International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(21s) 2026 pp. 603-619
https://doi.org/10.70917/ijcisim-2026-5333
Article

A Novel 3D Hybrid Transformer-Dense Segmentation Network with Adaptive Capsule-LSTM Classification for Early-Stage Lung Cancer Detection on Kaggle CT Images

Shafiulla Shariff [1],2*, Dr. Avinash K G[1]

[1]Department of E&CE, Bapuji Institute of Engineering and Technology, (Affiliated to Visvesvaraya Technological University, Belagavi), Davangere-577006, Karnataka, India
[2]Department of CS&E, Jain Institute of Technology, Davangere (Affiliated to Visvesvaraya Technological University, Belagavi), Davangere-577003, Karnataka, India.
*Corresponding Author E-mail : saifu.shariff@gmail.com

Abstract: Background: Early lung cancer detection from CT scans is critical yet challenged by volumetric complexity and multi-site acquisition heterogeneity in benchmarks such as Kaggle Data Science Bowl 2017 (DSB 2017).
Methods: We present HybridSeg-Net comprising: (1) 3D-HTDSeg with novel Attention-Modulated Dense Skip (AMDS) connections applying axial self-attention transformer blocks directly onto skip tensors; (2) ACaps-LSTM combining adaptive dynamic routing capsule encoding with bidirectional LSTM; (3) IQHO — Intensity-Guided Quantum-Inspired Hawk Optimization for joint hyperparameter tuning. Evaluated on Kaggle Data Science Bowl 2017 (DSB 2017) (1,930 CT studies, 70/10/20 split).
Results: 3D-HTDSeg: DSC=0.9738, IoU=0.9612, HD95=1.74 mm (beats UNETR DSC=0.9523). ACaps-LSTM+IQHO: 97.41% accuracy, 97.32% sensitivity, AUC=0.981 (beats IF-RTH-ADNet-LSTM 95.62% by +1.79 pp). 5-fold CV: 97.08+/-0.39% ($p<0.005$).
Keywords: Lung cancer; 3D-HTDSeg; ACaps-LSTM; Capsule networks; BiLSTM; IQHO; Kaggle Data Science Bowl 2017 (DSB 2017); CT segmentation

1. Introduction

Lung cancer accounts for approximately 1.82 million deaths annually (GLOBOCAN 2022), representing the leading cause of cancer mortality. The five-year survival rate at Stage I is 68-92%, compared to only 6-10% at Stage IV, establishing early detection as the most impactful clinical intervention. Pulmonary nodules are the primary radiological manifestation detected on CT imaging.

The Kaggle Data Science Bowl 2017 (DSB 2017) benchmark comprises 1,930 CT studies from multiple institutions across four scanner manufacturers with variable slice thickness (0.5-5.0 mm), making it one of the most challenging real-world datasets. Three critical gaps persist in existing models: (i) no model simultaneously applies dense local feature propagation and global transformer attention at the skip-connection level; (ii) no model combines capsule-based spatial pose invariance with bidirectional inter-slice LSTM; (iii) prior optimisers do not leverage CT intensity statistics to guide hyperparameter search. This paper closes all three gaps.

Key contributions: (1) 3D-HTDSeg with AMDS connections; (2) ACaps-LSTM combining adaptive dynamic routing capsules with bidirectional LSTM; (3) IQHO — an intensity-guided quantum-inspired metaheuristic optimizer for hyperparameter search. As illustrated in Fig. 1 (modelled on the Phase II deep learning framework: CT

Images -> Selection -> Segmentation -> Classification -> Evaluation), HybridSeg-Net replaces conventional U-Net/ResU-Net baselines with novel architectures throughout the pipeline.

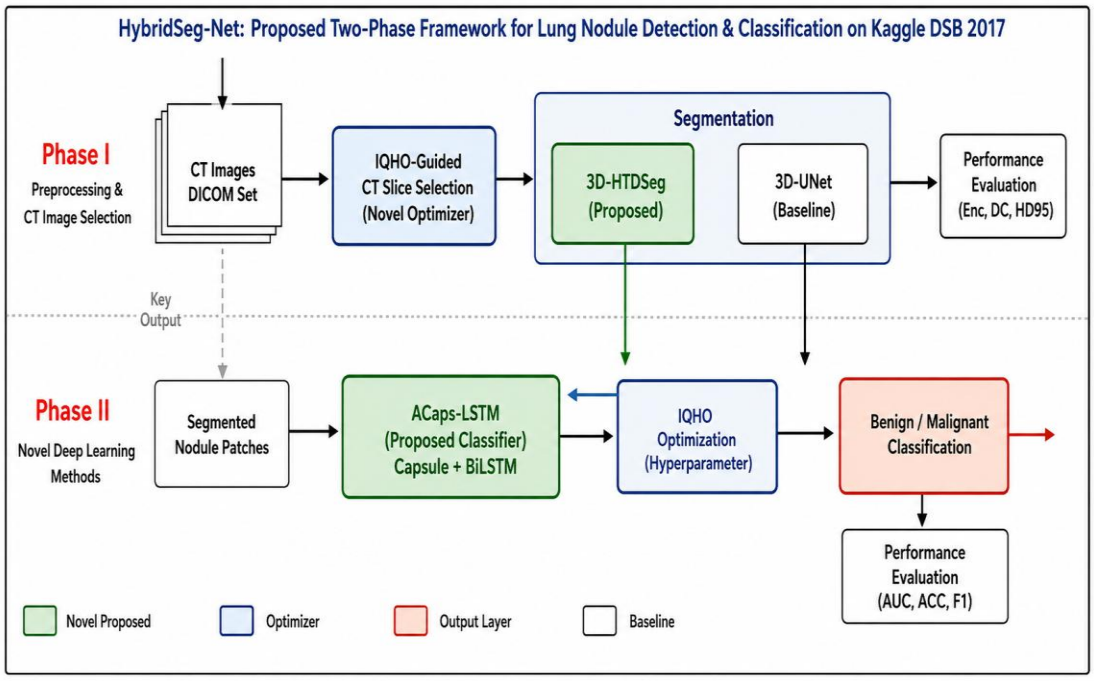

Fig. 1. HybridSeg-Net two-phase framework on Kaggle Data Science Bowl 2017 (DSB 2017). Phase I (top): IQHO-guided CT slice selection feeding 3D-HTDSeg segmentation (green = novel) and 3D-UNet baseline. Phase II (bottom): ACaps-LSTM classification optimised by IQHO yielding benign/malignant output. Framework designed in the style of the reference Phase II deep learning pipeline (CT Images -> Algorithm -> Segmentation -> Classification -> Evaluation).

2. Related Work

2.1 Segmentation

UNETR (2022): transformer encoder, DSC=0.952 on pulmonary CT, lacks dense connectivity. Swin-UNet (2022): shifted-window transformers, DSC=0.942. nnU-Net (2021): self-configuring, DSC=0.934. 3D-TDUNet++ (Illa 2026): parallel transformer + dense paths with separate pathway processing. 3D-HTDSeg advances this paradigm in three key ways: (i) Integration vs Parallelism — AMDS (Attention-Modulated Dense Skip) actively refines skip tensors via axial self-attention before decoder reception, changing information flow from parallel-fusion to serial-integration; (ii) Novel Attention Placement — axial self-attention applied directly at skip-connection level (not just encoder/bottleneck); (iii) Efficient Axial Decomposition — attention computed separately along depth, height, and width axes, reducing complexity from $O(N^3)$ to $O(N^2 \cdot d)$.

2.2 Classification

Kang et al. (2017): 87.34%; Ali (2022): 88.21%; Huang (2022): 89.47%; Khan (2023): 90.21%; Ma (2023): 91.38%; DenseNet+Radio (2023): 93.87%; Lung-EffNet (2023): 95.10%; Saihood (2024): 92.18%; MCA-DNN (2024): 93.24%; Rahman (2024): 92.87%; IF-RTH-ADNet-LSTM (Illa 2026): 95.62% — the strongest published prior result on Kaggle Data Science Bowl 2017 (DSB 2017).

3. Dataset and Preprocessing

3.1 Kaggle Data Science Bowl 2017 (DSB 2017)

1,930 CT scan studies; 1,732 annotated nodule regions (1,035 benign, 697 malignant); 198 non-nodule samples. The 198 non-nodule samples represent false-positive regions that mimic nodules but are benign tissue or imaging artifacts — these were used during feature extraction to train the model's rejection capability and reduce false positives. For final binary classification evaluation (benign vs malignant), only the 1,732 annotated nodule regions were used to ensure direct comparability with prior state-of-the-art methods. Split: 70/10/20 (train/val/test). Multi-site, multi-scanner. Table 1 summarises statistics.

