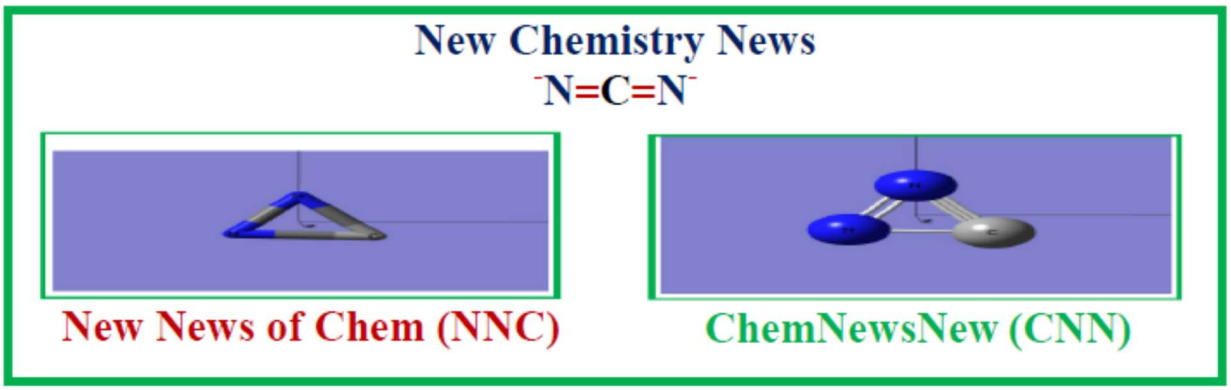
Available online at **www.joac.info**

ISSN: 2278-1862

Journal of Applicable Chemistry

2025, 14 (2): 548-591

(International Peer Reviewed Journal)



CNN – 66E

Iam (Intelligence Augmented /Assisted Method(s))

Transformers --architectures & Fits (2025)

Information Source	sciencedirect.com;	
S. Narasinga Rao M D Associate Professor, Emergency Medicine dept., Andhra Medical College, King George Hospital Visakhapatnam, A.P., India	K. Somasekhara Rao, Ph D Dept. of Chemistry, Acharya Nagarjuna Univ., Dr. M.R.Appa Rao Campus, Nuzvid-521 201, India	R. Sambasiva Rao, Ph D Dept. of Chemistry, Andhra University, Visakhapatnam 530 003, India
snrnaveen007@gmail.com (+91 98 48 13 67 04)	sr_kaza1947@yahoo.com (+91 98 48 94 26 18)	rsr.chem@gmail.com (+91 99 85 86 01 82)

Conspectus: The paper entitled "Attention is All You Need" by Ashish Vaswani et al. in the year 2017, brought renaissance in Text-sequence-data processing. Also, it has noteworthy influence in computational paradigm with other data structures. The new approach became a state-of-knowledge model and won the favour of data scientists in all application-domains. This model became popular as

Transformer net (TransF Net) or Transformer neural network (TransF NN). This network, TransF NN, consists of two modules, viz., attention layer and MLP-NN. They are instrumental in carrying out Natural Language processing (NLP). The evolution of architecture of TransF NN, attention mechanism, and hybridization with other approaches, during these few years, revolutionized computational science. By now, this approach is in the sought-after tools in extracting information/knowldege with multi-modal data (viz.Text, numerical time-series, sound (speech), image/video sequence, and tactile-sense-output) with local and global inter-dependencies.

The architectures of Transformer neural nets (TransF-NN) or Transformer nets (TransF-N) employed in this state-of-knowledge-methods-module for dataTOknowledge transformation are

- Sparse transformer (ResidualTop-C_sparse attention)
- Pancreas segmentation (PanSeg) Transformer
- Co-evolution Transformer
- EEG Conformer
- Swin Transformer
- HRSTNet: High-Resolution Swin Transformer Network
- HQRSTNet: High-Quality Resolution Swin Transformer
- EfficientNet and Swin Transformer
- Squeeze and Excitation-based UNet TRansformers (SE-UNETR)
- Squeeze and Excitation-based High-Quality Resolution Swin Transformer Network (SE-HQR-STNet)
- TCN
- RS-MOCO
- RBMDC-Net

The results of modelling of tasks (vide infra) with transformer NNs are more accurate in comparison with NNs, machine learning algorithms or discipline-wise theorical approaches.

