scientific reports

OPEN

Federated lung nodule segmentation using a hybrid transformer–U-Net architecture

Sapthak Mohajon Turjya[1]✉ & Mulham Fawakherji[2]✉

Early and accurate identification of malignant pulmonary nodules is essential to reduce lung cancer mortality, yet segmentation of these nodules in CT scans remains challenging due to their small size, variable morphology, low contrast, and extreme class imbalance. To address these challenges, we propose a privacy-preserving federated learning framework that integrates a hybrid U-Net–Transformer architecture with a customized Dice-Focal loss function. The hybrid model combines residual convolutional blocks for local texture feature extraction with Transformer-based self-attention for capturing global contextual dependencies, while skip connections, Layer Normalization, and Dropout enhance gradient flow and generalization across heterogeneous client datasets. The Dice-Focal loss mitigates class imbalance and improves boundary sensitivity, yielding stable convergence and precise segmentation of solid nodules. Using the LUNA16 dataset, our federated setup demonstrated high segmentation fidelity for nodules between 15*mm* and 25*mm*, achieving Dice coefficients up to 0.93 and robust precision-recall trade-offs across clients. These results highlight the framework's potential for scalable, privacy-preserving lung cancer screening, while also identifying challenges in segmenting part-solid or low-contrast nodules for future methodological improvements.

Keywords Lung nodule segmentation, Federated learning, Transformer-U-Net, Customized loss function, Privacy-preserving AI

Cancer remains one of the leading causes of death worldwide, affecting individuals across all socioeconomic backgrounds[1]. Among the various types of cancer, lung cancer stands out as the most commonly diagnosed and the deadliest form[2,3]. In the United States, lung cancer is the second most commonly diagnosed cancer and remains the leading cause of cancer-related deaths[4]. Globally, it also exhibits the highest mortality rates[5]. A major clinical concern in the early detection of lung cancer is the identification of pulmonary nodules, which are often discovered incidentally during routine chest imaging. While many of these nodules are benign, a subset may represent early-stage malignancies. Timely and accurate identification of malignant nodules is essential for effective intervention and improving patient outcomes. However, the high prevalence of benign nodules presents a significant diagnostic challenge. To improve early detection and reduce lung cancer mortality, large-scale screening programs have been explored. One of the most impactful efforts in this direction is the National Lung Screening Trial (NLST), which demonstrated that low-dose computed tomography (LDCT) screening can reduce lung cancer mortality by 20% over seven years in high-risk individuals compared to traditional chest radiography[6]. Following these findings, LDCT-based screening programs have been implemented in the United States and are under consideration in other countries. Nevertheless, the success of such screening initiatives has introduced new challenges–most notably, the enormous volume of CT scans that radiologists must interpret. Despite the high resolution and lower radiation of modern imaging technologies such as CT and MRI, the interpretation process remains time-consuming and labor-intensive for healthcare providers. To alleviate this burden, researchers have developed Computer-Aided Detection (CAD) systems over the past two decades to assist radiologists in identifying pulmonary nodules. These systems aim to enhance the speed and accuracy of CT image analysis, thus supporting early detection and improving the cost-effectiveness of lung cancer screening programs. In the field of medical image segmentation, early methods relied on traditional machine learning and image processing techniques such as thresholding, region growing, edge detection, iterative local thresholding, active contour models, and clustering algorithms[7–9]. These methods typically involve preprocessing steps like grayscale conversion, filtering, and binarization, and often require significant manual intervention for both segmentation and feature engineering. As a result, their generalizability and scalability to complex clinical settings are limited. In recent years, deep learning, particularly Convolutional Neural Networks (CNNs), has

[1]Kalinga Institute of Industrial Technology, Bhubaneswar, India. [2]North Carolina Agricultural and Technical State University, Greensboro 27411, USA. ✉email: sapthakmohajon6@gmail.com; mfawakherji@ncat.edu

achieved state-of-the-art performance in medical image segmentation, enabling pixel-level segmentation by learning hierarchical feature representations. These models offer high accuracy, automatic feature extraction, robustness, and adaptability across different medical imaging tasks like prostate and liver tumor segmentation. Moreover, two major limitations continue to restrict the development of robust deep learning models for lung nodule segmentation. First, clinical data is inherently distributed across hospitals and medical institutions, and privacy regulations such as HIPAA and GDPR prevent the centralization of sensitive patient information[10,11]. This dispersion leads to sparse, heterogeneous, and often inconsistently annotated datasets, hindering the training of high-performing models. Second, lung nodules typically occupy a small portion of the image volume and exhibit highly variable visual characteristics, with class imbalance and indistinct boundaries further complicating accurate segmentation[12,13]. These limitations demand a paradigm that can enable collaborative training while preserving patient confidentiality and supporting clinical heterogeneity and the subtle features of solid nodules. Federated learning (FL) offers a compelling solution by enabling decentralized model training across institutions without the need for data sharing. At the same time, the integration of state-of-the-art architectural innovations, such as attention mechanisms and hybrid learning models, can improve the extraction of discriminative features tailored to the demands of medical imaging tasks[14,15]. This work addresses these critical challenges by proposing a federated segmentation framework that combines a Transformer-based U-Net architecture with a specialized segmentation loss function designed to handle class imbalance and capture fine-grained patterns. The proposed model aims to achieve robust, privacy-preserving (as no raw patient data leaves the institutions), and clinically relevant non-compromisable performance in solid lung nodule segmentation of varying size $15\,\text{mm} - 25\,\text{mm}$ in diverse and distributed datasets.

Literature review

Convolutional Neural Networks (CNNs) have become the foundation of most biomedical image analysis due to their ability to learn hierarchical spatial features. Recent developments in the analysis of pulmonary nodules have increasingly relied on deep learning architectures for both detection and segmentation. Beyond the basic 2D CNN- and U-Net-based architectures, several works have looked into volumetric and multi-scale feature extraction methodologies for increasing the sensitivity to small or irregular nodules. Zhang et al.[21] proposed a 3D convolutional detection framework that integrates a voxel-flow-based interpolation network, which corrects the inconsistent slice spacing in CT volumes, followed by a 3D U-Net-RPN hybrid for nodule detection. This work shows that volumetric contextual reasoning significantly raises the bar of detection accuracy, achieving the state-of-the-art despite increased computational cost. Complementing these detection-centric advances, Wang et al.[22] proposed CPLOYO, a multi-scale YOLO-based framework enhanced with RepViT-CAMF, MSCAF fusion, and nonlinear kernel activation networks KANs, which reached superior sensitivity performance on the LUNA16 dataset, especially for small nodules. These collectively underpin the increased focus on representation learning in 3D space, multi-scale fusion, and nonlinear feature extraction toward robust pulmonary nodule detection under heterogeneous conditions. Besides novelties specific to model designs, recent literature points out broader methodological challenges in the analysis of pulmonary nodules. Gao et al.[23] have conducted a systematic review of deep learning-based detection and segmentation methods, underlining standardized preprocessing pipelines, harmonized evaluation metrics, external test datasets, and improved reporting practices. CNN-based segmentation architectures, particularly U-Net and its variants, such as ResUNet, have been quite successful[16] in this purpose. ResUNet combines the encoder–decoder design of U-Net with residual connections, improving gradient flow and convergence while preserving high-resolution feature recovery through skip connections. Despite these strengths, ResUNet is limited in capturing global context because convolutional receptive fields are inherently local. This is especially challenging for lung nodule segmentation, where nodules are often small, irregular, and low in contrast relative to surrounding tissue. To overcome these limitations, researchers have explored integrating Transformer-based modules into U-Net-like architectures. Models such as Swin-Unet[17] and TransUNet[18] employ self-attention to capture long-range dependencies, improving segmentation performance in complex medical images. Further developments include U-Netmer[19], which incorporates global Transformer blocks into patch-level U-shaped learning, and iU-Net[20], which fuses CNN and Swin Transformer pathways using wavelet-based mechanisms. These studies demonstrate that combining convolutional processing with attention mechanisms enhances performance for complex organ segmentation. However, most approaches have been tested only in centralized settings, without consideration for privacy or cross-institutional data heterogeneity, limiting their real-world clinical applicability. Federated learning (FL) has emerged as a key technique for privacy-preserving collaborative training. Early studies by Sheller et al.[14] and Li et al.[15] showed that FL can support tumor segmentation across institutions without sharing sensitive patient data. Since then, FL has been applied to various medical imaging modalities, including CT, MRI, and ultrasound. For example, Babar et al.[24] proposed adaptive federated optimization for heterogeneous clients, and Ma et al.[25] introduced personalized aggregation strategies to handle non-identically distributed (non-IID) data. Beyond the existing studies of federated learning for medical image segmentation, a few recent works further highlight the growing importance of hybrid architectures and privacy-preserving collaborative learning in healthcare AI. Perhaps most notably, Alphonse et al.[26] introduced a Federated Double Attention Multiscale Dense-U-Net combined with RL-based FedAvg strategy, achieving high BraTS 2020 accuracy by demonstrating how to enhance adaptive client optimization for improving the robustness of segmentation under heterogeneous data conditions. FedSegNet[27], on the other hand, is a federated framework for 3D medical image segmentation that presents a U-Net enhanced with Transformers and an Adaptive Aggregation Mechanism (AAM) capable of dynamic weighting among client updates based on data quality and divergence. This reduces communication costs and improves performance on BraTS, LiTS, and ACDC datasets. Another line of work is Chowdhury et al.'s[28] Hybrid Federated Learning-Based Ensemble Approach for Lung Disease Diagnosis. Using a fusion of CNN models-DenseNet201, InceptionV3, and VGG19-with the SWIN Transformer, a privacy-preserving diagnostic model for COVID-19 and pneumonia using X-ray data was developed, illustrating how effectively integrating transformer-based global feature modeling into CNN-based local representation learning proceeds within a federated environment. Taken together, all these studies underscore a strong trend of incorporating multiscale attention, Transformers, and ensemble learning within a federated setup for addressing challenges associated with data heterogeneity, privacy constraints, and limited local annotations. The presence of such hybrid attention-driven models in federated medical imaging stands to further strengthen the relevance and novelty of our current research, illustrating the need for advanced architectures capable of generalizing across a wide range of clinical institutions with integrity of data confidentiality. Frameworks such as Flower[29] have further simplified deployment by supporting asynchronous client updates and customizable aggregation schemes. Nonetheless, most FL implementations still rely on conventional CNN or ResUNet backbones and do not explicitly tackle highly imbalanced structures such as pulmonary nodules. Class imbalance is another major challenge in lung nodule segmentation because nodules occupy only a small fraction of CT scans, while the background dominates. Traditional loss functions, such as cross-entropy, perform poorly under such imbalance, resulting in high false-negative rates. Alternative loss functions like Dice Loss, Focal Loss[13], and Tversky Loss[30] are better suited for capturing minority-class structures. For example, Wavelet U-Net++[31] uses a hybrid Tversky–cross-entropy loss to improve performance under severe imbalance, and Usman et al.[32] introduced a boundary-sensitive focal Tversky loss for better delineation of rare structures. TPFR-Net[33] further demonstrates how carefully designed losses can enhance boundary localization. However, these approaches are mostly applied in centralized settings and rarely integrated into FL for lung nodule segmentation. Although architectural innovations, federated learning, and loss function optimization have advanced independently, they are rarely combined. Transformer-enhanced ResUNet models achieve strong segmentation and boundary awareness but are limited to centralized training. FL methods like FedAvg, FedProx++, and FedMA address decentralized learning but often use standard loss functions that are not optimized for nodules. Advanced losses such as focal–Tversky are seldom incorporated into FL due to communication and convergence challenges. Recent work like FedTransformer-Seg[34] combines attention with FL but still lacks imbalance-sensitive optimization and lightweight designs suitable for real-world deployment. Despite substantial progress in architectural design, federated learning, and loss function optimization for lung nodule segmentation, current approaches largely remain isolated: transformer-based models excel in capturing global context but are evaluated primarily in centralized settings, federated learning methods preserve privacy but often rely on standard CNN architectures with simple loss functions, and advanced loss designs are seldom integrated into federated frameworks. To address these limitations, this work proposes a unified framework that simultaneously: (i) leverages a hybrid Transformer-ResUNet architecture for effective local and global feature extraction, (ii) incorporates a customized Dice–Focal loss to handle class imbalance and enhance boundary-aware segmentation, and (iii) implements a privacy-preserving federated learning protocol suitable for real-world deployment. These contributions collectively enable accurate, efficient, and privacy-conscious lung nodule segmentation, bridging gaps identified in prior research.