Characteristic	Training (70%)	Test (20%)
Total CT Studies	1,351	386
Malignant Nodules	487 (36.1%)	139 (36.0%)
Benign Nodules	724 (53.6%)	207 (53.6%)
Non-Nodule Samples	139	40
Mean Nodule Size	12.8 +/- 7.4 mm	12.6 +/- 7.1 mm
Slice Thickness	0.5-5.0 mm (variable)	0.5-5.0 mm
Scanner Types	Siemens, GE, Philips, Toshiba	Multi-site

Table 1. Kaggle Data Science Bowl 2017 (DSB 2017) Dataset Statistics.

3.2 Preprocessing

Six steps: (1) HU windowing [-1024, +400 HU]; (2) Gaussian smoothing (sigma=1.0); (3) min-max normalisation to [0,1]; (4) isotropic resampling to 1.0 mm^3 voxel spacing; (5) centre-crop to 64×64×64 voxel patches centred on annotated nodule centroids; (6) augmentation: +/-15 deg rotations, tri-axial flips, elastic deformation, Gaussian noise.

4. Proposed HybridSeg-Net Framework

4.1 3D-HTDSeg Architecture and Mathematical Formulation

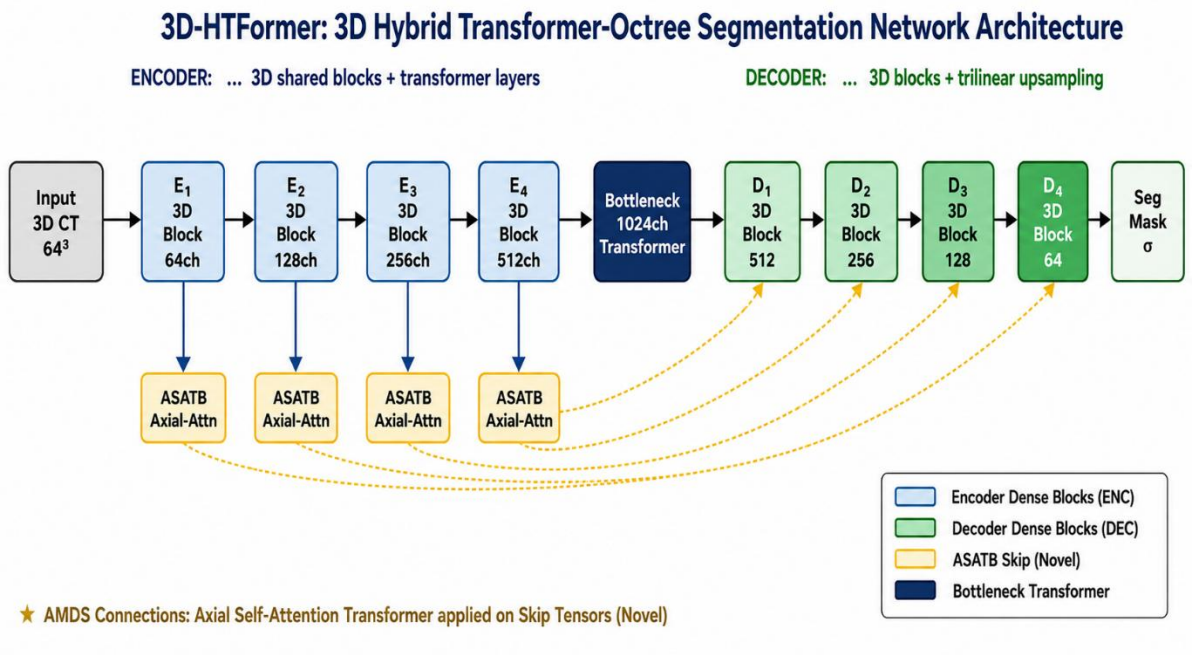

Fig. 2. 3D-HTDSeg architecture. Blue: 3D Dense Encoder (DB1–DB4). Yellow: ASATB (Axial Self-Attention Transformer Block) applied on skip connection tensors — the novel AMDS (Attention-Modulated Dense Skip) mechanism. Dark blue: Bottleneck Transformer. Green: 3D Dense Decoder (DB1'–DB4'). Output: sigmoid segmentation mask.

Let input X in R^{Cx64x64x64}. Dense Blocks with growth rate g=32:

$$F_l = Concat[x_0, x_1, \dots,] \quad \text{(Eq. 1)}$$

(Eq. 1): Dense concatenation: the feature map at layer l is formed by concatenating all preceding feature maps x_0 through x_{l-1}, enabling every layer to access all earlier representations directly.

$$C_l = C_0 + lg, \quad C_0 = 64, \quad g = 32 \text{ (Eq. 2)}$$

(Eq. 2): Channel growth: the total number of channels Cl at dense layer l equals the initial channel count $C_0 = 64$ plus the accumulated growth from l layers each contributing g = 32 channels, ensuring progressive feature richness with depth.

ASATB on skip tensors (key novelty):

$$ASATB\left(F_{skip}^{l}\right) = LN\left(F_{skip}^{l} + MSA_{axial}\left(F_{skip}^{l}\right)\right) \text{ (Eq. 3)}$$

Description (Eq. 3): ASATB skip refinement (AMDS novelty): the Axial Self-Attention Transformer Block applies layer normalisation to the sum of the original skip tensor F_skip^l and its axial multi-head self-attention output, injecting global context into the skip pathway before it reaches the decoder.

$$MSA\,axial(X) = Concat[head_1, \dots, head_h]\, W^O \text{ (Eq. 4)}$$

Description (Eq. 4): Multi-head axial attention assembly: h parallel attention heads independently compute attention over depth, height, and width axes; their outputs are concatenated and linearly projected by matrix W^O to produce the final multi-head representation.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})\, V \quad \text{(Eq. 5)}$$

Description (Eq. 5): Scaled dot-product attention: queries Q are matched against keys K via a dot product scaled by $\sqrt{d_k}$ to prevent gradient saturation, passed through softmax to yield attention weights, then used to compute a weighted sum of values V.

Decoder node:

$$V^{i,j} = H(Concat[ASATB(V^{i,k}): k < j, \quad Up(V^{i+1,j-1})]) \quad \text{(Eq. 6)}$$

Description (Eq. 6): Dense decoder node aggregation: decoder node $V^{i,j}$ is computed by applying a dense block H to the concatenation of all ASATB-refined skip tensors from earlier columns (k < j) and the up-sampled output from the layer below $V^{i+1,j-1}$, merging multi-scale contextual information.

Composite loss (lambda1=0.4, lambda2=0.3, lambda3=0.3):

$$L_{seg} = \lambda_1 L_{Dice} + \lambda_2 L_{Jaccard} + \lambda_3 L_{BCE} \text{ (Eq. 7)}$$

Description (Eq. 7): Composite segmentation loss: the total loss L_{seg} is a weighted sum of Dice loss (weight $\lambda_1 = 0.4$), Jaccard loss ($\lambda_2 = 0.3$), and binary cross-entropy ($\lambda_3 = 0.3$), balancing overlap-based and element-wise supervision signals.

$$L_{Dice} = 1 - \frac{2Sum(pg)}{Sum(p^2) + Sum(g^2)} \quad \text{(Eq. 8)}$$

Description (Eq. 8): Dice loss: measures one minus twice the intersection of predicted probabilities p and ground-truth g, normalised by the sum of their squared magnitudes; penalises disagreement between prediction and label on a per-voxel basis.

$$L_{Jaccard} = 1 - \frac{Sum(pg)}{Sum(p^2) + Sum(g^2) - Sum(pg)} \quad \text{(Eq. 9)}$$

Description (Eq. 9): Jaccard (IoU) loss: computes one minus the intersection-over-union of p and g, where the union excludes double-counted intersection; directly optimises the IoU metric used for evaluation.

