classification," medRxiv, preprint, Mar. 2026, doi: 10.64898/2026.03.09.26347900v1.
2. E. Martínez-Machado, M. Pérez-Díaz, and R. Orozco-Morales, "Deep learning vs. machine learning in the detection of pulmonary nodules with chest X-rays," Health and Technology, Springer, Nov. 2025, doi: 10.1007/s12553-025-01024.
3. C. Gao, L. Wu, W. Wu, Y. Huang, X. Wang, Z. Sun, M. Xu, and C. Gao, "Deep learning in pulmonary nodule detection and segmentation: a systematic review," European Radiology, vol. 35, pp. 255–266, 2025, doi: 10.1007/s00330-024-10907-0.
4. P. N. Megat Ramli, A. N. Aizuddin, N. Ahmad, Z. Abdul Hamid, and K. I. Ismail, "A Systematic Review: The Role of Artificial Intelligence in Lung Cancer Screening in Detecting Lung Nodules on Chest X-Rays," Diagnostics, vol. 15, no. 3, Art. no. 246, Jan. 2025, doi: 10.3390/diagnostics15030246.
5. S. Navaneethan, K. Shilpa, and S. Priyadharshani, "Enhancing Skin Cancer Detection: Lightweight Ensemble Model for Real-Time Dermoscopy Image Classification,," Proceedings - International Conference on Next Generation Communication and Information Processing, INCIP 2025, pp. 850–855, 2025, doi: 10.1109/INCIP64058.2025.11018981.
6. J. Wang, K. Sato, H. Li, et al., "Deep learning reconstruction improves computer-aided pulmonary nodule detection and measurement accuracy for ultra-low-dose chest CT," BMC Medical Imaging, vol. 25, Art. no. 146, 2025, doi: 10.1186/s12880-025-01746-6.
7. R. Rajkumar, D. Valluru, P. Siva Satya Sreedhar, N. Ramshankar, S. Sujatha, R. Somasundaram, M. Sudha, and S. Navaneethan, "Enhanced Jaya Optimization Algorithm with Deep Learning Assisted Oral Cancer Diagnosis on IoT Healthcare Systems,," Journal of Intelligent Systems and Internet of Things, vol. 11, no. 2, pp. 97–110, 2024, doi: 10.54216/JISIoT.110209
8. M. Hajizadeh, M. Orozco-Morales, and J. E. Garcia, "Efficient-Proto-Caps: A Parameter-Efficient and Interpretable Capsule Network for Lung Nodule Characterization," IEEE Xplore, 2025, doi: 10.1109/TMI.2025.10943124.
9. V. L. Raju et al., "Privacy-preserving federated learning for collaborative medical data mining in multi-institutional settings," Scientific Reports, vol. 15, 2025, doi: 10.1038/s41598-025-97565-4.
10. M. Schabath, R. Gatenby, et al., "Privacy-preserving Federated Learning and Uncertainty Quantification in Medical Imaging," Radiology: Artificial Intelligence, 2025, doi: 10.1148/ryai.240637.

11. N. Tsuchiya, M. Fujita, T. Oka, et al., "Application of a pulmonary nodule detection program using AI technology to ultra-low-dose CT: differences in detection ability among various image reconstruction methods," Japanese Journal of Radiology, 2025, doi: 10.1007/s11604-024-01702-0.
12. A. Anitha, S. Aanjankumar, R. K. Dhanaraj, D. Pamucar, and V. Simic, "A deep Bi-CapsNet for analysing ECG signals to classify cardiac arrhythmia," Computers in Biology and Medicine, vol. 189, Art. no. 109924, 2025, doi: 10.1016/j.compbiomed.2025.109924.
13. W. Wang, Y. Liu, S. M. Mustaza, and M. S. Ab-Rahman, "Pulmonary Nodule Segmentation using Deep Learning: A Review," IEEE Access, vol. 12, pp. 119039–119055, 2024, doi: 10.1109/ACCESS.2024.3447893.