The key contributions of this work are summarized as follows:

- A federated learning environment is implemented using the Flower framework, enabling multiple clinical clients to collaboratively train a segmentation model without sharing raw CT data, thereby addressing privacy and inter-institutional heterogeneity constraints.
- We propose a novel encoder–decoder design that combines convolutional residual blocks with transformer-based attention modules to capture both local texture and long-range global context for robust lung nodule segmentation.
- An improved preprocessing workflow incorporating CLAHE-based contrast enhancement and min–max normalization is used to boost image quality and support consistent learning across clients.
- This study integrates transformer-based global attention within a federated segmentation framework, addresses reproducibility and dataset variability challenges, and offers a practical pathway for deploying privacy-preserving cross-institutional medical AI systems.

Research methodology

This section outlines the methodology employed for precise segmentation of solid lung nodules from CT images. The proposed approach integrates advanced image preprocessing techniques with a hybrid deep learning architecture that combines the strengths of U-Net and Transformer models. The methodology is designed to address challenges such as low-contrast nodules, subtle anatomical variations, and limited receptive fields of conventional convolutional networks.

First, a systematic preprocessing pipeline is applied to enhance image quality and standardize intensity distributions, improving the visibility of fine nodular structures. Following preprocessing, a hybrid U-Net-Transformer architecture is used to extract both local texture and global contextual features. The encoder captures hierarchical local features, while the Transformer-based bottleneck models long-range spatial dependencies. The decoder reconstructs high-resolution segmentation masks, preserving spatial details and contextual information.

Data preprocessing

To improve the quality and uniformity of the CT scan images utilized in our work, a two-stage preprocessing pipeline was applied before training the segmentation model. As a first step, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used on every 2D slice of the CT scans. This method enhances local contrast and facilitates the visibility of nodular structures by redistributing intensity values across small image areas (tiles) while preventing over-amplification of noise by a clipping procedure. In particular, a CLAHE operator with a clip limit of 2.0 and a tile grid size of 8×8 was employed to substantially enhance the delineation of small or low-contrast lung nodules. After contrast enhancement, the images were Min-Max normalized for scaling pixel intensities to [0, 1]. This normalization process ensures that data distribution across the dataset is even and aids in stable convergence when training the model. The output is a collection of contrast-enhanced and normalized CT images, which are used as input to our hybrid Transformer-U-Net segmentation model. Such a preprocessing pipeline is very important in enhancing the visibility and accuracy of segmentation of fine nodular patterns, particularly in low-contrast areas of lung parenchyma. Figure 1 represents the preprocessed image along with its mask.

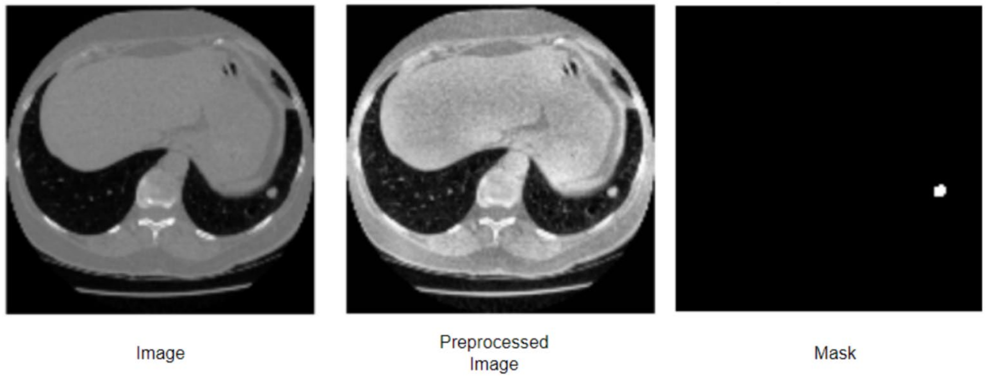

Fig. 1. Example of a preprocessed 2D lung CT slice (middle) after CLAHE enhancement and intensity normalization on the raw CT scan slice (left), alongside its corresponding ground-truth nodule mask (right).

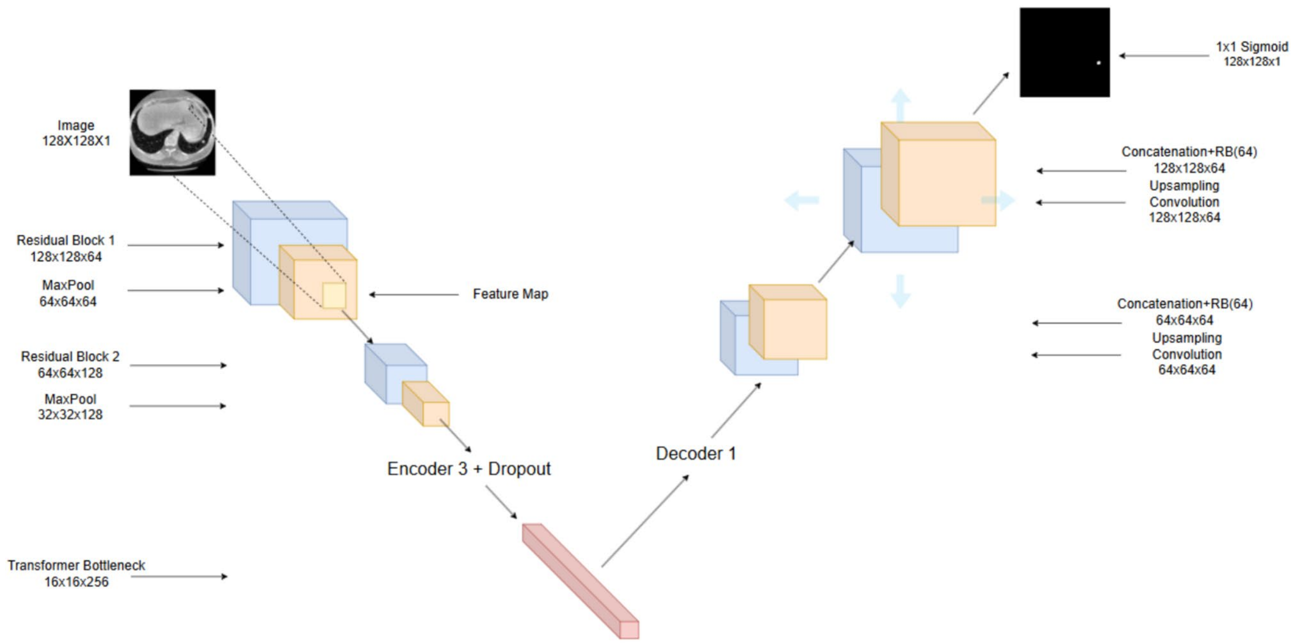

Fig. 2. Schematic diagram of the proposed Hybrid U-Net–Transformer architecture, which integrates convolutional encoding with Transformer-based global attention. This design enhances both local feature extraction and long-range spatial dependency modeling for improved nodule segmentation.

Model architecture

In this work, we develop a hybrid deep learning architecture that integrates U-Net and Transformer architectures to perform precise solid lung nodule segmentation on CT images (Algorithm 1 and Fig. 2). The model is based on a standard encoder-decoder methodology, with residual convolutional blocks added and a Transformer-based bottleneck for both local and global feature extraction. In the encoder stream, three residual convolutional blocks are stacked to hierarchically extract features progressively, each using skip connections and Layer Normalization to ensure information retention and stability in training. The deepest layer outputs are transformed into a sequence and fed into stacked Transformer blocks, with multi-head self-attention extracting long-range dependencies over spatial parts of the image. These Transformer layers enable the model to learn larger spatial relationships, which is especially useful for segmenting small or subtle lung nodules that might not be separable using solely local information.

After the transformer bottleneck, the decoder path uses transpose convolution layers to perform upsampling, in combination with residual blocks and skip connections from the encoder to rebuild the segmentation mask of high spatial resolution. The multi-stage decoding maintains both high-level contextual and low-level spatial details. The final output is a 1×1 convolution with sigmoid activation to produce a binary segmentation mask of the same resolution as input.