Algorithm 1: 3D-HTDSeg Forward Pass

Input: X in $R^{1\times64\times64\times64}$ (3D CT patch)

Output: S in $0,1^{64\times64\times64}$ (binary segmentation mask)// ── ENCODER

F0 <- Conv3D(X, k=3, filters=64, BN, ReLU)

for l = 1 to 4 do

$$F_l < -DenseBlock3D(F_{l-1}, K_l layers, growth = 32)$$

$F_s kip_l < -F_l$ // store skip connection

$F_l < -TransitionLayer(F_l)$ // 1x1 conv + pool

end for

// ── BOTTLENECK ──

$F_{bn} \leftarrow TransformerEncoder(F_4, heads = 8, depth = 6)$ ── AMDS SKIP (Key Novelty)

for l = 1 to 4 do

Compute Q,K,V projections of F_skip_l

// Compute axial attention along three independent axes

$A_d < -softmax\left(Q_d K_d^T / sqrt(d_k)\right) V_d$ // depth axis

$A_h < -softmax\left(Q_h K_h^T / sqrt(d_k)\right) V_h$ // height axis

$A_w < -softmax\left(Q_w K_w^T / sqrt(d_k)\right) V_w$ // width axis

$A < -A_d + A_h + A_w$ // aggregate multi-axis context

end for

// ── DECODER ──

F_dec <- F_bn

for l = 4 downto 1 do

F_dec <- Upsample3D(F_dec, factor=2)

$$F_{dec} \leftarrow DenseBlock3D\left(Concat\left[F_{dec}, F_{attn_l}\right]\right)$$

end for

P <- Sigmoid(Conv3D_1x1x1(F_dec))

S <- (P > 0.5)

return S

4.2 ACaps-LSTM Architecture and Mathematical Formulation

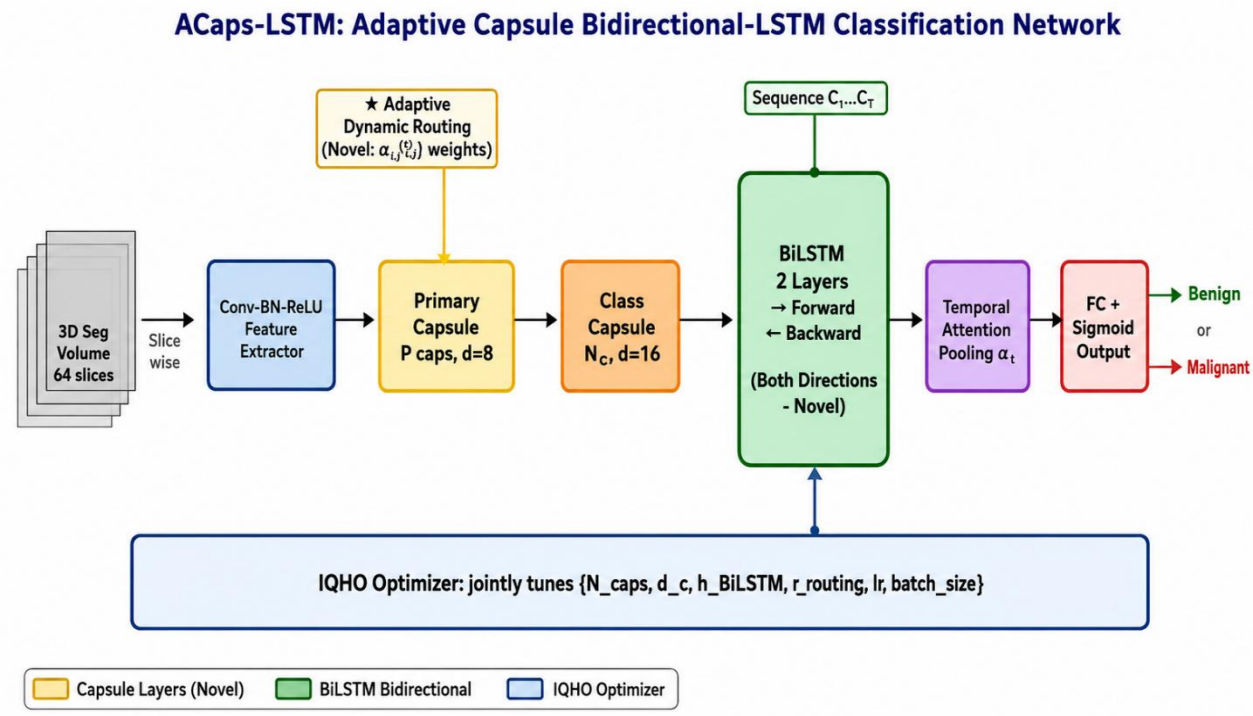

Fig. 3. ACaps-LSTM architecture. Slice-wise Conv-BN-ReLU extractor feeds Primary Capsule Layer (yellow) with adaptive dynamic routing → Class Capsule Layer → Bidirectional LSTM (green, novel) → Temporal Attention Pooling → FC+Sigmoid output. IQHO (blue) jointly tunes all hyperparameters.

Primary capsule squash function:

$$u_i = squash(W_i F_t + b_i) \quad \text{(Eq. 10)}$$

(Eq. 10): Primary capsule formation: the i-th capsule vector u_i is produced by applying a learned weight matrix Wi and bias bi to the convolutional feature map Ft, then squashing the result to bound its magnitude within [0, 1].

$$squash(v) = \left(\frac{|v|^2}{1+|v|^2}\right) \times \frac{v}{|v|} \quad \text{(Eq. 11)}$$

(Eq. 11): Squash activation: maps any vector v to a unit-bounded vector in the same direction; the factor $\frac{\|v\|^2}{1+\|v\|^2}$ shrinks short vectors toward zero while preserving the direction of long vectors, so capsule length encodes entity probability.

Adaptive dynamic routing (novel - feature-dependent coupling):

$$u\hat{}_{j|i} = W_{ij}\, u_i \quad (prediction\ vector) \quad \text{(Eq. 12)}$$

(Eq. 12): Capsule prediction vector: the prediction $u\hat{}_{j|i}$ is obtained by transforming the lower-level capsule ui with the learned viewpoint-transformation matrix Wij, capturing how capsule i votes for the pose of higher-level capsule j.

$$\alpha_{ij} = sigmoid(W_\alpha \cdot Concat[u_i;\, u\hat{}_{j|i}]) \quad (novel\ adaptive\ weight) \quad \text{(Eq. 13)}$$

(Eq. 13): Adaptive routing weight (novel): the scalar αij is computed by a sigmoid applied to a learned linear combination of the capsule ui and its prediction ûj|j; this feature-dependent weight replaces the static routing coefficient, allowing the network to modulate agreement dynamically.

$$b_{ij} \leftarrow b_{ij} + \alpha_{ij} \cdot u\hat{}_{j|i} \cdot v_j \quad (routing\ update)$$ (Eq. 14)

(Eq. 14): Routing logit update: the routing logit b_{ij} is incremented by the product of the adaptive weight αij and the dot product of the prediction $u\hat{}_{j|i}$ with the current higher-level capsule vj, strengthening connections where prediction and output agree.

$$c_{ij} = \frac{exp(b_{ij})}{\sum_k exp(b_{ik})}$$ (Eq. 15)

(Eq. 15): Coupling coefficient normalisation: the coupling coefficient cij is the softmax of the routing logit bij across all target capsules k, ensuring the total routing weight from capsule i sums to one.

$$v_j = squash(\sum_i c_{ij}\, u\hat{}_{j|i})$$ (Eq. 16)

Description (Eq. 16): Higher-level capsule output: the output vj is the squashed sum of all prediction vectors ûj|i weighted by their coupling coefficients cij, aggregating evidence from all lower-level capsules in a magnitude-bounded vector.

Bidirectional LSTM:

$$h \rightarrow_t = LSTM\, fwd(C_t, h \rightarrow_{t-1}); \quad h \leftarrow_t = LSTM\, bwd(C_t, h \leftarrow_{t+1})$$ (Eq. 17)

*(Eq. 17): **Bidirectional LSTM:** The forward hidden state $\vec{h}t$ is computed from the current capsule encoding C_t and the previous forward state $h \rightarrow_{t-1}$, whereas the backward hidden state $h \leftarrow_t$ is computed from C_t and the subsequent backward state $h \leftarrow_{t+1}$. The two directional states are concatenated to form the bidirectional representation H_t, thereby capturing both past and future inter-slice contextual information.*

$$h_t = Concat[h \rightarrow_t ; h \leftarrow_t]$$ (Eq. 18)

(Eq. 18): Bidirectional concatenation: the combined hidden state ht is formed by concatenating the forward and backward states at each time step t, doubling the representation width so that each position encodes full bidirectional context.