- Computational Quantum Chemistry (CQC): To predict quantum chemical energies and physical-/chemical-/physico-chemical/chemico-physical energies/properties
  - Total Molecular energy, orbitals-energies, HOMO-LUMO energy gap, dipole moment, electron density, ESP, bond energies, electronic spectra, NMR, Reaction Pathways, Reaction Mechanisms, Transition State

- Quantum Monte Carlo (QMC) Simulations
  - Approximation of Ground-State Energy in Many-Body Systems

- Chemistry
  - Structure-activity relationship (SXR)
  - Biological activity (e.g., drug efficacy, toxicity, or binding affinity) in drug design, materials science, and toxicology
  - Molecular Property Prediction:
    - Boiling point, toxicity, or binding affinity
    - Solubility of Organic Compounds

        - Nonlinear Solubility is due to nonlinear interactions like hydrogen bonding, van der Waals forces, and entropy changes upon dissolution,
            - SFHformer captures these characteristics through its hybrid spatial and frequency domain approach
    - Predicting Reaction Rates in a Multi-Step Reaction Network
    - Adsorption Isotherms
        - Predicting Adsorption on a Surface (Langmuir-Freundlich Isotherms)
    - Predicting Toxicity of Chemical Compounds
            - for Safety Assessment
    - Nonlinear Thermodynamic Equilibria: Chemical Equilibrium in a Multi-Component System:
        - At equilibrium, the functional relation between the concentrations of reactants and products is nonlinear, more so when there are competing reactions or there are changes in phase.
            - SFHformer accounts for both local interactions (e.g., bond strengths, charge distributions) and global properties (e.g., temperature, pressure effects). That is why, it predicts equilibrium concentrations more accurately, outperforming theoretical and other computing models
    - Predicting Reaction Kinetics (Rate Constants)
    - Predicting Molecular Properties for Chemical Engineering
    - Predicting Thermodynamic properties (e.g., heat capacity, entropy, enthalpy) and kinetic properties (e.g., reaction rates, activation energies)
    - Predicting Boiling points, melting points

- Nanomaterials (Physics, chemistry Biology)
    - Nanomaterials exhibit unique physical properties (like superconductivity, magnetic behavior) that are difficult to predict due to their complexity and the high number of influencing variables (like size, shape, surface properties) involved.
    - Nanomaterials exhibit responses like optical absorption, band gaps, and thermal conductivity
    - (Smart)-Nanomaterial Discovery: use of Nanocatalysts in industrial process viz. hydrogen production, carbon capture, and chemical synthesis is the need of the hour.
    - New nano-materials with desirable properties (e.g., photovoltaic devices, superconductors)
    - Rational design of nanomaterials in specific applications, such as catalysis, electronics, and energy storage
    - Nano-Scale materials for Optical Sensing and Imaging
    - Predicting Nanoparticle-Polymer Interactions in Drug Delivery Systems

- Fusion Reactors (e.g., tokamaks)
    - Predicting Reactor Decommissioning and Safety
    - Modeling Plasma Behavior in Fusion Research
    - Design of fusion reactors and energy optimization in nuclear fusion research, contributing to the goal of clean, sustainable energy

- Fusion reactors require materials that can withstand extreme conditions such as high temperatures, radiation, and corrosive environments

- Predicting Optical Properties of Molecules Using Excited States
  - Predicting the optical properties of molecules, such as absorption spectra and fluorescence. This is essential for the design of optical materials and pharmaceuticals
  - Design of materials for light-emitting devices and photodynamic therapy.
  - Molecular sensors and optical materials of advanced materials for solar cells, LEDs, and bioimaging, contributing to innovations in photonics and quantum materials
- Environmental monitoring
  - Chemical Sensor Data Analysis: The time-series sensor data from gas sensors (e.g., detecting gases like CO_2, NO_2, or O_3) over time and their corresponding chemical concentrations

Medical Diagnosis

- Medical Diagnosis using Images
  - Lung Cancer Detection (CT Scan Image) and Classification
  - Tumor Segmentation (MRI Brain Scan)
  - Medical Diagnosis of Diabetic Retinopathy Using Eye Imaging Data
  - Early Diagnosis of Alzheimer's Disease Using MRI Brain Imaging
  - Brain Tumor Detection Using MRI and PET Scans
- Diagnosis using Time-Series Data
  - Sepsis Prediction (Time-Series Data)
  - Heart Disease Prediction (ECG Time-Series Data)
- Clinical Decision Support Systems (CDSS)
  - Multi-Modal Diagnosis (Image + Time-Series Data)