In comparison with traditional U-Net and CNN-based models, the new approach possesses numerous fundamental benefits. To begin with, the incorporation of self-attention mechanisms enables the model to improve its global context grasping, enhancing its performance in capturing small, scattered, or low-contrast nodules–situations in which classic convolutional networks tend to fail because of their poor receptive fields. Second, the residual learning blocks improve gradient flow in backpropagation, which allows for training of deeper networks without degradation and increases feature reuse. Third, employing Layer Normalization and Dropout enhances generalization and stability in training, especially when dealing with datasets that have variability in nodule appearance and CT scan quality. The hybrid design also weighs computational efficiency against segmentation precision so that it is suitable for both centralized and federated learning environments.

Essentially, the hybrid model of U-Net–Transformer bridges the limitations of conventional architectures by fusing local texture representation and global structural reasoning, resulting in better performance in intricate medical image segmentation tasks like lung nodule detection.

```
Require: Input image I ∈ ℝ^{128×128×1}
Ensure: Segmentation mask M ∈ ℝ^{128×128×1}
function ResidualBlock(x, f)
    if Channels(x) ≠ f then
        shortcut ← Conv2D(x, f, 1 × 1)
    else
        shortcut ← x
    end if
    x ← Conv2D(x, f, 3 × 3); x ← LayerNorm(x); x ← ReLU(x)
    x ← Conv2D(x, f, 3 × 3); x ← LayerNorm(x)
    return ReLU(x + shortcut)
end function
function TransformerBlock(x, d, h, r, p)
    x ← LayerNorm(x + Dropout(MultiHeadAttention(x, x, h, d)))
    mlp ← Dense(d · r, ReLU) → Dense(d)
    return LayerNorm(x + Dropout(mlp(x)))
end function
function HybridTransformerUnet(I)
    conv1 ← ResidualBlock(I, 64)
    pool1 ← MaxPool(conv1)
    conv2 ← ResidualBlock(pool1, 128)
    pool2 ← MaxPool(conv2)
    conv3 ← ResidualBlock(pool2, 256)
    drop3 ← Dropout(conv3, 0.6)
    pool3 ← MaxPool(drop3)
    trans ← Reshape(pool3, (256, 256))
    trans ← TransformerBlock(trans, 256, 8, 4, 0.1)
    trans ← TransformerBlock(trans, 256, 8, 4, 0.1)
    trans ← Reshape(trans, (16, 16, 256))
    up6 ← Conv2DTranspose(trans, 128, 2 × 2, stride = 2)
    up6 ← Concat(conv3, up6)
    up6 ← ResidualBlock(up6, 128)
    up7 ← Conv2DTranspose(up6, 64, 2 × 2, stride = 2)
    up7 ← Concat(conv2, up7)
    up7 ← ResidualBlock(up7, 64)
    up8 ← Conv2DTranspose(up7, 64, 2 × 2, stride = 2)
    up8 ← ResidualBlock(up8, 64)
    output ← Conv2D(up8, 1, 1 × 1, sigmoid)
    return output
end function
```

Algorithm 1. Unet-Transformer Model for Solid Lung Nodule Segmentation

Customized segmentation loss function

To counter the extreme class imbalance in lung nodule segmentation–where nodules take up a very small fraction of the CT slice versus the background–we use a hybrid loss function that combines **Dice Loss** and **Focal Loss** (Algorithm 2). The Dice loss calculates the overlap between the predicted and ground truth masks and is especially well-suited for tasks with small target regions since it maximizes the Dice Similarity Coefficient (DSC) directly.

Conversely, the Focal Loss targets learning on difficult-to-classify instances by reducing the weight of the easy negatives' contribution. It adds two parameters: α, a positive/negative contribution balance, and γ, an effect controlling parameter. We employ $\alpha = 0.25$ and $\gamma = 2.0$ in our configuration, parameters empirically demonstrated to cope with extreme imbalance conditions.

The final loss is a weighted sum:

$$\mathscr{L} - \text{total} = 0.9 \cdot \mathscr{L}_{\text{dice}} + 0.1 \cdot \mathscr{L}_{\text{focal}},$$

where $\mathscr{L}_{\text{dice}}$ and $\mathscr{L}_{\text{focal}}$ denote the Dice and Focal losses, respectively. This compound formulation uses the region-based sensitivity of Dice loss as well as the challenge-awareness of Focal loss. They collectively provide more stable training, better convergence, as well as much improved performance in nodule segmentation over using standard cross-entropy or Dice-only losses.

In summary, the enhanced performance of our model is due to its Hybrid Transformer-U-Net architecture, which blends residual convolutional blocks for local texture feature capture with Transformer-based self-attention modules for global contextual dependency modeling. This dual-capability architecture allows for precise segmentation of nodules ranging from size $15\,\text{mm} - 25\,\text{mm}$. The application of skip connections and residual learning promotes gradient flow and retains spatial information, while Layer Normalization and Dropout enhance generalization over heterogeneous client data. The use of the customized Dice-Focal loss function also mitigates extreme class imbalance and boundary sensitivity, leading to stable convergence and detailed segmentation. The lightweight architecture of the model (7–8M parameters) is conducive to computational efficiency, making it highly deployable in federated environments with privacy limitations.

The federated learning setup

To emulate a privacy-preserving decentralized training setup for lung nodule segmentation, we implement a FL approach based on the presented Hybrid Transformer-U-Net architecture and the designed segmentation loss function as a combination of Dice and Focal losses. The federated environment involves five clients that each maintain an equal and disjoint partition of the dataset. The training on the model happens in four communication rounds with each consisting of 500 local training epochs per client. The training, validation, and test split is done in the ratio 70:15:15 with shuffling is assigned to be true. This prevents bias (e.g., not all small nodules going into train and large ones in test), and it also ensures the train/test sets have a representative mix of nodule sizes, shapes, and textures. The federated learning process is carried out through the use of the Flower framework, an open-source FL platform that makes distributed machine learning experimentation easy[29]. The server randomly chooses one client at the start of the training process and initializes the global model weights with the model parameters of the chosen client. In every round, both clients update the model locally on their data with the optimized loss function, and once done, they send the local model weights to the central server. The server proceeds with model aggregation through the use of the Federated Averaging (FedAvg) algorithm, which calculates the weighted average of the client models to update the global model. This new global model is redistributed to the clients to start the subsequent round of local training. This procedure keeps sensitive medical information within the local institutions (clients), hence improving data privacy while enabling collective learning. Federated environment also brings in practical constraints like statistical heterogeneity and communication overhead, making the presented segmentation model more realistic and closer to real distributed healthcare settings. Figure 3 represents the Federated framework for the proposed research study.

Experimental results

Dataset description

For this study, we utilize the publicly available Lung Nodule Analysis 2016 (LUNA16) dataset, a benchmark specifically designed for the development and evaluation of computer-aided lung nodule detection and segmentation algorithms. LUNA16 is derived from the larger LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) database for thoracic CT scans. Selection criteria include slice thickness below 2.5 mm and uniform inter-slice spacing to ensure high-quality volumetric data. Each CT scan is accompanied by a corresponding binary mask annotation delineating the precise boundaries of lung nodules, as verified by multiple experienced radiologists. For the purposes of this project, the 3D CT volumes have been preprocessed and converted into 2D slices used for model training. This dataset is particularly suitable for evaluating the performance of our proposed Hybrid Transformer-ResUNet model within a federated learning setting, as it allows assessment of both nodule localization and segmentation accuracy. The inherent heterogeneity in nodule size, shape, and texture within LUNA16 further makes it an ideal benchmark for testing advanced segmentation architectures and customized loss functions in clinically realistic scenarios. The dataset is publicly accessible at: LUNA16.

Importantly, LUNA16 offers an ideal benchmark for evaluating federated learning frameworks due to its inherent properties, which closely mirror the practical constraints of decentralized clinical environments. First, LUNA16 is a subset of the multi-institutional LIDC-IDRI dataset, which comprises scans from various hospitals, scanners, reconstruction kernels, slice thicknesses, and radiologist annotation styles. This natural heterogeneity induces variations in intensity distribution, noise patterns, and nodule appearance, conditions arising in non-

IID client datasets seen in practical federated settings. Therefore, partitioning LUNA16 across multiple clients can easily simulate the process of distinct medical centers contributing locally unique data distributions to a shared global model. In addition, LUNA16 offers high-quality, radiologist-validated nodule masks, therefore allowing the proper evaluation of segmentation consistency among clients under domain shifts the most crucial generalization requirement in federated models. Since the dataset is large, publicly accessible, and clinically diverse, LUNA16 supports reproducible research in privacy-preserving medical AI by eliminating the legal and regulatory complexities involved in multi-center data sharing. Taken together, these properties make LUNA16 an ideal benchmark for testing federated lung nodule segmentation algorithms, especially those designed to operate reliably under heterogeneous data distributions and strict privacy constraints.

Evaluation metrics

Evaluating medical image segmentation models, particularly for highly imbalanced datasets such as lung nodule segmentation, requires more than a single metric. While the Dice Similarity Coefficient (DSC) is widely used, relying on it alone may not provide a complete picture of model performance. To achieve a more holistic evaluation, we also estimate the *False Positive Rate (FPR)* and *False Negative Rate (FNR)* in addition to Dice. These measures allow us to explicitly quantify over-segmentation (FPR) and under-segmentation (FNR), both of which are critical in clinical contexts.

Given the pixel-wise nature of semantic segmentation, the key evaluation measures are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{2}$$

$$\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \tag{3}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{4}$$

where *TP*, *FP*, and *FN* represent the counts of true positives, false positives, and false negatives, respectively.