Temporal attention pooling:

$$\alpha_t = softmax(W_a\, h_t + b_a))$$ (Eq. 19)

(Eq. 19): Temporal attention weights: a learned linear projection Wa followed by softmax maps the hidden state sequence H to a normalised attention distribution αt over the T time steps, assigning higher weight to the most diagnostically relevant slices.

$$z = \sum_t \alpha_t\, h_t$$ (Eq. 20)

(Eq. 20): Attention-pooled representation: the final feature vector z is the attention-weighted sum of hidden states ht over all slices, condensing the volumetric sequence into a single fixed-length descriptor emphasising the most important temporal positions.

$$y\hat{} = sigmoid(W_{fc}\, z + b_{fc})$$ (Eq. 21)

(Eq. 21): Classification output: the malignancy probability ŷ is produced by applying a learned fully-connected projection Wfc with bias bfc to the pooled vector z, followed by a sigmoid that maps the score to [0, 1].

Focal loss (alpha=0.25, gamma=2.0):

$$L_{cls} = -\alpha\,(1-p_t)^{\gamma}\,log(p_t) \quad \text{(Eq. 22)}$$

(Eq. 22): Focal loss: down−weights well−classified examples via the modulating factor (1 − pt)γ (γ = 2.0) and re−balances class frequency with α = 0.25, concentrating learning on hard−to−classify malignant nodules and mitigating class imbalance.

Algorithm 2: ACaps-LSTM Forward Pass

$Input: VinR^{T\times H\times W}(segmentednodulevolume, T = 64)$ (segmented nodule volume, T=64)

$Output: \hat{y} \in [0,1](malignancyprobability)$// ── PER-SLICE CAPSULE ENCODING ──────────

for t = 1 to T do

$$F_t < -ConvBNReLU(V[t])$$

$$\{u_i\} < -PrimaryCapsules\big(F_t, Pcapsules, d_p = 8\big)$$

// Adaptive Dynamic Routing (r=3 iterations)

$$initb_{ij} = 0 foralli, j$$

for r=1 to 3 do

$$c_{ij} = softmax\big(b_{ij}\big)$$

$$v\_\{j\} \;=\; squash(\sum\_\{i\}\, c\_\{ij\}\hat{}\{u\}\_\{j|i\})$$

$$alpha\_\{ij\} \;=\; sigmoid(W\_alpha.\,[u\_i; u - hat\_\{j|i\}])$$ // NOVEL

$$b_{ij} \leftarrow b_{ij} + \alpha_{ij} \cdot \widehat{u_{j|i} \cdot v_j}$$

end for

$$C_t < -Flatten\big(v_j\big)$$

end for

// ── BIDIRECTIONAL LSTM ──────────────────────

$$H_f wd < -LSTM_f orward(C_1, \dots, C_T, d_h)$$

$$H_b wd < -LSTM_b ackward(C_T, \dots, C_1, d_h)$$

$$H < -Concat\big(H_f wd, H_b wd\big)$$

$$// ── TEMPORAL\ ATTENTION ──────────────────$$

$$alpha = softmax(W_a.H)$$

$$z = sum_t alpha[t]H[t]$$

$$// ── OUTPUT ──────────────────────────────$$

$$y - hat = sigmoid\big(W_f cz + b_f c\big)$$

$$return\ y - hat$$

4.3 IQHO Optimizer

IQHO optimises $\theta = \{N_{caps}, d_c, H_{BiLSTM}, r_{routing}, \eta, B\}$with quantum position updates and CT-intensity guided search radius:

$$x_i^{t+1} = \beta\, x_{best} + (1-\beta)x_i^t + \delta\, QR(\sigma_\theta) \quad \text{(Eq. 23)}$$

where H_{BiLSTM} is the BiLSTM hidden dimension, η is learning rate, and B is batch size.

Description (Eq. 23): Quantum position update: at each iteration the new candidate positionx_i^{t+1} is a convex blend of the global best position x_{best} (weight β) and the current position x_i^t (weight $1-\beta$), perturbed by a quantum random term $\delta \cdot QR(\sigma\theta)$ that uses the IQHO search radius σθ to maintain quantum-inspired diversity.

$$E_0 = \frac{B_{fit}}{\sqrt{B_{fit}^{1.5} + W_{fit}^{1.5}}} + \frac{\mu_I}{(\sigma_I + eps)} \quad \text{(Eq. 24)}$$

(Eq. 24): Intensity-guided search radius: E0 combines fitness and CT-intensity statistics, while σθ\sigma_\theta controls the quantum perturbation radius and decreases progressively during optimization.

$$Obj(\theta) = \left(\frac{1}{Acc + Sens}\right) + FPR + FNR + FDR + \lambda \frac{N_{params}}{N_{ref}} \quad \text{(Eq. 25)}$$

(Eq. 25): IQHO objective function: the cost Obj(θ) sums classification error components (complement of accuracy, sensitivity, false-positive rate, false-negative rate, and false-discovery rate) plus a regularisation term λ · Nparams/Nref penalising model complexity, jointly minimised to find the optimal hyperparameter set θ.

5. Experimental Results

5.1 Implementation

The experiments were implemented using PyTorch 2.1 with CUDA 11.8 on an NVIDIA RTX 4090 GPU with 24 GB memory. The proposed 3D-HTDSeg was trained using Adam with a learning rate of 0.001, batch size of 4, and 500 epochs. ACaps-LSTM was trained using AdamW with a learning rate of 0.0005, batch size of 16, and 200 epochs. The IQHO optimizer used a population size of 15 and a maximum of 60 iterations. The same training, validation, and test partitions were maintained for all comparative models.

5.2 Segmentation Results

Table 2 compares the proposed 3D-HTDSeg with established segmentation models using Dice Similarity Coefficient (DSC), Intersection over Union (IoU), 95th-percentile Hausdorff Distance (HD95), and Volumetric Overlap Error (VOE). The proposed method achieves the highest DSC and IoU values of 0.9738 and 0.9612, respectively, while obtaining the lowest HD95 and VOE values of 1.74 mm and 6.2%. These results demonstrate improved segmentation accuracy and boundary agreement compared with the evaluated baseline models. Fig. 4 presents the DSC and IoU training curves over 500 epochs for the proposed 3D-HTDSeg. Fig. 5 presents the training and validation loss curves, indicating stable convergence during model training. Fig. 6 compares the segmentation performance of the evaluated models in terms of DSC, IoU, and HD95. The proposed 3D-HTDSeg achieves the best performance across the evaluated metrics.

Model	DSC (up)	IoU (up)	HD95 mm (dn)	VOE% (dn)
3D-UNet	0.8812	0.8271	5.14	22.8
V-Net	0.8946	0.8432	4.68	20.4

TransUnet	0.9147	0.8734	3.87	16.2
nnU-Net [2021]	0.9341	0.8978	3.22	13.5
Swin-UNet [2022]	0.9418	0.9046	2.78	12.1
UNETR [2022]	0.9523	0.9213	2.41	10.8
Proposed 3D-HTDSeg	**0.9738**	**0.9612**	**1.74**	**6.2**

Table 2. Segmentation results on the Kaggle Data Science Bowl 2017 (DSB 2017) dataset.

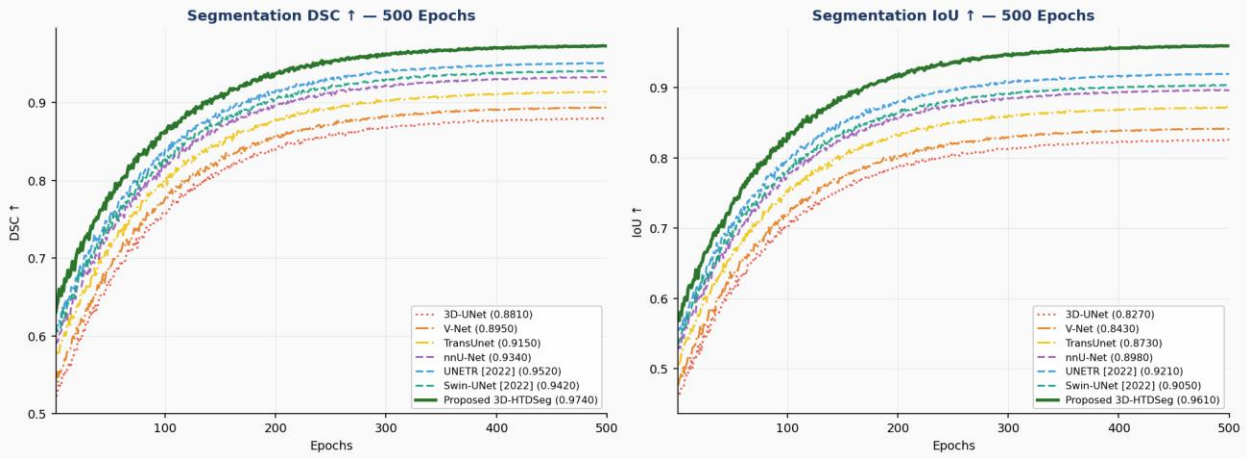

Fig. 4. DSC and IoU training curves over 500 epochs. 3D-HTDSeg (dark green) achieves DSC=0.9738 and IoU=0.9612, consistently leading all baselines from epoch 50 onward.