Bio-Medical research

- AlphaFold, a model based on Transformer architecture
  - To predict 3D-protein structures from amino acid sequences
- Activity of bio-molecules
  - Predicting
    - Functional activity of an enzyme
    - Binding Affinity for Kinase Inhibitors
    - Anticancer Activity (e.g., Inhibition of Kinase Activity)
    - Inhibition of HIV-1 Protease
    - Activity of Enzyme Inhibitors
    - Binding Affinity for HIV Protease Inhibitors
    - Drug-Protein Interaction in Alzheimer's Disease
    - Drug Resistance in Cancer Therapy (Optimizing Anti-Cancer Drug Binding Using Attention Maps)
    - ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) Properties
    - Toxicity levels (e.g., LD50 values, mutagenicity, or carcinogenicity) for Chemical Compounds (chemical structures represented by SMILES or molecular graphs)
    - Protein-Ligand Binding Affinity for Drug Discovery

- Protein-Protein Interactions in Disease Pathways
  - Personalized Medicine for Cancer Treatment Using Genomic Data
  - Drug Repurposing for Rare Diseases
- Detecting Mutations in Genes Associated with Brain Diseases

Keywords: Artificial intelligence (AI); Heuristic expert systems—Integrated expert systems--Classical Neural Nets (MLP; SOM; ARTMAP) -- Capsule Neural Nets — Attention_Mechanism--TransFormer_Nets — Hybrid_TransFormer_Networks -- Artificial General intelligence (AGI); Vitual reality (VR) – Meta Verse (MV)--

CNN : [C [Computations; Computer; Chemistry, Cell, Cellestial, Cerebrum] NN [New News; News New; Neural Nets; Nature News; News of Nature;]]
Fits : [Figure Image Table Script;]

K(nowledge)Lab
rsr.chem1979

Transformer Net	**2025-13**	**Cancer**

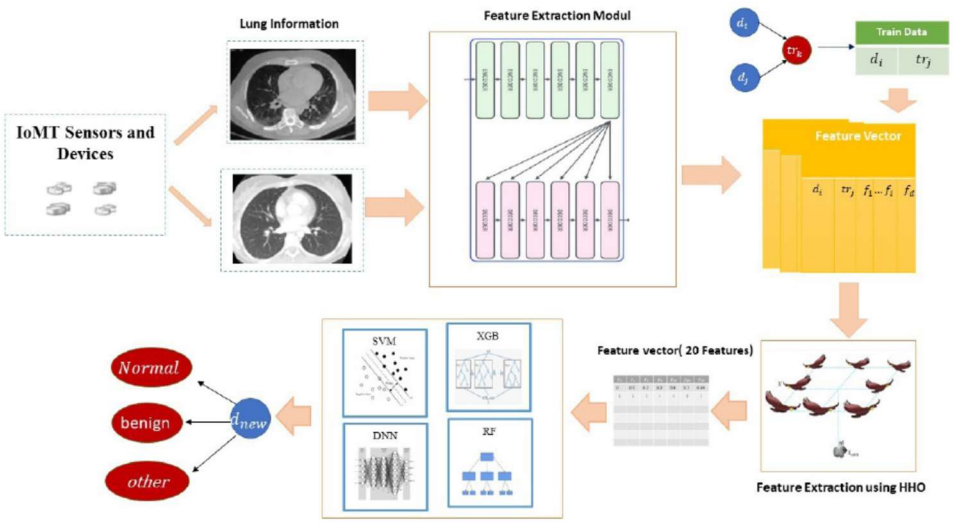


Fig. 1. Overview of the proposed method's stages.

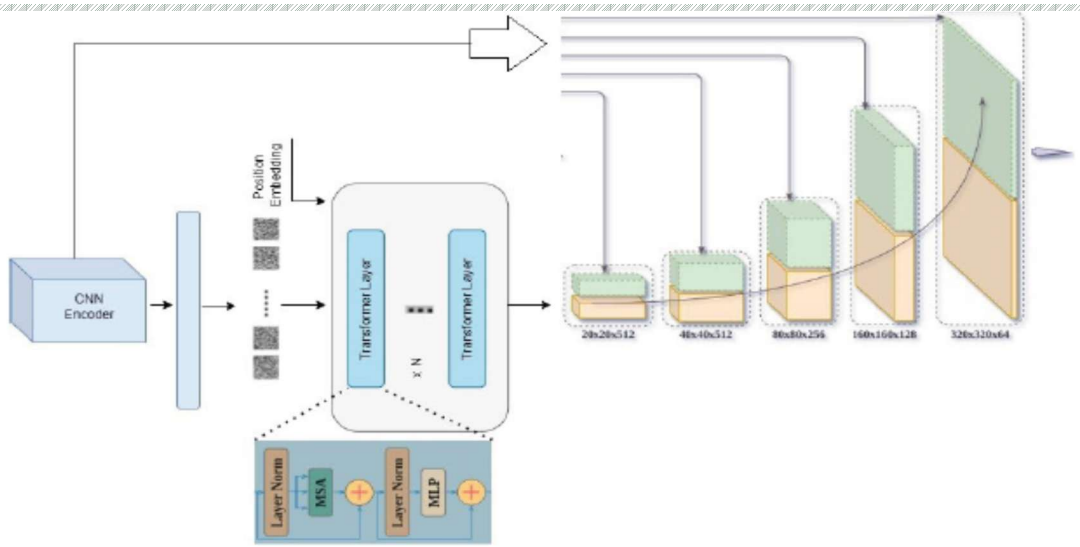


Fig. 2. Details of the Transformer network used for feature extraction.