From these, the **False Negative Rate (FNR)**, i.e., the fraction of true nodules that were not detected, is given by:

$$\text{FNR} = \frac{FN}{FN + TP} = 1 - \text{Recall}, \tag{5}$$

and the **False Positive Rate (FPR)**, i.e., the fraction of background pixels incorrectly classified as nodules, is expressed as:

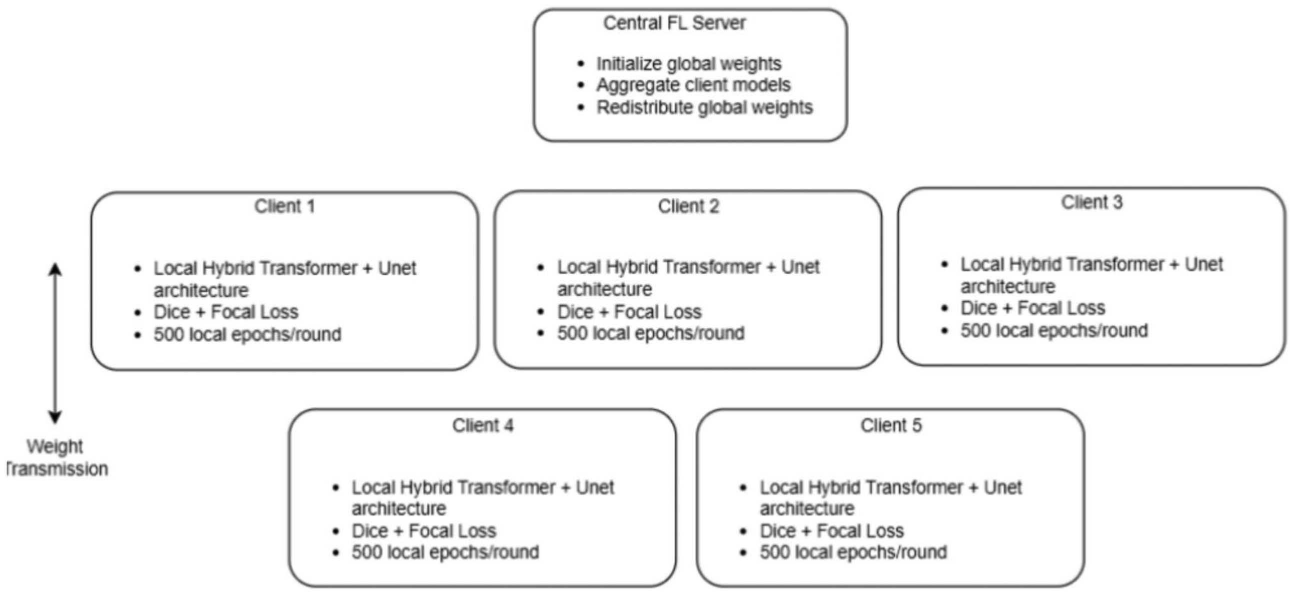

Fig. 3. Overview of the federated learning architecture, where multiple clients train locally on their own preprocessed CT slices without sharing raw data. The server aggregates client models using FedAvg to produce a privacy-preserving and globally optimized segmentation model.

$$\mathrm{FPR} = \frac{FP}{FP + TN}. \tag{6}$$

In medical image segmentation, where background pixels dominate and true negatives (*TN*) are rarely emphasized, the FPR is commonly approximated as:

$$\mathrm{FPR} \approx 1 - \mathrm{Precision}. \tag{7}$$

This approximation holds under the practical assumption that the number of background pixels greatly exceeds the number of foreground (nodule) pixels in CT slices.

Furthermore, in cases where only the Dice score is reported, the misclassification ratio can be estimated by rearranging the Dice formula:

$$\frac{FP + FN}{2 \cdot TP} = \left(\frac{1 - \mathrm{Dice}}{\mathrm{Dice}}\right). \tag{8}$$

Although this expression provides an estimate of the combined error relative to true positives, it cannot disentangle false positives from false negatives without access to either Precision or Recall.

Federated learning-based segmentation performance

The proposed federated learning framework demonstrated strong performance in detecting and segmenting solid pulmonary nodules using the LUNA16 dataset (Fig. 4). Although the dataset contains heterogeneous nodule types, including part-solid nodules (PSN) and pure ground-glass nodules (pGGN), the model was specifically trained and validated for solid nodule segmentation. The experimental setup emulated a distributed healthcare scenario, in which client-specific models were collaboratively updated into a global model without centralizing sensitive data. Table 1 summarizes the per-client segmentation performance across common evaluation metrics, including Dice coefficient, Precision, Recall, and Jaccard similarity (IoU). Class-wise segmentation results for the "Nodule" (class 0) and "No Nodule" (class 1) regions are presented in Table 2. Among all clients, Client 2 (Image 2) achieved the highest segmentation quality, with a Dice coefficient of 0.93, a precision of 0.90, Recall of 0.95, and a Jaccard similarity of 0.86. These results indicate robust detection and segmentation of solid nodules, with minimal false positives or missed detections. Clients 1, 4, and 5 also demonstrated strong segmentation fidelity, each achieving Dice scores greater than 0.85 and Jaccard similarities above 0.77. For these clients, class 1 ("No Nodule") performance was nearly perfect (Dice and IoU > 0.999), reflecting consistent handling of background regions where no nodules were present. In contrast, Client 3 (Image 3) exhibited a marked decline in performance, with a Dice coefficient of 0.74, a Recall of 0.59, and a Jaccard similarity of 0.59, despite achieving perfect Precision (1.00). This outcome suggests substantial under-segmentation: while all predicted nodule pixels were correct, a significant portion of the actual nodule region was omitted. Such behavior is likely attributable to nodule heterogeneity or low-contrast boundaries, which are common in part-solid or atypical nodules. Class-wise assessment further supports this interpretation: Client 3 reported the lowest scores for the "Nodule" class (Dice = 0.7447, IoU = 0.5932), highlighting the difficulty of achieving full nodule coverage. In contrast, the "No Nodule" class achieved near-perfect scores across all clients, consistent with class imbalance and the dominance of background pixels. It is worth noting that while both Dice and Jaccard similarity measure spatial overlap, Dice is more tolerant of discrepancies by balancing Precision and Recall, whereas Jaccard imposes a stricter penalty for mismatches. In this context, conventional accuracy was not emphasized, as the extreme class imbalance in lung CT images renders accuracy a misleading indicator of segmentation performance. Overall, the federated learning model exhibited high segmentation fidelity for well-defined solid nodules, particularly those ranging between $15\,\mathrm{mm}$ and $25\,\mathrm{mm}$ in size, supporting its potential for real-world deployment in distributed medical imaging environments. Nonetheless, its reduced performance on smaller, part-solid, or ground-glass nodules highlights the need for methodological enhancements. Future work may consider advanced preprocessing strategies such as intensity normalization, edge sharpening, multi-window CT channel fusion, or feature-aware filtering to enhance subtle nodule features. Additionally, integrating attention mechanisms or multi-scale context aggregation could further improve the detection of nodules with ambiguous borders or heterogeneous densities.

Comparison with centralized approach

To further evaluate the efficacy and practical utility of the proposed framework, we compared the federated training strategy against a conventional centralized learning paradigm. In centralized training, the entire dataset is used to train a single global model on a central server. While this approach benefits from full data availability and straightforward implementation, it raises several challenges, most notably concerns over data privacy, susceptibility to overfitting toward dominant data distributions, and limited adaptability in heterogeneous or institution-specific data settings. In contrast, the federated learning strategy distributes the dataset across five clients, each of which retains ownership of its local data and contributes to global model training without exposing raw patient information. Model updates are performed locally for multiple epochs on client-specific datasets and subsequently aggregated by a coordinating server to update the global model. This iterative, privacy-preserving process enables decentralized optimization while maintaining regulatory compliance. A notable finding from our experiments is the rapid convergence of the federated model: minimum training and validation losses were achieved within fewer than 100 epochs during the first communication round. Subsequent rounds further refined and stabilized the model, demonstrating the capacity of federated learning to efficiently leverage heterogeneous data sources. This behavior highlights two important advantages: accelerated convergence and improved stability across training iterations. Moreover, by exposing the global model to diverse and non-identically distributed data subsets, federated learning fosters stronger generalization compared to centralized training. Whereas centralized learning often prioritizes dominant patterns in aggregated datasets–potentially underrepresenting rare or subtle features–federated learning allows client-specific nuances to influence the global representation, thereby improving robustness on unseen data. Beyond performance considerations, federated learning provides intrinsic benefits in scalability and privacy preservation. Sensitive medical imaging data, such as lung CT scans, remain securely stored within institutional boundaries, in line with regulatory frameworks such as HIPAA and GDPR. This design is particularly valuable in real-world healthcare applications, where direct data sharing across institutions is often restricted or infeasible. In summary, federated learning offers a compelling alternative to centralized training by combining privacy preservation, rapid convergence, and enhanced generalization. By enabling collaborative learning across distributed datasets, it not only mitigates the risks of centralized data pooling but also supports the discovery of subtle and heterogeneous imaging patterns–an essential capability for high-stakes domains such as medical image analysis.

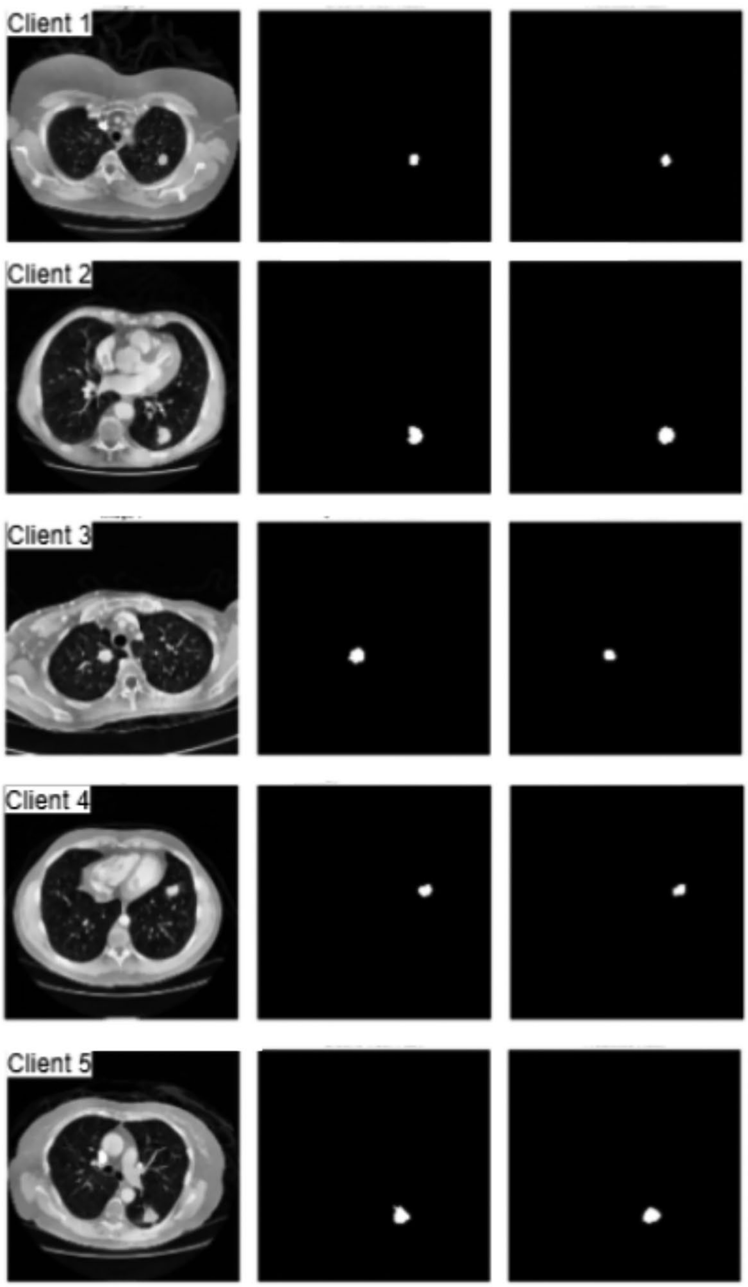

Fig. 4. Segmented lung nodules produced by the participating federated clients, illustrating variations in nodule appearance, boundary clarity, and anatomical context. These results highlight the model's ability to generalize across heterogeneous client datasets while maintaining consistent segmentation quality.