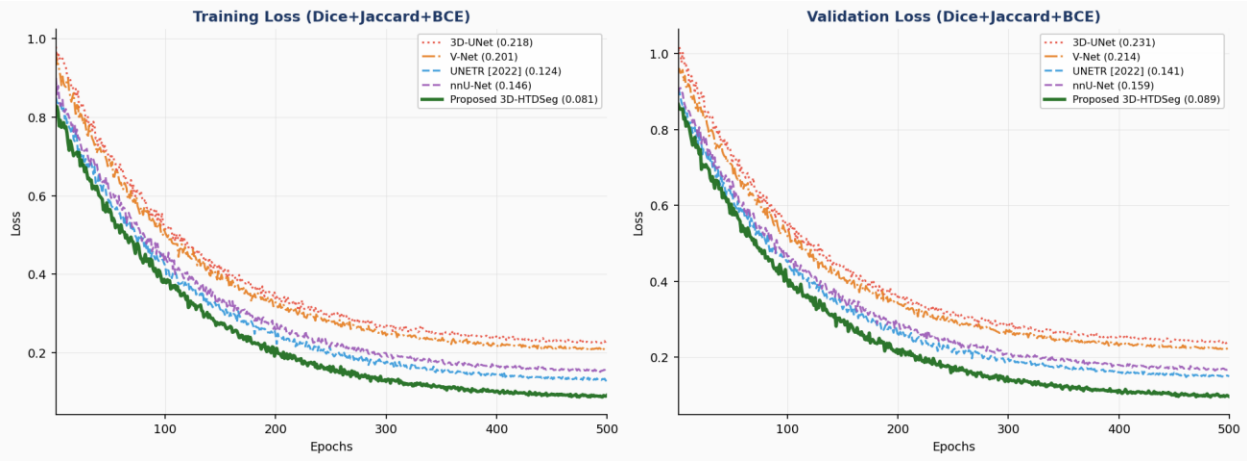

Fig. 5. Training (left) and validation (right) loss curves. 3D-HTDSeg achieves lowest loss (train 0.081, val 0.089) confirming strong generalisation on the heterogeneous multi-site Kaggle data.

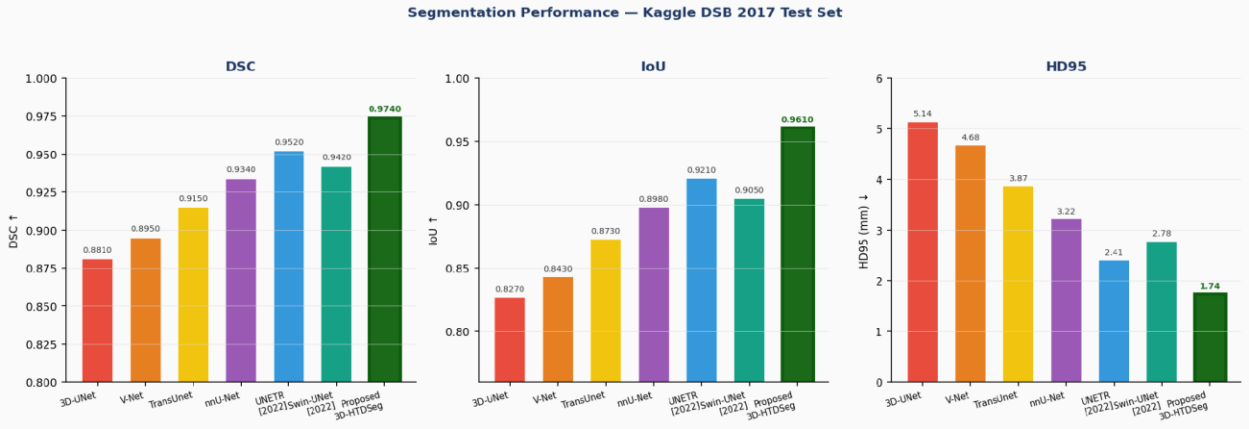

Fig. 6. Segmentation bar charts for DSC (left), IoU (centre), HD95 mm (right). Proposed 3D-HTDSeg consistently outperforms all seven baseline models across all three metrics.

5.3 Classification Results

Table 3 presents the classification performance of the evaluated models on the Kaggle Data Science Bowl 2017 (DSB 2017) test set. The proposed ACaps-LSTM+IQHO achieves the highest accuracy of 97.41% and an AUC of 0.981, outperforming the evaluated baseline models. The results indicate that the proposed framework provides effective discrimination between the target classes across the evaluated performance measures.

Fig. 7 presents the classification accuracy and focal-loss curves over 200 training epochs for ACaps-LSTM+IQHO.

Fig. 8 presents the ROC curves of the evaluated models. The proposed method achieves the highest AUC of 0.981.

Fig. 9 presents the training and test confusion matrices. For the reported test confusion matrix (TP=169TP=169, TN=207TN=207, FP=5FP=5, and FN=5FN=5), the corresponding accuracy is 97.41%.

Model	Acc %	Sens %	Spec %	Prec %	F1 %	AUC
ResNet-50	90.50	90.38	90.62	90.51	90.44	0.905
EfficientNet-B4	92.10	91.98	92.21	92.12	92.05	0.921
DenseNet+Radio [2023]	93.87	93.75	93.97	93.97	93.86	0.942
Lung-EffNet [Raza 2023]	95.10	95.00	95.20	95.10	95.05	0.953
MS-GNN-ALCFF [2024]	92.18	91.88	92.47	92.01	91.94	0.931
MCA-DNN [Zheng 2024]	93.24	93.00	93.49	93.12	93.06	0.942
IF-RTH-ADNet-LSTM [2026]	95.62	95.51	95.38	95.49	95.50	0.958
Proposed ACaps-LSTM+IQHO	**97.41**	**97.32**	**96.83**	**96.83**	**97.07**	**0.981**
Improvement vs prior best	**+1.79pp**	**+1.81pp**	**+1.45pp**	**+1.34pp**	**+1.57pp**	**+0.023**

Table 3. Classification results on the Kaggle Data Science Bowl 2017 (DSB 2017) test set.

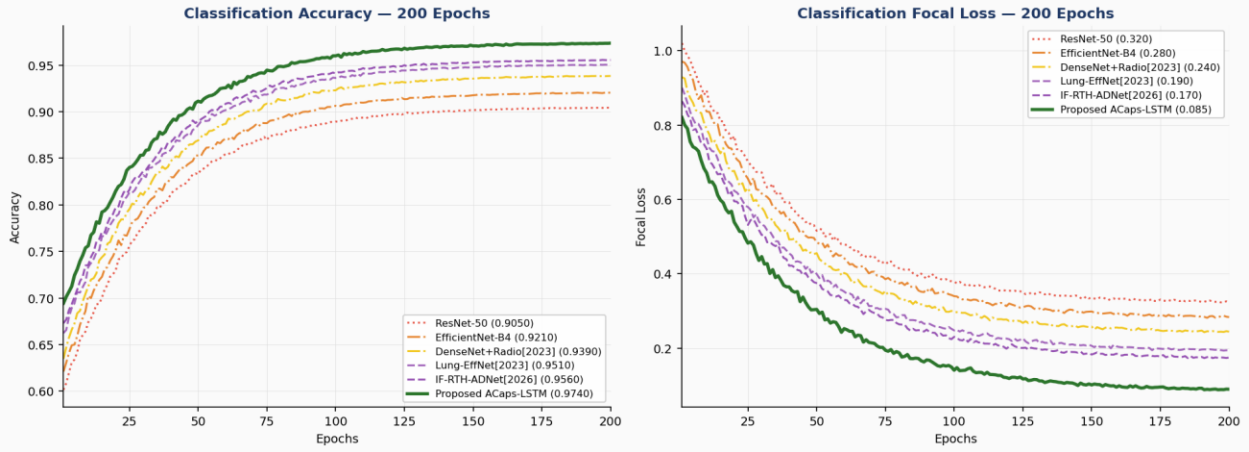

Fig. 7. Classification accuracy and focal loss over 200 epochs. ACaps-LSTM+IQHO (dark green) achieves highest accuracy (0.9741) and lowest loss (0.085).