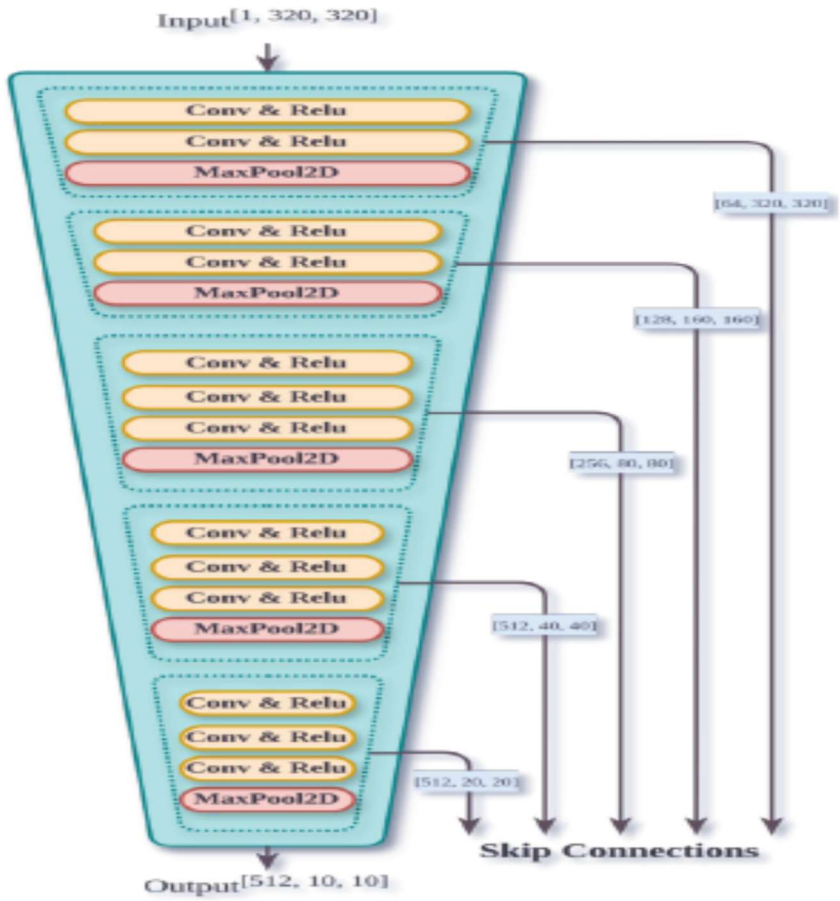


Fig. 3. Convolutional network architecture used in feature extraction phase.

```
Input: Dataset
Output: One sample with the best features
////////// filter phase//////////
-Weigh features of dataset according to F-Score algorithm
-Keep high weights according to dimensions of dataset
////////// generate generation//////////
-Generate the first generation among selected features randomly from the previous step
-for (each hawk (Xi)) do
    -calculate the fitness of sample
    -While (fitness<0.65) do
-select one random feature and replace it
-calculate the fitness of sample
////////// hho phase//////////
-While (stopping condition is not met) do
    -for s=2 to number_of_samples do
        Calculate the fitness values of hawks
        Set Xrabbit as the location of rabbit (best location)
        for (each hawk (Xi)) do
            Update the initial energy E0 and jump strength J ▷
            E0=2rand()-1, J=2(1-rand())
            Update the E
            if (|E|≥ 1) then ▷ Exploration phase
                Update the location vector
            if (|E|< 1) then ▷ Exploitation phase
                if (r ≥0.5 and |E|≥ 0.5) then ▷ Soft besiege
                    Update the location vector
                else if (r ≥0.5 and |E|< 0.5) then ▷ Hard besiege
                    Update the location vector
                else if (r < 0.5 and |E|≥ 0.5) then ▷ Soft besiege with progressive rapid dives
                    Update the location vector
                else if (r < 0.5 and |E|< 0.5) then ▷ Hard besiege with progressive rapid dives
                    Update the location vector
```

Fig. 5. Pseudocode of the proposed feature selection algorithm.