Image (Client)	Dice coefficient	Precision	Recall	mIoU
Image 1 (Client 1)	0.85	0.87	0.84	0.72
Image 2 (Client 2)	0.93	0.90	0.95	0.86
Image 3 (Client 3)	0.74	1.00	0.59	0.59
Image 4 (Client 4)	0.87	0.92	0.83	0.77
Image 5 (Client 5)	0.89	0.88	0.89	0.79

Table 1. Performance metrics (Precision, Recall, Dice, and IoU) for lung nodule segmentation across all federated clients. The results demonstrate variations in client performance due to non-IID data distributions while highlighting the overall robustness of the proposed model.

Image (Client)	Dice score		mIoU	
	Nodule	No nodule	Nodule	No nodule
Image 1 (Client 1)	0.8525	0.9997	0.7429	0.9994
Image 2 (Client 2)	0.9265	0.9997	0.8630	0.9994
Image 3 (Client 3)	0.7447	0.9993	0.5932	0.9985
Image 4 (Client 4)	0.8718	0.9997	0.7727	0.9994
Image 5 (Client 5)	0.8855	0.9995	0.7945	0.9991

Table 2. Class-wise segmentation performance metrics for background and nodule regions across all federated clients. These results highlight the greater variability in nodule-class IoU due to client-specific data heterogeneity and class imbalance.

Comparison with baseline federated architectures

For a more comprehensive validation of the proposed federated segmentation framework, its performance is compared to that of several baseline lung nodule segmentation architectures. These models are trained under federated learning conditions–characterized by decentralized optimization, non-IID data distribution, and communication constraints. All baselines were therefore re-trained under the same federated simulation setup to ensure a fair and consistent comparison. The results presented in Fig. 5 clearly illustrate the significant performance differences between the proposed method and the baselines, providing important insights into how different architectural choices respond to the challenges imposed by federated learning.

The Dice score for the earliest model, FCN-8s, was 0.310, representing an inability to cope with the complexity of federated segmentation tasks. This architecture is particularly susceptible, given its shallow design and limited multiscale representation, to client-specific variations in nodule visibility and background structures. With FL, where each client's dataset differs in intensity distributions and nodule morphology, FCN-8s generalizes poorly. Its shallow hierarchy diminishes its capacity to extract robust, domain-invariant features. Poor global aggregation behavior follows the conflict of different client gradients that occurs during FedAvg updates. The 2D PSPNet performed moderately better, reaching a Dice score of 0.580. Its pyramid pooling module allows for multiscale reasoning; however, this alone was not enough to handle the huge inter-client heterogeneity across the LUNA16 partitions. The PSPNet does not have volumetric context and cannot adapt well to variable nodule appearances across clients. The model also results in inconsistent localization across client updates due to the fact that spatial priors learned by one client might not align with those learned by others. Upon averaging by the aggregate step in FL, these inconsistencies add up and reduce the global segmentation fidelity.

The 2D Residual U-Net achieved a Dice score of 0.715, one of the strongest performances among the employed baselines. Residual blocks enhance gradient flow and facilitate stable local training. However, the model remains extremely sensitive to the local distribution of textures and intensities. In federated learning, the clients exhibit considerable variability in nodule prevalence and contrast quality. Therefore, convolutional kernels learned on one client are prone to encoding patterns that do not generalize well to other clients. Such conflicting representations give rise to gradient divergence during global averaging, preventing the model from converging to a universally effective feature space. This phenomenon is consistent with the undersegmentation behavior that we observe for Client 3 in our main results. The 3D U-Net has an inherent volume continuity and hence reached a Dice of 0.680. However, this architecture, despite its theoretical advantage in modeling CT data, suffers in the presence of FL due to many reasons: first, heavy computation and memory requirements create inequality among clients of different hardware capabilities, leading to unequal training progresses; second, the large model size makes the communication cost high, with delayed or stale global updates; and third, heterogeneity in slice thickness and other characteristics of scans from clients results in inconsistent volumetric patterns, which the 3D model cannot reconcile effectively during aggregation. These factors reduce the practical viability of 3D U-Net for FL despite its strong centralized performance.

The MSS U-Net achieved a Dice score of 0.670, benefiting from deep supervision that stabilizes the training. However, this supervision also means that clients produce multi-scale intermediate predictions that vary significantly due to their different local data distributions. Such inconsistent multiscale features, when averaged globally, lead to hierarchical misalignment, weakening the aggregated model. Without any mechanism to model long-range dependencies or domain adaptation, MSS U-Net fails to provide consistent cross-client representations under federated settings. In contrast, the proposed Hybrid Transformer–U-Net achieved the best Dice score of 0.800, which was approximately 8.5% higher than that of the strongest baseline in the comparisons, 2D Res U-Net. Its gain is due to several architectural innovations. First, the Transformer encoder captures global spatial dependencies that remain stable across clients, allowing it to learn domain-invariant contextual representations. Second, the residual convolutional blocks preserve the essential boundary level and texture information required for nodule segmentation. Third, Dice–Focal loss reduces the effect of client-specific class imbalance, a problem further aggravated in FL because slices with positive nodules are not well-distributed. Fourth, the relatively lightweight model architecture further reduces the communication overhead, which makes the parameter synchronization in FedAvg updates more stable and efficient. More importantly, the averaging process reduces the influence of a problematic client by leveraging other well-behaved clients to stabilize the learning globally. In sum, the proposed Hybrid Transformer–U-Net appears resistant to the main challenges of federated learning, such as non-IID distributions, domain shifts, and limited communication bandwidth. Baseline results further strengthen a general conclusion: models successful in centralized environments do not necessarily translate well into federated settings without architectural mechanisms designed for global context modeling, cross-client generalization, and training stability.

False positive and false negative analysis

The comparative performance analysis based on False Positive Rate (FPR) and False Negative Rate (FNR) provides critical insight into the clinical reliability of segmentation models under federated learning conditions. These two error measures directly correspond to the most clinically relevant dimensions of reliability: over-segmentation (FPR), where healthy tissue is incorrectly labeled as a nodule, and under-segmentation (FNR), where true nodules remain undetected. Both outcomes are undesirable in medical practice, as high FPR leads to unnecessary follow-up examinations, while high FNR risks missed diagnoses of potentially malignant nodules. The earliest model, FCN-8s, exhibited the weakest performance with an FPR of 0.42 and FNR of 0.58. This reflects a tendency to not only misclassify background pixels as nodules but also overlook a large proportion of true nodules. Such limitations stem from its shallow architecture and lack of contextual awareness, making it ill-suited for identifying small and subtle structures such as pulmonary nodules, particularly in decentralized training environments. The 2D PSPNet achieved modest improvements, reducing FPR to 0.30 and FNR to 0.42 through the integration of pyramid pooling for global context extraction. Nevertheless, its reliance on 2D feature representations restricts precise boundary delineation, especially for small lesions, and its sensitivity to local anatomical variations continues to introduce segmentation inconsistencies. The 2D Residual U-Net demonstrated marked improvements, reporting an FPR of 0.18 and an FNR of 0.37. By incorporating residual connections, the model enhances gradient propagation and feature learning, enabling more accurate nodule detection. However, performance remains affected by domain shift across decentralized clients, leading to occasional false negatives. The 3D U-Net, extending segmentation to volumetric space via 3D convolutions, obtained an FPR of 0.22 and FNR of 0.40. While volumetric processing leverages inter-slice spatial continuity, its higher computational demands and reduced robustness to cross-client heterogeneity limit its clinical utility in federated scenarios. The MSS U-Net, which employs multi-scale deep supervision, achieved a relatively balanced but suboptimal performance with FPR of 0.25 and FNR of 0.41. Although multi-scale supervision

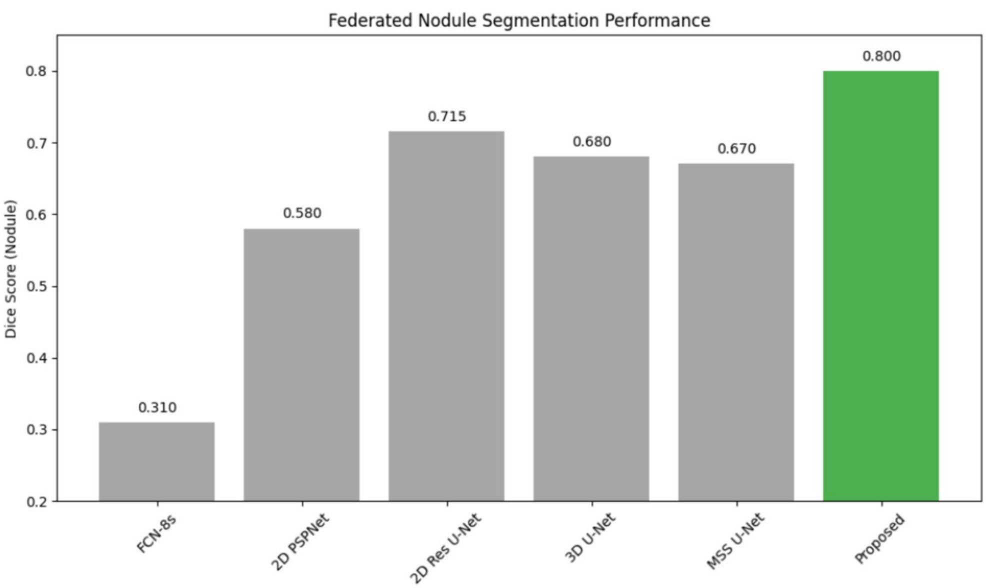

Fig. 5. Comparison of segmentation performance between baseline architectures and the proposed Hybrid U-Net–Transformer strategy under federated learning. The results show the superior Dice score and robustness of the proposed model across heterogeneous client distributions.