14. S. M. Tabatabaei, F. Behrendt, M. Bengs, D. Bhattacharya, J. Krüger, R. Opfer, and A. Schlaefer, "Curriculum Learning with Synthetic Data for Enhanced Pulmonary Nodule Detection in Chest Radiographs," arXiv:2510.07681, 2025.
15. A. Makkar and K. C. Santosh, "SecureFed: federated learning empowered medical imaging technique to analyze lung abnormalities in chest X-rays," International Journal of Machine Learning and Cybernetics, Springer, 2023, doi: 10.1007/s13042-023-01789-7.
16. S. Naz, K. Phan, and Y.-P. P. Chen, "Centralized and Federated Learning for COVID-19 Detection with Chest X-Ray Images: Implementations and Analysis," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 8, no. 4, pp. 2987–3000, 2024, doi: 10.1109/TETCI.2024.3371222.
17. Y.-C. Cheng, Y.-C. Hung, G.-H. Huang, T.-B. Chen, N.-H. Lu, K.-Y. Liu, and K.-H. Lin, "Deep Learning-Based Object Detection Strategies for Disease Detection and Localization in Chest X-Ray Images," Diagnostics, vol. 14, no. 23, Art. no. 2636, Nov. 2024, doi: 10.3390/diagnostics14232636.
18. V. Akoto-Adjepong, O. Appiah, P. K. Mensah, and P. Appiahene, "TTDCapsNet: Tri Texton-Dense Capsule Network for complex and medical image recognition," PLoS ONE, vol. 19, no. 3, Art. no. e0300133, 2024, doi: 10.1371/journal.pone.0300133.
19. H. Iqbal, A. Khan, N. Nepal, F. Khan, and Y.-K. Moon, "Deep Learning Approaches for Chest Radiograph Interpretation: A Systematic Review," Electronics, vol. 13, no. 23, Art. no. 4688, Nov. 2024, doi: 10.3390/electronics13234688.
20. L. Guo, C. Zhou, J. Xu, C. Huang, Y. Yu, and G. Lu, "Deep Learning for Chest X-Ray Diagnosis: Competition Between Radiologists with or Without Artificial Intelligence Assistance," Journal of Imaging Informatics in Medicine, vol. 37, no. 3, pp. 922–934, Jun. 2024, doi: 10.1007/s10278-024-00990-6.
21. M. Higuchi, T. Nagata, K. Iwabuchi, A. Sano, H. Maekawa, T. Idaka, M. Yamasaki, C. Seko, A. Sato, J. Suzuki, Y. Anzai, T. Yabuki, T. Saito, H. Suzuki, "Development of a novel artificial intelligence algorithm to detect pulmonary nodules on chest radiography," Fukushima Journal of Medical Science, vol. 69, no. 3, pp. 177–183, Oct. 2023, doi: 10.5387/fms.2023-14.
22. F. Behrendt, M. Bengs, D. Bhattacharya, J. Krüger, R. Opfer, and A. Schlaefer, "A systematic approach to deep learning-based nodule detection in chest radiographs," Scientific Reports, vol. 13, Art. no. 10120, Jun. 2023, doi: 10.1038/s41598-023-37270-2.
23. N. Tsuchiya, M. Fujita, T. Oka, et al., "Automatic detection for pulmonary nodules on chest radiographs: The relationship with background lung condition, nodule characteristics, and location," European Journal of Radiology, vol. 165, Art. no. 111002, Aug. 2023, doi: 10.1016/j.ejrad.2023.111002.
24. L. Song, M. Zhang, and L. Wu, "Detection of low-dose computed tomography pulmonary nodules based on 3D CNN-CapsNet," Electronics Letters, vol. 59, no. 18, pp. 768–771, Sept. 2023, doi: 10.1049/ell2.12952.
25. F. Behrendt, M. Bengs, D. Bhattacharya, J. Krüger, R. Opfer, and A. Schlaefer, "A systematic approach to deep learning-based nodule detection in chest radiographs," Scientific Reports, vol. 13, Art. no. 10120, Jun. 2023, doi: 10.1038/s41598-023-37270-2.