Table 1. Details of the dataset used

Class	Database	Train	Test	Total
Pneumonia	COVIDx-CT	3419	854	4273
Tuberculosis	PTB	401	100	501
Covid-19	SARS-CoV-2 CT	1986	496	2482
Lung cancer	CIA	4355	1088	5043
Normal	-	4302	1292	5594

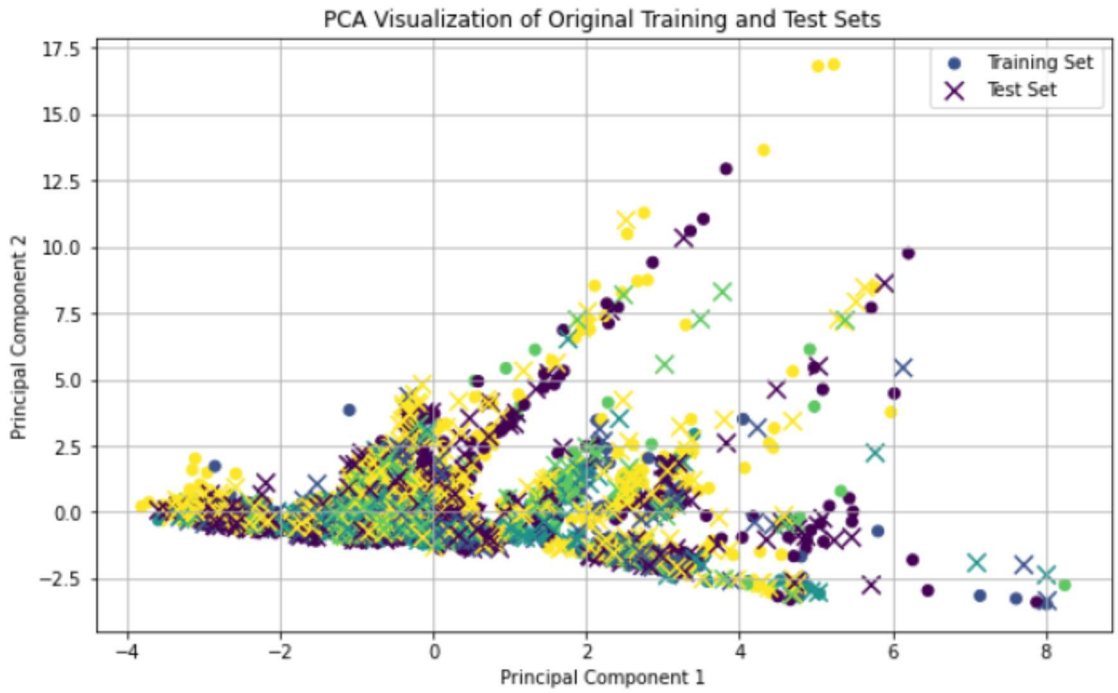


Fig. 6. The distribution of training and testing data after dimensionality reduction using PCA.

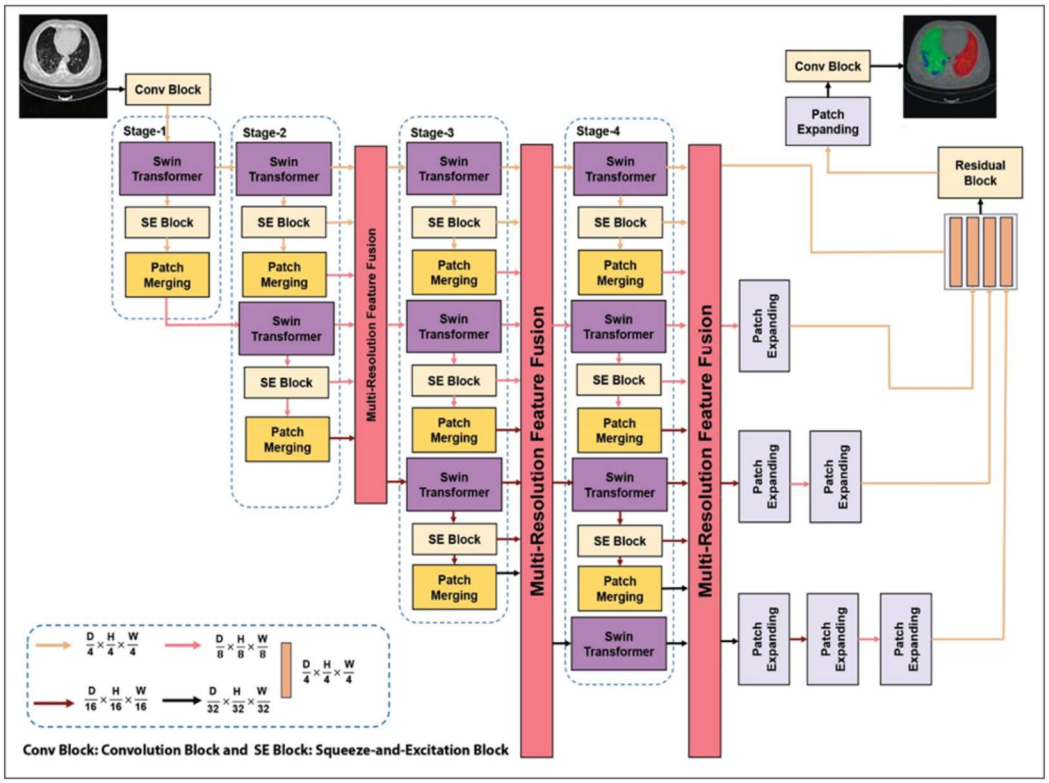
Table 2. Evaluation of different classifiers into the dataset with and without feature selection