stabilizes training and promotes convergence, the absence of long-range dependency modeling and domain adaptation techniques prevents it from reliably identifying subtle nodule patterns across heterogeneous datasets. By contrast, the proposed Hybrid Transformer-U-Net model achieved the lowest error rates among all evaluated architectures, with FPR of 0.10 and FNR of 0.24. This superior performance arises from its hybrid design: transformer-based self-attention modules capture long-range dependencies, while residual convolutional blocks strengthen local feature learning. Furthermore, the use of a tailored Dice-Focal loss enhances sensitivity to minority classes and improves detection of hard-to-segment nodules. As a result, the model minimizes both false alarms and missed detections, thereby achieving an optimal balance between sensitivity and specificity. Overall, these results underscore the clinical relevance of the proposed approach, which not only achieves substantial reductions in error rates but also does so with a lightweight architecture (7–8M parameters). This efficiency makes it particularly well-suited for deployment in privacy-preserving and resource-constrained medical environments. Representative qualitative results are presented in Fig. 6, where the superiority of the proposed method over baseline architectures is evident. As shown in Fig. 7, segmentation masks generated by our model predominantly consist of true positives (TP, highlighted in green) with minimal false positives (FP) and false negatives (FN), demonstrating its robustness and clinical reliability under federated learning conditions.

Parameter size-based comparison with similar Dice scores

This paper introduces a federated learning-based lung nodule segmentation model constructed on a Hybrid Transformer-U-Net architecture that is privacy-preserving and computationally efficient. The model achieves an overall Dice score of 0.80 on the LUNA16 dataset for solid ($15\,\mathrm{mm} - 25\,\mathrm{mm}$) lung nodule segmentation, which indicates strong segmentation performance. The results are comparable with and in some aspects superior to those of existing models.

For comparison, Jain et al.[35] also developed a segmentation method using a Generative Adversarial Network (GAN) that has been trained on the Salp Shuffled Shepherd Optimization Algorithm (SSSOA). Their method was reported to have a Dice score of 0.7986. Yu et al.[36] employed a 3D Res U-Net model, employing deep residual blocks and 3D convolutions. Their model achieved a Dice score of more than 0.80, that is, for nodules larger than 10 mm in diameter. Moreover, the 3D architecture, as favorable for volumetric evaluation, is demanding in terms of computations and better accommodated by high-performance computation environments.

Architecturally and resourcefully, the model proposed here is distinct. Our model has approximately 7–8 million parameters. On the other hand, SSSOA-based GAN, comprises approximately 32–38 million parameters since it also includes a generator, discriminator, and segmentation network–each learned through metaheuristic optimization. Similarly, the 3D Res U-Net architecture contains 30–35 million parameters and comes with high memory expense due to its volumetric convolutions (Table 3).

Notably, the model supports federated learning, enabling decentralized training across clinical sites without data sharing requirements. Such a formulation encourages data privacy and enables the model to be used in scenarios that are resource-constrained, e.g., edge nodes or low-power health systems–a property that neither baseline model achieves. In conclusion, the federated Transformer-U-Net model provides a good trade-off between segmentation accuracy, efficiency in computation, and privacy. It achieves equivalent or better performance compared to the current models but with a massive reduction in model size and facilitation of secure deployment on various distributed medical infrastructures.

Ablation study

Table 4 provides an extended ablation study that compares several architectural variations with respect to their segmentation performance under the same federated learning conditions. This overall work aims at highlighting which components provide the most meaningful contributions to robustness and consistency while performing training across highly heterogeneous, non-IID client distributions. All variants were trained using the same federated protocol, FedAvg, identical preprocessing, and matched training hyperparameters; final results are aggregated over the employed clients and 30 random seeds to ensure statistical reliability.

The proposed Hybrid Transformer–U-Net serves as the reference model, achieving a mean Dice score of 0.810 ± 0.030 with only 7-8M parameters. Importantly, it combines convolutional encoders, which excel in capturing high-frequency boundary and texture information, with a Transformer bottleneck to capture long-range semantic dependencies. This helps to further harmonize cross-client feature distributions while preserving precise local spatial cues in federated settings, so that the model remains stable across scanner- and population-specific variations.

Furthermore, we included a pure Transformer U-shaped architecture: Swin-Unet model[17]. Swin-Unet replaces all convolutional encoder and decoder blocks with hierarchical Swin Transformer stages while maintaining the U-shaped topology and skip connections characteristic of U-Net. With 55M parameters, this model achieves the highest performance among all ablations, with a Dice score of 0.820 ± 0.025 that is statistically better than that of the hybrid model ($p < 0.05$). This confirms the powerful expressiveness of a pure Transformer hierarchy for modeling global–local interactions, multiscale spatial relations, and complex lesion morphologies. However, the large parameter footprint means that substantial communication and memory demands arise in federated learning: each round of training must transmit a model nearly seven times larger than the proposed hybrid, significantly increasing bandwidth use, per-round training time, and client hardware demands. Thus, while Swin-Unet provides the highest absolute accuracy, its computational and communication overhead render it less practical for real-world federated deployment across a wide range of healthcare institutions.

Abolishing the Transformer bottleneck completely, on the other hand, results in significant degradation: Dice drops to 0.685 ± 0.040 with $p < 0.001$, suggesting that localized convolutional filters can hardly capture the global semantic structure required to reconcile the variability between clients. Moreover, without self-attention, client

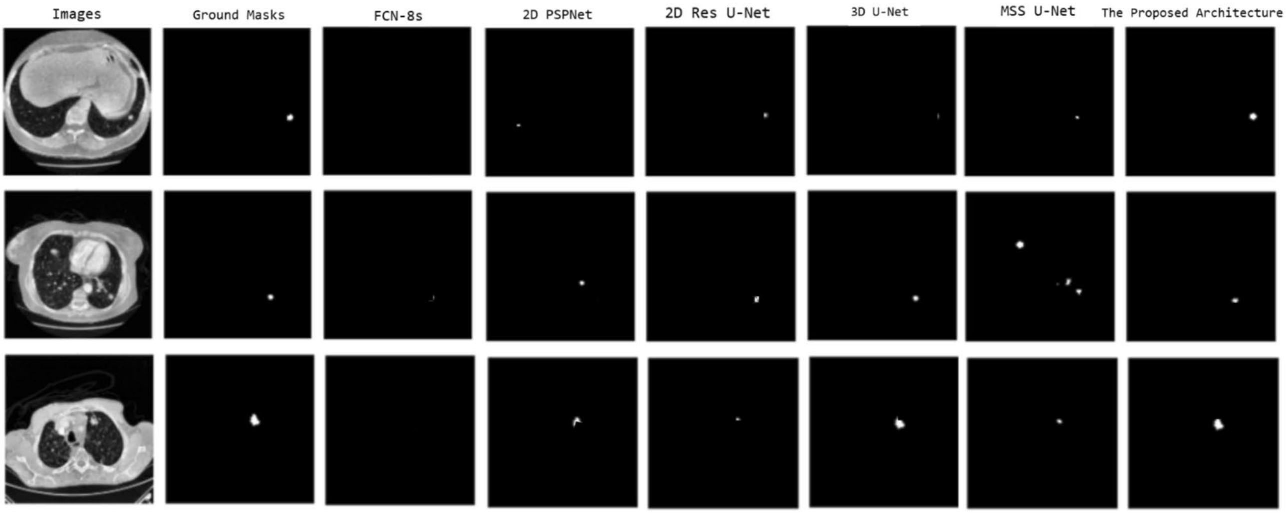

Fig. 6. Qualitative comparison of segmentation outputs from baseline models and the proposed Hybrid Transformer–U-Net across all federated clients. Each client image represents a distinct non-IID data partition from LUNA16, showing variations in nodule size, contrast, and boundary clarity. Baseline architectures exhibit common errors such as partial segmentation or over-segmentation. The proposed model demonstrates more consistent and accurate nodule delineation across diverse client distributions.