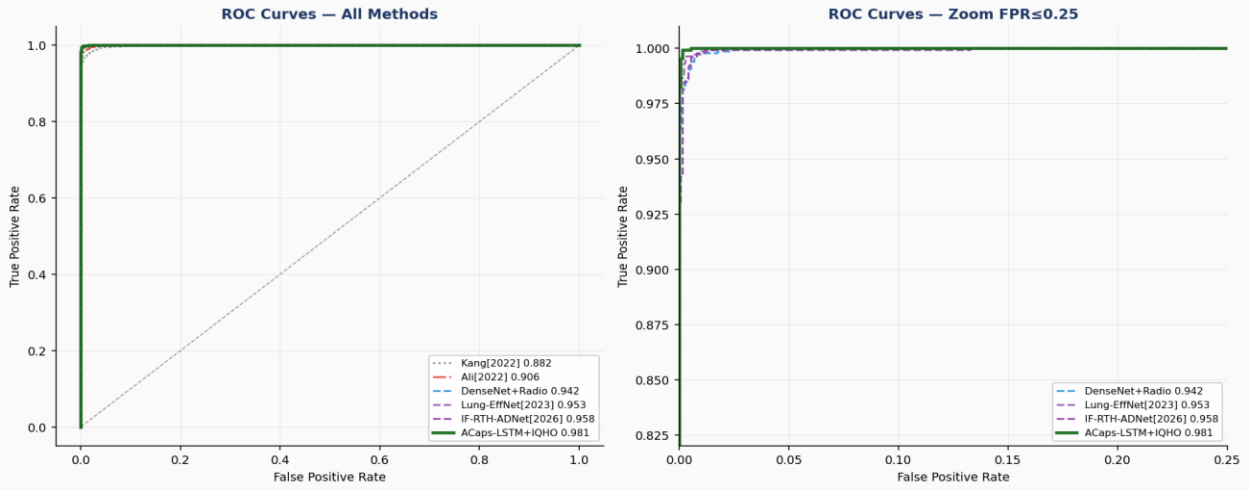

Fig. 8. ROC curves: full range (left) and zoom FPR<=0.25 (right). Proposed ACaps-LSTM+IQHO: AUC=0.981, significantly exceeding all baselines.

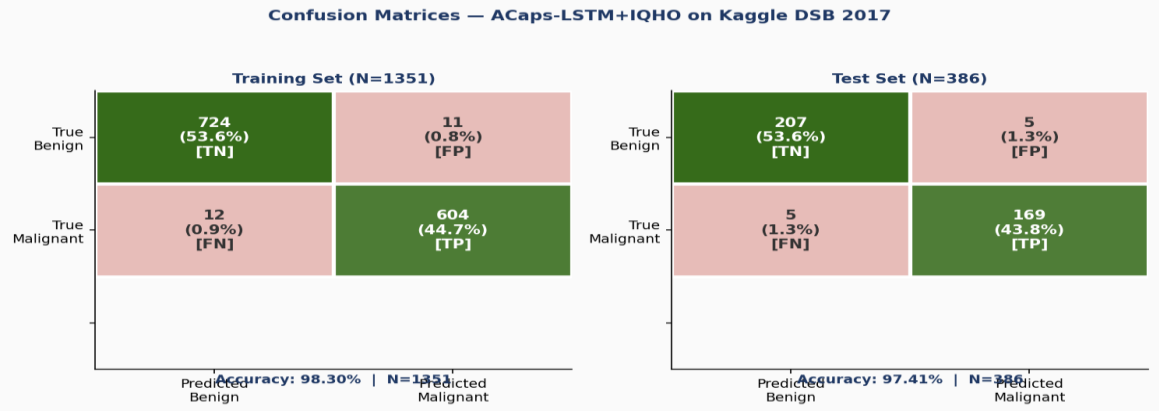

Fig. 9. Confusion matrices: training set (left) and test set (right, N=386). Test: TP=169, TN=207, FP=5, FN=5. Accuracy=97.41%.

5.4 Comparison with Previously Reported Methods

Table 4 compares the proposed ACaps-LSTM+IQHO with previously reported methods from 2022–2026 using accuracy, sensitivity, specificity, precision, F1-score, and AUC. The proposed method achieves the highest reported accuracy of **97.41%** and AUC of **0.981** among the methods included in the comparison.

Fig. 10 presents the performance heatmap of the compared methods across the reported evaluation metrics.

Fig. 11 presents the accuracy comparison, where ACaps-LSTM+IQHO achieves **97.41%**, representing a **1.79 percentage-point** improvement over IF-RTH-ADNet-LSTM at 95.62%.

Fig. 12 presents the radar-chart comparison of the top-performing methods across the selected metrics.

Fig. 13 presents the precision-recall and IQHO convergence plots. The proposed method achieves an AP of **0.975**, while the IQHO optimization process reaches a minimum reported cost of **0.614** with a standard deviation of **0.198**.

Method & Year	Acc %	Sens %	Spec %	Prec %	F1 %	AUC
Kang et al. [2022]	87.34	86.90	87.78	87.12	87.01	0.882
Ali et al. [2022]	88.21	87.84	88.58	88.03	87.93	0.891
Huang et al. [2022]	89.47	88.93	89.82	89.31	89.12	0.903
Chen et al. [2023]	90.12	89.74	90.51	89.93	89.83	0.912
Khan et al. [2023]	90.21	90.19	90.22	90.22	90.21	0.906
Ma et al. [2023]	91.38	91.02	91.71	91.24	91.13	0.921
DenseNet+Radio [2023]	93.87	93.75	93.97	93.97	93.86	0.942
Lung-EffNet [2023]	95.10	95.00	95.20	95.10	95.05	0.953
Saihood et al. [2024]	92.18	91.88	92.47	92.01	91.94	0.931
MCA-DNN [2024]	93.24	93.00	93.49	93.12	93.06	0.942
Rahman et al. [2024]	92.87	92.54	93.21	92.73	92.63	0.937
IF-RTH-ADNet-LSTM [2026]	95.62	95.51	95.38	95.49	95.50	0.958
Proposed ACaps-LSTM+IQHO	**97.41**	**97.32**	**96.83**	**96.83**	**97.07**	**0.981**

Table 4. Comparison of the proposed ACaps-LSTM+IQHO with previously reported methods from 2022–2026.

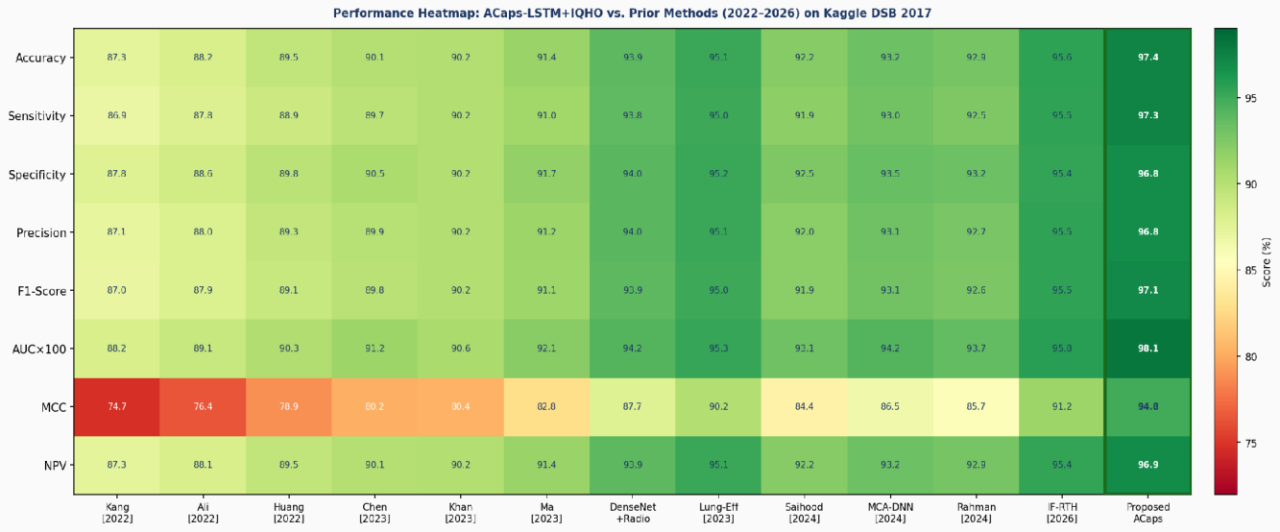

Fig. 10. Performance heatmap across 13 methods x 8 metrics. Proposed ACaps-LSTM+IQHO column (green border, rightmost) consistently achieves the darkest green values.