Dataset	Classifier	Without				With			
		Acc (%)	Sp (%)	Sn (%)	F1-score	Acc (%)	Sp (%)	Sn (%)	F1-score
Pneumonia	SVM	73	68	86	87	97	95	97	97
	RF	91	93	93	93	98	96	98	98
	XGBoost	93	94	92	93	98	97	98	98
	DNN	87	89	89	89	96	95	97	96
Tuberculosis	SVM	83	85	84	85	97	95	98	96
	RF	89	91	92	92	97	96	98	97
	XGBoost	89	92	91	92	98	96	99	98
	DNN	84	86	86	86	96	95	97	96
Covid 19	SVM	59	58	61	63	96	95	98	98
	RF	90	93	93	94	97	97	99	99
	XGBoost	94	94	93	93	98	98	99	99
	DNN	69	82	87	87	96	95	98	98
Lung cancer	SVM	66	68	69	69	98	97	98	97
	RF	90	92	91	91	99	98	99	97
	XGBoost	93	94	97	97	99	98	99	98
	DNN	78	69	88	88	98	97	98	97
Normal	SVM	65	64	66	66	97	96	98	98
	RF	91	93	91	91	98	97	99	98
	XGBoost	92	93	95	95	98	99	99	99
	DNN	72	58	86	86	97	96	98	97