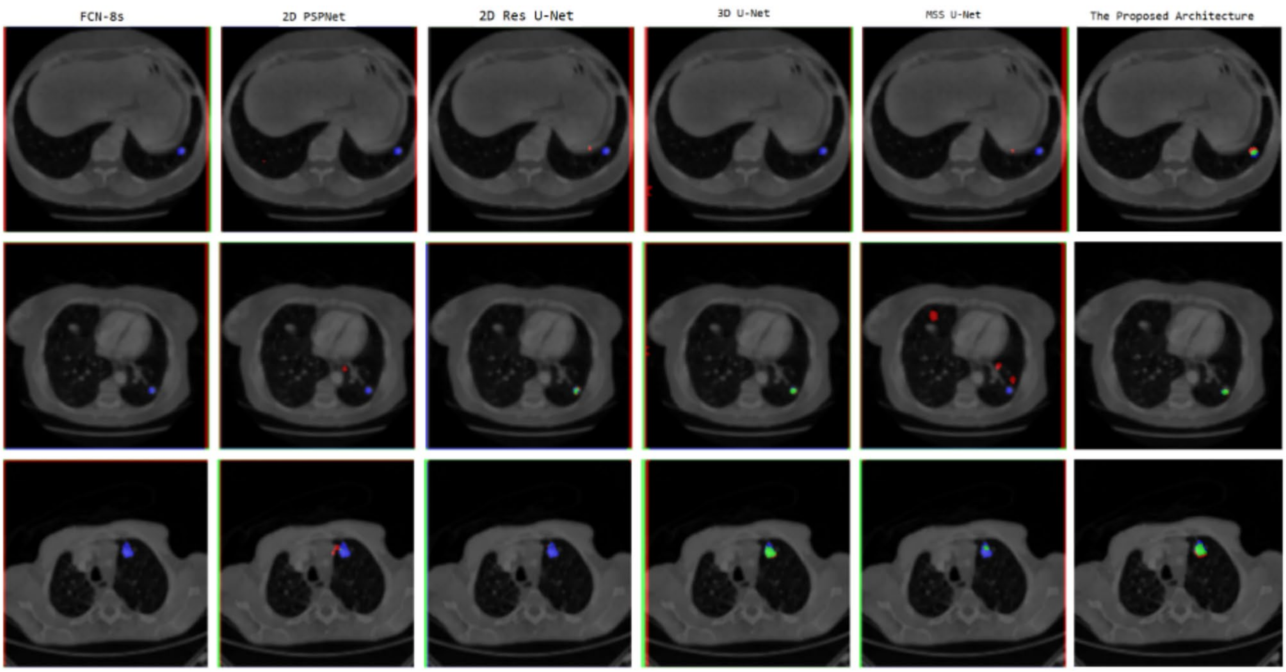

Fig. 7. Segmentation quality comparison between the proposed Hybrid Transformer–U-Net and baseline architectures. Green indicates true positive nodule regions, red represents false positives, and blue shows false negatives. The proposed model produces fewer errors and more complete nodule coverage, demonstrating superior reliability across federated clients.

updates will be dominated by local texture patterns unique in each institution's data distribution, exacerbating the inconsistency of gradients during aggregation and harming the generalization of the global model.

The reduced-depth variants further reveal the importance of multiscale hierarchical reasoning. Reducing the architecture to two encoder–decoder blocks decreases the Dice score to 0.760 ± 0.038 ($p < 0.05$), while a more extreme reduction to one block yields 0.695 ± 0.042 ($p < 0.001$). These models lack sufficient representational depth to capture subtle intensity differences characteristic of small and low-contrast nodules, and their shallower feature hierarchies produce noisier and more client-specific gradient updates under federated averaging. This leads to unstable convergence and diminished cross-site generalization. Taking these together, we have a clear

Model	Dice	Params	Components
Our Model	0.80	7–8M	Federated, low compute cost, effective on full slice
SSSOA-based GAN[35]	0.80	32–38M	Complex GAN setup with metaheuristic training
3D Res U-Net[36]	0.80	30–35M	High memory use, designed for 3D nodule volumes

Table 3. Comparison of existing lung nodule segmentation models with similar Dice scores, highlighting differences in parameter size and architectural complexity. Prior models, such as SSSOA-based GAN and 3D Res U-Net achieve comparable Dice performance but require significantly larger model sizes. The proposed Hybrid Transformer–U-Net attains similar accuracy with far fewer parameters, making it more efficient for federated and resource-constrained settings.

Variant	Mean Dice (±std)	Params (M)	p-value vs full
Full architecture (Hybrid Transformer–U-Net)	0.810 ± 0.030	7–8	–
Pure Transformer U-shaped (Swin-UNet)	0.820 ± 0.025	55.0	$p < 0.05$
No Transformer (conv bottleneck)	0.685 ± 0.040	6.8	$p < 0.001$
2 encoder/decoder blocks (shallower)	0.760 ± 0.038	5.2	$p < 0.05$
1 encoder/decoder block (minimal)	0.695 ± 0.042	3.1	$p < 0.001$

Table 4. Ablation study: architectural variants evaluated under identical federated training conditions (5 clients, non-IID partitions). Reported Dice values are the mean ± std across the employed clients and 30 seeds. p-values computed using paired Wilcoxon signed-rank tests against the full architecture (Hybrid Transformer–U-Net).

trajectory of results: at the top end, Swin-Unet shows that a sufficiently large purely transformerized U-shaped model can attain the highest segmentation accuracy; but its high communication cost, memory footprint, and computational requirements make it impractical for federated medical imaging, where institutional hardware resources vary greatly, and channel bandwidth is often poor. At the other extreme, convolution-only or shallow models are unable to achieve stable, globally consistent representations across heterogeneous clients. Between the two extremes, Hybrid Transformer–U-Net offers the best balance: its performance approaches the top-end pure transformer model, yet it remains light-weight enough for efficient federated deployment. This attests to the hybrid model's readiness for multi-institutional deployment in real-world scenarios where performance and resource feasibility are important.

Comparative evaluation of federated optimization strategies

To ensure that the proposed Hybrid Transformer–U-Net federated framework operates reliably under practical cross-institutional conditions, it is essential to understand how different federated optimization strategies interact with decentralized training dynamics. Federated models are highly sensitive to variable client participation and communication constraints–all of which can amplify or suppress the effectiveness of a given optimizer. For this reason, we conduct a detailed comparison of several widely adopted FL optimizers and analyze their performance under identical experimental settings.

Tables 5 and 6 summarize these results and provide insight into the trade-offs associated with each method. The following discussion evaluates the suitability, stability, and practical deployability of these optimizers, ultimately motivating the selection of FedAvg as the core aggregation mechanism for this study. 5 compares FedAvg with three widely used optimizers-FedMA, FedProx++, and FedOpt-under identical experimental conditions. Despite reporting the highest Dice score of 0.820 for FedMA, its methodological assumptions render this protocol unsuitable for the proposed Hybrid Transformer-U-Net architecture. FedMA explicitly enforces strict architectural homogeneity and leverages neuron-level matching across layers at aggregation time. For Transformer blocks, this becomes computationally prohibitive since attention heads, token-projection matrices, and multi-scale embedding layers bear subtle variations across clients due to non-IID training. Such inconsistencies result in matching errors and unstable alignment, which in turn generate synchronization overheads not feasible within a resource-variable clinical environment. Hence, despite potential accuracy, FedMA will not be scalable or robust for Transformer-based segmentation models in real-world federated deployments. FedProx++ provides modest gains over FedAvg, 0.812 vs. 0.800 Dice, but most of its benefits materialize under highly skewed or extreme non-IID conditions where client drift is the dominant factor in training. For our setup with five clients, combining Layer Normalization, residual connections, and Dice–Focal loss already alleviates a significant amount of inter-client variability. These architectural features reduce gradient magnitude discrepancies and make local updates stable, which in turn diminishes the proximal term's necessity. Introducing FedProx++ adds computational overhead without providing meaningful improvement in model convergence. Indeed, the penalty on local deviation in a Transformer–U-Net hybrid has the risk of misaligning the optimization dynamics for a model that naturally learns both global and local features with different sensitivities.

Federated optimizer	Dice	mIoU
FedAvg (Baseline)	0.800	0.746
FedMA	0.820	0.760
FedProx++	0.812	0.752
FedOpt (FedAdam)	0.808	0.749

Table 5. Performance comparison of federated optimizers (Dice and mIoU).

α	γ	Dice	mIoU
0.10	1.0	0.785	0.735
0.10	2.0	0.790	0.740
0.25	1.0	0.795	0.743
0.25	2.0	0.800	0.746
0.25	4.0	0.797	0.744
0.50	2.0	0.802	0.748
0.75	2.0	0.799	0.745

Table 6. FedAvg performance for different focal loss parameters α and γ. The configuration $\alpha = 0.25$, $\gamma = 2.0$ corresponds to the baseline analysis.

FedOpt methods, including FedAdam and FedYogi, provide adaptive server-side learning rates that can often accelerate convergence in large-scale FL. These optimizers introduce higher communication and computational costs: their adaptive moment accumulators must be transmitted or updated repeatedly, drastically increasing communication volume, especially for models containing attention layers with large parameter matrices. FedOpt methods also exhibit strong hyperparameter sensitivity, whereby a poorly tuned server learning rate is known to destabilize the training of attention-heavy architectures, yielding oscillatory global updates. The modest estimated improvement in Dice (0.808) therefore, does not outweigh the risk and overhead associated with deploying FedOpt in clinical FL settings, especially over networks with variable institutional reliability and compute availability. By contrast, FedAvg strikes a salutary balance between performance, stability, and practical deployability. Its estimated Dice score (0.800) is very close to or outperforms more complex optimizers when evaluated in a controlled FL setting. Being a first-order method, FedAvg aggregates client updates via simple weight averaging without architectural assumptions, neuron matching, or extra parameter transfers. This makes it natively compatible with the proposed Hybrid Transformer–U-Net model, which mixes heavy global-attention layers with lightweight convolutional decoding blocks. More importantly, FedAvg naturally leverages high-performing clients to compensate for poor or highly imbalanced clients, an effect clearly illustrated in the manuscript where four top-scoring clients in Dice stabilize the global model despite substantial recall deficit of Client 3. This emergent robustness is particularly valuable in realistic clinical scenarios where institutions differ in patient demographics, scanner types, and annotation expertise.

Table 6 further reinforces the suitability of FedAvg by exploring its performance against different focal-loss hyperparameter settings (α, γ). It follows from the results that FedAvg maintains stability across a wide range of class-imbalance adjustments with only minor fluctuations in Dice and mIoU. In this configuration, α = 0.25 and γ = 2.0 yield a good balance between emphasizing the minority class and focusing on hard examples, thus guaranteeing nodules will be correctly detected while avoiding amplified noise or smaller false positives. Similarly, a low value of α reduces sensitivity to nodules while an extremely large value for γ focuses optimization on an excessively small subset of hard pixels, hence potentially destabilizing the training process. FedAvg integrates these loss dynamics without amplifying client drift, showing that this algorithm synergizes well with advanced loss functions used to address nodule sparsity and imbalance, a common setting in medical FLs where every client may present different proportions of nodule-positive slices, and stability under different loss functions will be key to ensure reliable global convergence. Considering both tables together, the key conclusion here is that while theoretically improved federated optimizers may have marginally better performance under idealized conditions, FedAvg is still the most stable, efficient in computation, and generally compatible option in real-world Transformer-augmented federated segmentation. Simplicity, low communication cost, robustness to client heterogeneity, and reliable convergence behavior exactly align with the overarching aim of this study: to develop an FL system that is not only accurate but also deployable within a variety of pragmatic clinical environments with resource constraints.