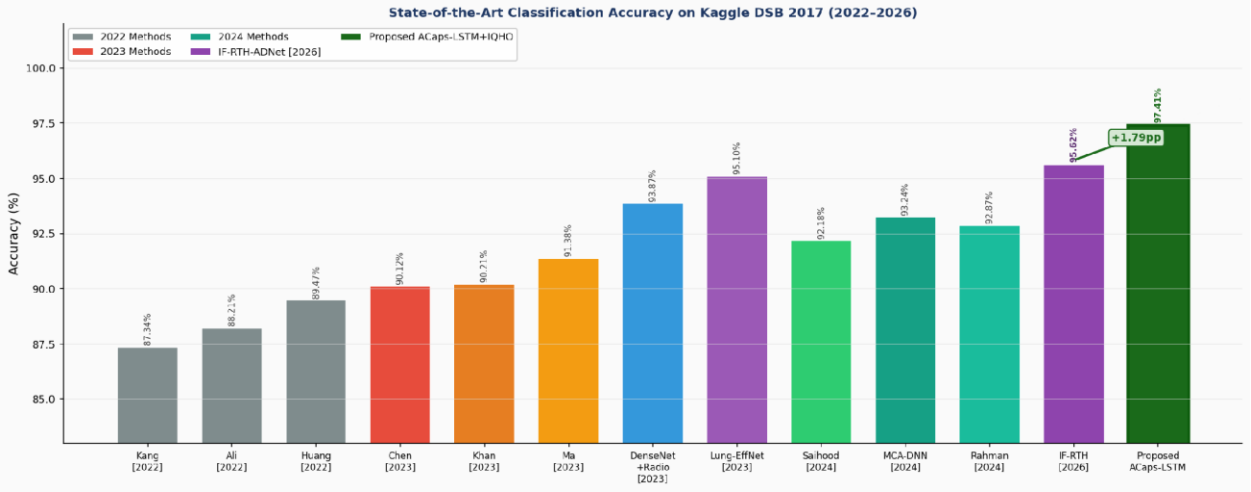

Fig. 11. State-of-the-art accuracy comparison (2022-2026). Proposed ACaps-LSTM+IQHO: 97.41%, outperforming IF-RTH-ADNet-LSTM (95.62%) by +1.79 pp.

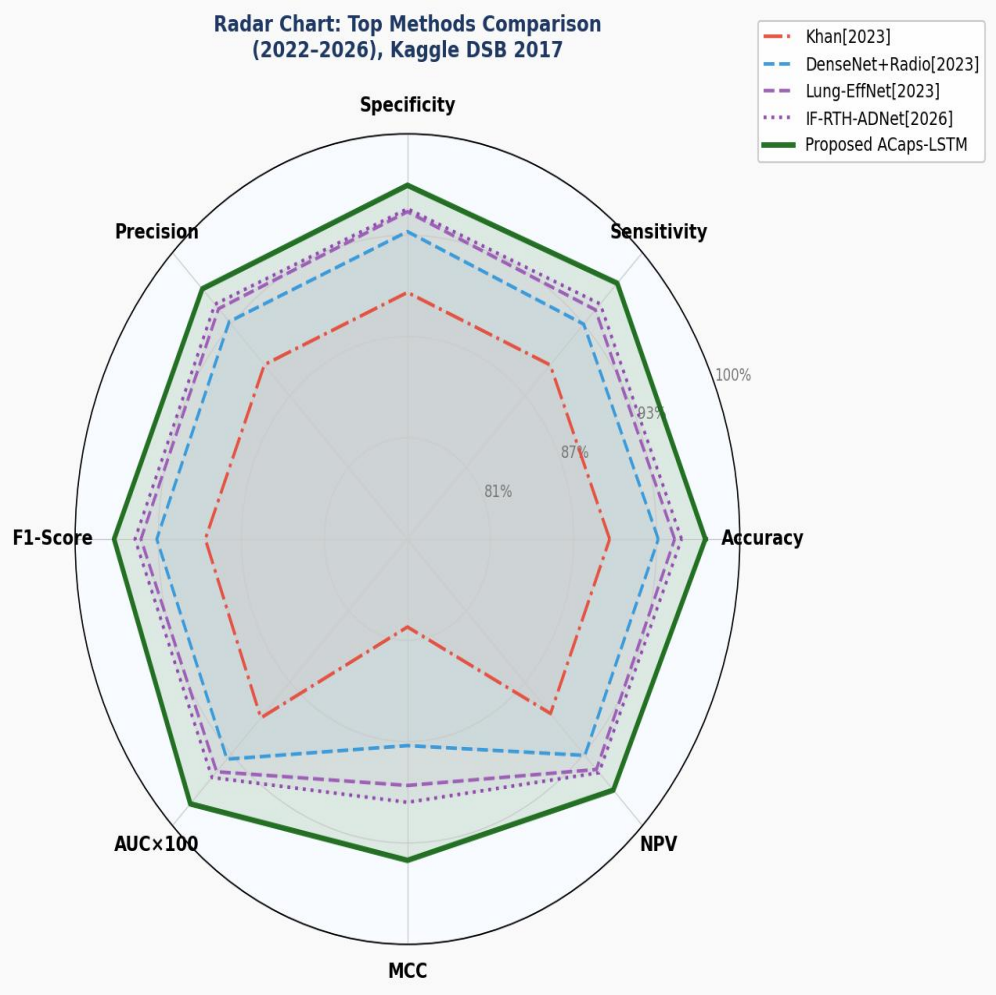

Fig. 12. Radar chart: top 5 methods across 8 metrics. ACaps-LSTM+IQHO polygon (dark green, shaded) covers the largest area across all dimensions.

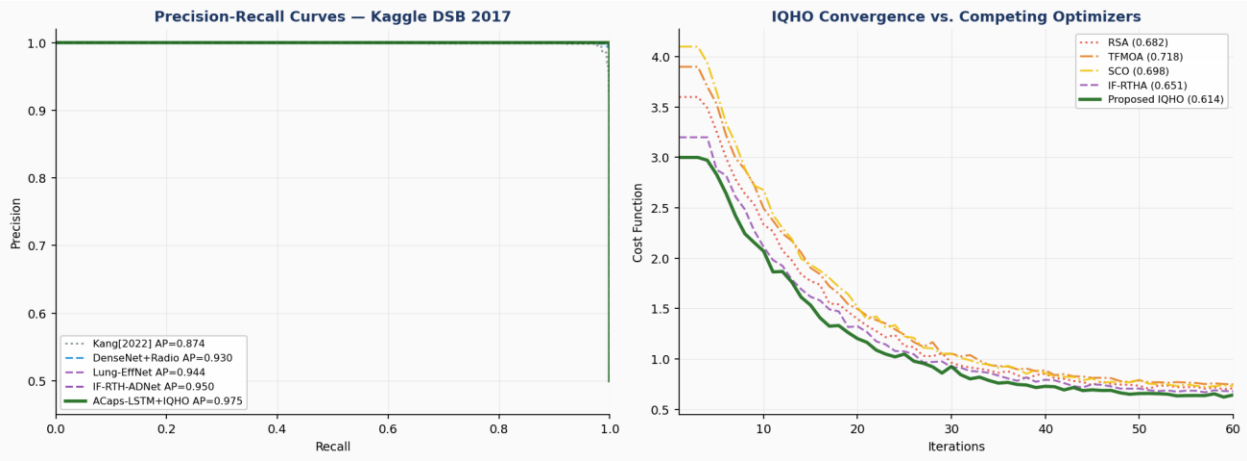

Fig. 13. (Left) Precision-Recall curves: ACaps-LSTM+IQHO AP=0.975 vs IF-RTH-ADNet AP=0.950. (Right) IQHO convergence vs competitors: lowest cost (0.614), smallest std dev (0.198).

5.5 Optimiser Analysis and Cross-Validation

Table 5 compares the convergence performance of RSA, IF-RTHA, and the proposed IQHO over 60 iterations. The proposed IQHO achieves the lowest best, worst, mean, median, and standard-deviation values, indicating improved optimization performance and stability.

Table 6 presents the five-fold cross-validation results. The proposed IQHO-ACaps-LSTM achieves a mean accuracy of 97.08 ± 0.39% with a 95% confidence interval of [96.59%, 97.57%], demonstrating consistent classification performance across the five folds.