Table 3. Comparison with other methods on the dataset

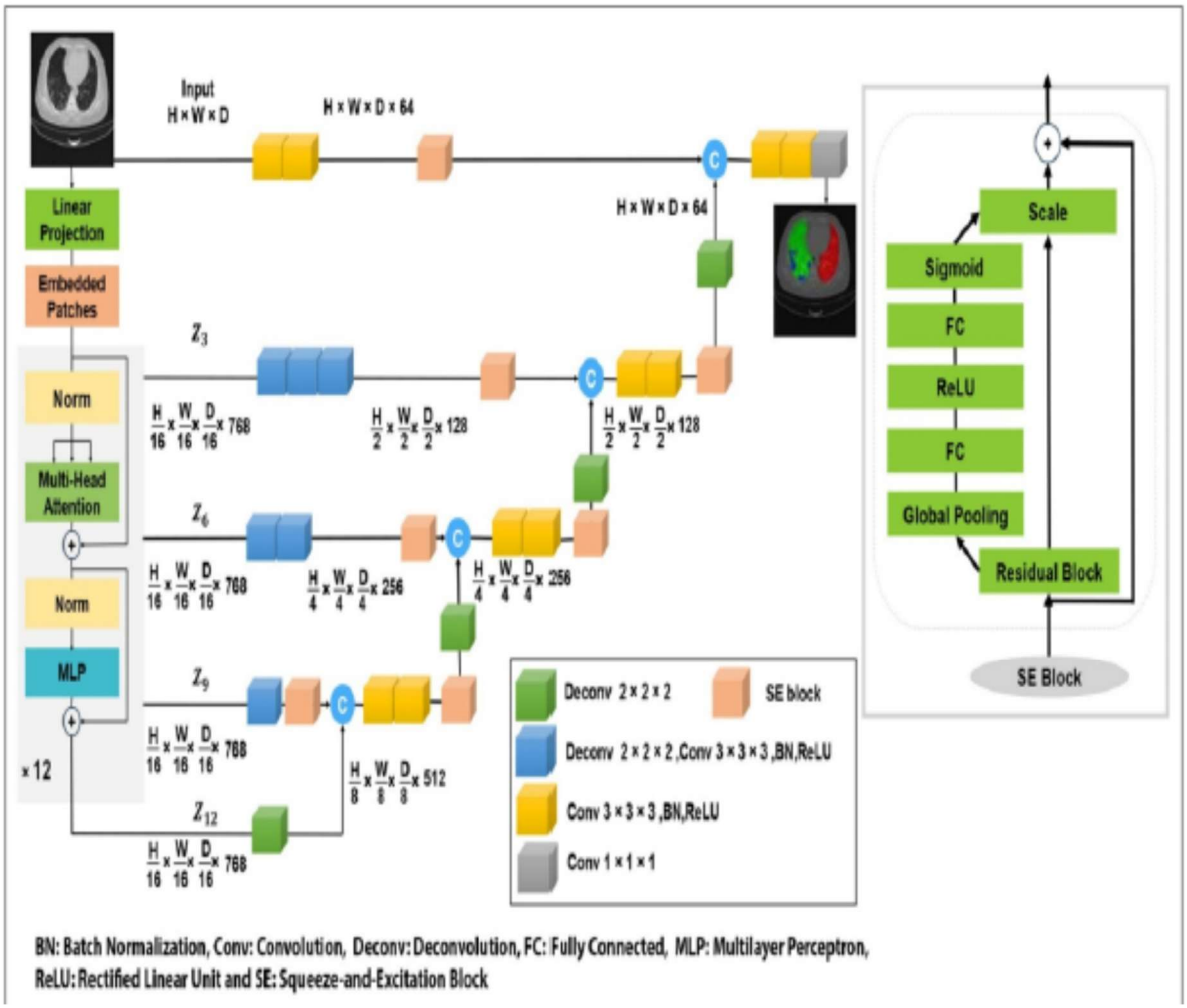
Method	Class	Accuracy (%)
Deep learning [20]	2	89.5
DCNN [21]	2	93.64
CNN [22]	2	98
AE-CNN [23]	2	80.29
VDSNet [24]	2	73
eKNN with ACO [25]	2	97.5
Ensemble learning [26]	2	98.56
3DDCNN [27]	2	98.51
Proposed method	5	98.53

Transformer Net	2025-17	Covid-19

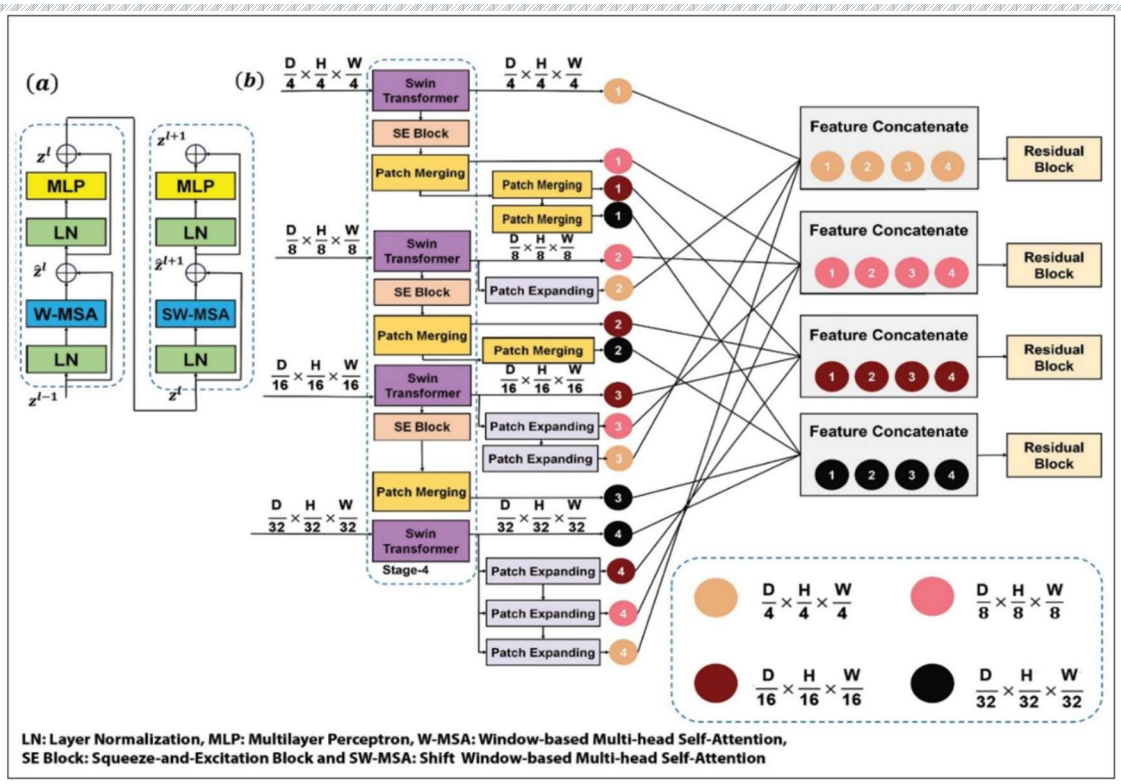


The structure of the Squeeze and Excitation-based High-Quality Resolution Swin Transformer Network (SE-HQRSTNet) architecture.