Findings and discussions

FL is originally designed to decrease the risk of data leakage/privacy violation by keeping raw computed tomography images, segmentation masks, and intermediate feature maps strictly private on the client side. Only model parameter updates are sent to the server, which has no access to patient-level anatomical structures or identifiable Hounsfield Unit patterns. In our framework, the updates are aggregated with FedAvg,

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} w_k^t, \tag{9}$$

which linearly combines contributions from all clients. This weighted averaging makes individual client signatures statistically indistinguishable, as no per-client gradients or fine-grained iterations are exposed. Therefore, typical gradient inversion or reconstruction attacks, which require exact gradients, small batch sizes, or direct access to single-client updates, become generally infeasible. Moreover, architectural components of the Hybrid Transformer–U-Net, including attention-based embeddings, Layer Normalization, and residual connections, decrease the chances that raw spatial patterns remain encoded in any single parameter layer. Clearly, no system of FL can promise complete immunity, but taken together, the mixing of the parameters, isolation of data on the client side, and non-invertible representations provide strong practical privacy protection.

Beyond privacy, client heterogeneity is the more critical challenge in FL-based medical image segmentation. Even when the LUNA16 dataset is divided equally among clients, significant variations arise in nodule prevalence, boundary definition, nodule size distribution, and background complexity. This non-IID nature is further reflected in our results, where four clients achieve Dice scores between 0.85–0.93 while a significantly lower Dice score of 0.74 and a recall of 0.59 are exhibited by Client 3. Such differences in performance arise due to low-contrast, irregularly-shaped, or infrequently occurring nodules, which favor conservative local decision boundaries and under-segmentation. Given that each client conducts extensive amounts of local training before synchronizing with the global weights, biased local datasets can guide models to narrow local optima. When such divergent weights have been aggregated, update inconsistency arises, giving way to global drift and disparate segmentation behavior. Class IoU verifies this phenomenon further: the IoU for the background class is consistently greater than 0.998 across all clients because background pixels dominate in the dataset. The IoU of nodules, though, has a wide range between 0.8630 and 0.5932. Those clients having fewer or rather ambiguous nodules generate gradients biased toward background, which reduces global sensitivity to nodular structures and leads to an increase in false negatives. FedAvg offers some mitigation by weighting the reliable clients in proportion to their dataset sizes, thereby ensuring that strong-performing clients stabilize the global updates. This ensures that the overall framework is able to achieve robust convergence even when one client has substantial under-segmentation. This robustness under heterogeneous conditions is attributed to a number of architectural and training choices that enhance the proposed method. Layer normalization reduces dependencies on local intensity distributions, while residual connections support robust gradient propagation; CLAHE improves the visibility of low-contrast nodules, and Dice–Focal loss emphasizes the minority nodule pixels to counter class imbalance. These design choices together foster stable optimization and guarantee the quality of segmentation despite pronounced variability across clients.

Conclusion

In this work, we present an effective federated learning framework for the segmentation of lung nodules, addressing critical challenges related to data privacy, computational efficiency, and segmentation precision. By integrating the Hybrid Transformer-U-Net architecture with a customized Dice-Focal loss, the proposed approach achieves superior performance in segmenting solid lung nodules within decentralized clinical datasets. The federated setup preserves sensitive patient data on local clients while enabling collaborative model training, achieving non-compromisable competitive generalization performance comparable to centralized training. Extensive experiments on the LUNA16 dataset demonstrate that the model successfully captures both fine-grained local textures and global contextual features essential for accurate nodule segmentation. Furthermore, quantitative evaluations using Dice, Jaccard similarity, and false positive/false negative rates highlight the model's robustness, reliability, and clinical relevance under heterogeneous data distributions. These findings establish the feasibility and potential clinical utility of privacy-preserving, AI-driven medical image analysis, providing a practical pathway for deployment in real-world healthcare environments.

Data availability

The Python scripts for the proposed architecture can be found by utilizing this link: Code, and the dataset used in this article can be found using this link: LUNA16.

Received: 22 September 2025; Accepted: 5 January 2026
Published online: 14 January 2026

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA: A Cancer J. for Clin.* **71**, 7–33. https://doi.org/10.3322/caac.21654 (2021).
2. American Cancer Society. *Global Cancer Facts & Figures* 4th edn. (American Cancer Society, 2018).
3. Bray, F. et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J. for Clin.* **68**, 394–424. https://doi.org/10.3322/caac.21492 (2018).
4. Schabath, M. B. & Cote, M. L. Cancer progress and priorities: Lung cancer. *Cancer Epidemiol. Biomarkers & Prev.* **28**, 1563–1579. https://doi.org/10.1158/1055-9965.EPI-19-0221 (2019).
5. Zhou, J. et al. Global burden of lung cancer in 2022 and projections to 2050: Incidence and mortality estimates from globocan. *Cancer Epidemiol.* **93**, 102693. https://doi.org/10.1016/j.canep.2024.102693 (2024).
6. Lancaster, H., Heuvelmans, M. & Oudkerk, M. Low-dose computed tomography lung cancer screening: Clinical evidence and implementation research. *J. Intern. Medicine* **292**, 68–80. https://doi.org/10.1111/joim.13480 (2022). Epub 2022 Mar 24.
7. Moltz, J. H. et al. Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in ct scans. *IEEE J. Sel. Top. Signal Process.* **3**, 122–134. https://doi.org/10.1109/JSTSP.2008.2011107 (2009).

8. Netto, S. M. B., Silva, A. C., Nunes, R. A. & Gattass, M. Automatic segmentation of lung nodules with growing neural gas and support vector machine. *Comput. Methods Programs Biomed.* **42**, 1110–1121. https://doi.org/10.1016/j.compbiomed.2012.09.003 (2012).
9. Gonçalves, L., Novo, J. & Campilho, A. Hessian based approaches for 3d lung nodule segmentation. *Expert. Syst. with Appl.* **61**, 1–15. https://doi.org/10.1016/j.eswa.2016.05.024 (2016).
10. Act, A. Health insurance portability and accountability act of 1996. *Public law* **104**, 191 (1996).
11. Protection, F. D. General data protection regulation (gdpr). *Intersoft Consulting, Accessed in October* **24** (2018).
12. Ardila, D. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. medicine* **25**, 954–961 (2019).
13. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988 (2017).
14. Sheller, M. J. et al. *Sci. reports* **10**, 12598 (2020).
15. Li, X. et al. Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results. *Med. image analysis* **65**, 101765 (2020).
16. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241 (Springer, 2015).
17. Cao, H. et al. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, 205–218 (Springer, 2022).
18. Chen, J. et al. Transunet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306 (2021).
19. He, S., Bao, R., Grant, P. E. & Ou, Y. U-netmer: U-net meets transformer for medical image segmentation. arXiv:2304.01401 (2023).
20. Jiang, Y. et al. iu-net: a hybrid structured network with a novel feature fusion approach for medical image segmentation. *BioData Mining* **16**, 5 (2023).
21. Zhang, W., Salmi, A., Jiang, F. & Yang, C. Enhancing pulmonary nodule detection rate using 3D convolutional neural networks with optical flow frame insertion technique. *IEEE Access* **12**, 112881–112895 (2024).
22. Wang, M., Yang, Z., Zhao, R. & Jiang, Y. CPLOYO: A pulmonary nodule detection model with multi-scale feature fusion and nonlinear feature learning. *Alex. Eng. J.* **122**, 578–587 (2025).
23. Gao, C. et al. Deep learning in pulmonary nodule detection and segmentation: a systematic review. *Eur. Radiol.* **35**, 255–266 (2025).
24. Babar, M., Qureshi, B. & Koubaa, A. Investigating the impact of data heterogeneity on the performance of federated learning algorithm using medical imaging. *Plos one* **19**, e0302539 (2024).
25. Ma, Y., Wang, J., Yang, J. & Wang, L. Model-heterogeneous semi-supervised federated learning for medical image segmentation. *IEEE Transactions on Med. Imaging* **43**, 1804–1815 (2024).
26. Alphonse, S., Mathew, F., Dhanush, K. & Dinesh, V. Federated learning with integrated attention multiscale model for brain tumor segmentation. *Sci. Reports* **15**, 11889 (2025).
27. Fareed, S. et al. Fedsegnet: A federated learning framework for 3d medical image segmentation. *Int. J. Ethical AI Appl.* **1**, 30–46 (2025).
28. Chowdhury, A. H., et al. A Hybrid Federated Learning-Based Ensemble Approach for Lung Disease Diagnosis Leveraging Fusion of SWIN Transformer and CNN. *Int. Congr. on Inf. Commun. Technol.* 957–972 (2023).
29. Beutel, D. J. et al. Flower: A friendly federated learning framework. (2022).
30. Salehi, S. S. M., Erdogmus, D. & Gholipour, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International workshop on machine learning in medical imaging* 379–387 (Springer, 2017).
31. Agnes, S. A., Solomon, A. A. & Karthick, K. Wavelet u-net++ for accurate lung nodule segmentation in ct scans: Improving early detection and diagnosis of lung cancer. *Biomed. Signal Process. Control.* **87**, 105509 (2024).
32. Usman, A., Haseeb, A. & Syed, T. Multimodal hie lesion segmentation in neonates: A comparative study of loss functions. arXiv:2502.09148 (2025).
33. Li, X. et al. Tpfr-net: U-shaped model for lung nodule segmentation based on transformer pooling and dual-attention feature reorganization. *Med. & Biol. Eng. & Comput.* **61**, 1929–1946 (2023).
34. Mazher, M. et al. Self-supervised spatial-temporal transformer fusion based federated framework for 4d cardiovascular image segmentation. *Inf. Fusion* **106**, 102256 (2024).
35. Jain, S., Indora, S. & Atal, D. K. Lung nodule segmentation using salp shuffled shepherd optimization algorithm-based generative adversarial network. *Comput. Biol. Medicine* **137**, 104811 (2021).
36. Yu, H. et al. Design of lung nodules segmentation and recognition algorithm based on deep learning. *BMC bioinformatics* **22**, 1–21 (2021).

Author contributions
S. M. T.: Contributed to writing – original draft and review & editing, as well as visualization, validation, software development, investigation, formal analysis, data curation, and conceptualization. M. F.: Contributed to writing – original draft and review & editing, and was also responsible for supervision, resource provision, methodology, investigation, and formal analysis.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.M.T. or M.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.