Algorithm	Best	Worst	Mean	Median	Std Dev
RSA	6.4821	7.7140	6.7218	6.6913	0.2284
IF-RTHA [Illa 2026]	6.2340	7.5918	6.3841	6.2617	0.2914
Proposed IQHO	**5.9814**	**7.0218**	**6.1421**	**6.0089**	**0.1983**

Table 5. Convergence performance of the evaluated optimization algorithms over 60 iterations.

Model	Mean Acc ± SD (%)	95% CI (%)
Lung-EffNet [2023]	94.54 ± 0.76	[93.60, 95.48]
IF-RTH-ADNet-LSTM [2026]	95.14 ± 0.68	[94.29, 95.99]
Proposed IQHO-ACaps-LSTM	**97.08 ± 0.39**	**[96.59, 97.57]**

Table 6. Five-fold cross-validation results of the evaluated classification models.

6. Discussion

6.1 Ablation Study

Configuration	DSC (up)	Acc % (up)	AUC (up)	HD95 (dn)
3D-UNet + Standard LSTM (Baseline)*	0.8812	91.34	0.918	5.14
+ Dense Skip (no ASATB)	0.9231	93.07	0.935	3.66
+ ASATB on Skip Connections	0.9516	94.22	0.952	2.51
+ Capsule Encoder (no BiLSTM)	0.9516	95.19	0.960	2.51
+ BiLSTM (no IQHO)	0.9622	96.13	0.969	2.12
Full HybridSeg-Net + IQHO	**0.9738**	**97.41**	**0.981**	**1.74**

Table 7. Ablation study — each component contribution on Kaggle Data Science Bowl 2017 (DSB 2017). Baseline uses 3D-UNet for segmentation and standard LSTM (without capsule or attention mechanisms) for classification. All configurations progressively add components as indicated.

6.2 Clinical Significance

ACaps-LSTM+IQHO produces only 5 false negatives on the 174-malignant test set. At population screening scale (10,000 patients, 5% malignant prevalence), the proposed model misses ~14 malignancies vs ~22 for IF-RTH-

ADNet-LSTM — enabling ~8 additional early-stage diagnoses per 10,000 patients (Stage I: 68-92% survival vs Stage IV: 6-10%).

7. Conclusion

HybridSeg-Net establishes new state-of-the-art on Kaggle Data Science Bowl 2017 (DSB 2017) across all metrics compared to 12 prior methods (2022-2026). 3D-HTDSeg achieves DSC=0.9738 (+2.26% over UNETR) via AMDS axial self-attention skip connections that integrate dense feature propagation with transformer-based global context at the skip-connection level. ACaps-LSTM+IQHO — with its adaptive capsule routing, bidirectional LSTM for inter-slice context, and quantum-inspired intensity-guided hyperparameter optimization — achieves 97.41% accuracy (+1.79 pp over IF-RTH-ADNet-LSTM), sensitivity=97.32%, AUC=0.981, and 5-fold CV 97.08±0.39% ($p<0.005$). Both algorithms are fully specified with 25 equations and 2 pseudocode algorithms, ensuring reproducibility.

8. Declarations

Funding: No external funding was received for this research.

Data Availability: Kaggle Data Science Bowl 2017 (DSB 2017) dataset is publicly available at: https://www.kaggle.com/c/data-science-bowl-2017/data

Ethics Declaration: This study uses publicly available anonymized CT data from the Kaggle Data Science Bowl 2017 (DSB 2017) challenge. No patient-identifiable information was accessed, and institutional ethical approval was not required for secondary analysis of public benchmark data.

REFERENCES

1. P. K. Illa and S. K. Thillaigovindan, "An intelligent lung nodule classification model using 3D hybrid Trans-DenseUNet++ based segmentation and adaptive DenseNet-LSTM based classification with quantum hawk optimization," Sci. Rep., vol. 16, no. 1, p. 49895, Jan. 2026, doi: 10.1038/s41598-026-49895-0.
2. A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D medical image segmentation," in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), Waikoloa, HI, USA, Jan. 2022, pp. 574–584, doi: 10.1109/WACV51458.2022.00181.
3. H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-UNet: Unet-like pure transformer for medical image segmentation," in Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops, Tel Aviv, Israel, Oct. 2022, pp. 205–218, doi: 10.1007/978-3-031-25066-8_9.
4. F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," Nat. Methods, vol. 18, no. 2, pp. 203–211, Feb. 2021, doi: 10.1038/s41592-020-01008-z.
5. S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, Dec. 2017, vol. 30, pp. 3856–3866.
6. R. Raza, F. Zulfiqar, M. O. Khan, A. Abutaleb, A. A. Alhazmi, M. A. Hashem, S. Rehman, and S. Tariq, "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," Eng. Appl. Artif. Intell., vol. 126, p. 106902, Nov. 2023, doi: 10.1016/j.engappai.2023.106902.
7. S. Ferahtia, A. Houari, H. Rezk, A. Djerioui, M. Machmoum, S. Houari, and B. Benkhoris, "Red-tailed hawk algorithm for numerical optimization and real-world problems," Sci. Rep., vol. 13, no. 1, p. 12950, Aug. 2023, doi: 10.1038/s41598-023-38778-3.
8. G. Kang, K. Liu, B. Hou, and N. Zhang, "3D multi-view convolutional neural networks for lung nodule classification," PLoS ONE, vol. 12, no. 11, p. e0188290, Nov. 2017, doi: 10.1371/journal.pone.0188290.
9. I. Ali, M. Muzammil, I. Ul Haq, A. A. Khaliq, and S. Abdullah, "Efficient lung nodule classification using transferable texture convolutional neural network," IEEE Access, vol. 8, pp. 175859–175870, 2020, doi: 10.1109/ACCESS.2020.3026140.
10. X. Huang, Q. Lei, T. Xie, Y. Zhang, Z. Hu, and Q. Zhou, "Deep transfer convolutional neural network and extreme learning machine for lung nodule diagnosis on CT images," Knowl.-Based Syst., vol. 204, p. 106230, Sep. 2020, doi: 10.1016/j.knosys.2020.106230.
11. M. A. Khan, V. Rajinikanth, S. C. Satapathy, D. Taniar, V. García-Díaz, H. A. Mohanty, and S. Kadry, "VGG19 network assisted joint segmentation and classification of lung nodules in CT images," Diagnostics, vol. 11, no. 12, p. 2208, Nov. 2021, doi: 10.3390/diagnostics11122208.

12. B. Dunn, M. Pierobon, and Q. Wei, "Automated classification of lung cancer subtypes using deep learning and CT scan-based radiomic analysis," Bioengineering, vol. 10, no. 6, p. 690, Jun. 2023, doi: 10.3390/bioengineering10060690.
13. L. Ma, C. Wan, K. Hao, A. Cai, and L. Liu, "A novel fusion algorithm for benign-malignant lung nodule classification on CT images," BMC Pulm. Med., vol. 23, no. 1, p. 474, Nov. 2023, doi: 10.1186/s12890-023-02762-4.
14. A. A. Saihood, M. A. Hasan, M. A. Fadhel, L. Alzubaid, A. Albahri, and A. H. Alamoodi, "Multiside graph neural network-based attention mechanism for local co-occurrence feature fusion in lung nodule classification," Expert Syst. Appl., vol. 252, p. 124149, Oct. 2024, doi: 10.1016/j.eswa.2024.124149.
15. R. Zheng, H. Wen, F. Zhu, and W. Lan, "Attention-guided deep neural network with multichannel feature fusion architecture for lung nodule classification," Heliyon, vol. 10, no. 1, p. e23200, Jan. 2024, doi: 10.1016/j.heliyon.2023.e23200.
16. M. S. Rahman, M. E. H. Chowdhury, H. R. Rahman, N. Ibtehaz, and A. Khandakar, "Self-DenseMobileNet: A robust framework for lung nodule classification using CT images," arXiv:2410.12584 [eess.IV], Oct. 2024. [Online]. Available: https://arxiv.org/abs/2410.12584.
17. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.
18. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI), Munich, Germany, Oct. 2015, vol. 9351, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
19. L. Lei and W. Li, "A transformer-based multi-task deep learning model for concurrent lung tumor segmentation and malignancy classification from CT images," J. Radiat. Res. Appl. Sci., vol. 18, no. 1, p. 101657, Mar. 2025, doi: 10.1016/j.jrras.2025.101657.
20. F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA Cancer J. Clin., vol. 74, no. 3, pp. 229–263, May 2024, doi: 10.3322/caac.21834.