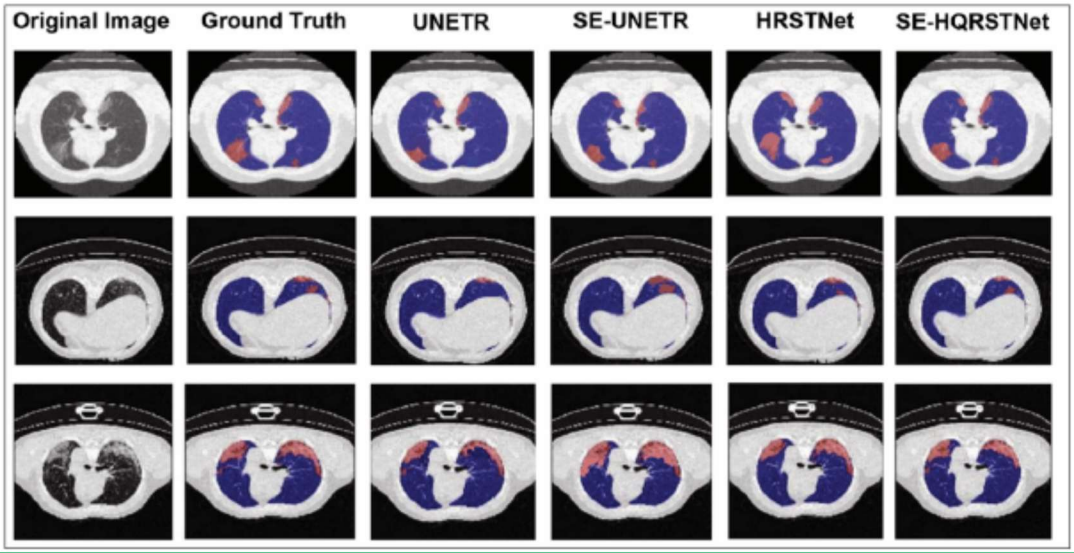
H:height;W: width; D: depth



Squeeze and Excitation-based UNet TRansformers (SE-UNETR) architecture. H:height;W: width; D: depth



(a) The Swin Transformer block, and (b) multi-resolution feature fusion (MRFF) block. H:height;W: width; D: depth



Visual comparison between the ground truth and prediction of the models segmentation for 3 computed tomography (CT) scan samples of COVID-19 patients. SE: Squeeze-and-Excitation, UNETR: UNEt TRansformers, HRSTNet: High-Resolution Swin Transformer Network, HQRSTNet: High-Quality Resolution Swin Transformer Network

Table 4: Comparison of results of our models with previous studies

Author	Dataset	Splitting Type	Method: Dice
Müller et al. (3)	COVID-19-CT-Seg	5-Fold	3D U-Net:0.761
Ma et al. (4)	COVID-19-CT-Seg	5-Fold	nnU-Net:0.673
Wang et al. (5)	COVID-19-CT-Seg	5-Fold	3D U-Net: 0.704
Singh et al. (25)	COVID-19-CT-Seg	Train:70% Validation:10% Test:20%	LungINFseg:0.8034
Aswathy et al. (7)	COVID-19-CT-Seg	Train:60% Validation:20% Test:20%	Cascaded 3D U-Net:0.820
Our method	COVID-19-CT-Seg	5-Fold	UNETR:0.8519 SE-UNETR:0.8581 HRSTNet: 0.8663 SE-HQRSTNet: 0.8684
Zheng et al. (26)	MosMed	5-Fold	3D CU-Net:0.668
Our method	MosMed	5-Fold	UNETR:0.6901 SE-UNETR:0.6935 HRSTNet: 0.7072 SE-HQRSTNet: 0.7089

SE: Squeeze-and-Excitation ,UNETR: UNEt TRansformers , HQRSTNet: High-Quality Resolution Swin Transformer Network, HRSTNet: High-Resolution Swin Transformer Network.

Transformer Net	2025-23	

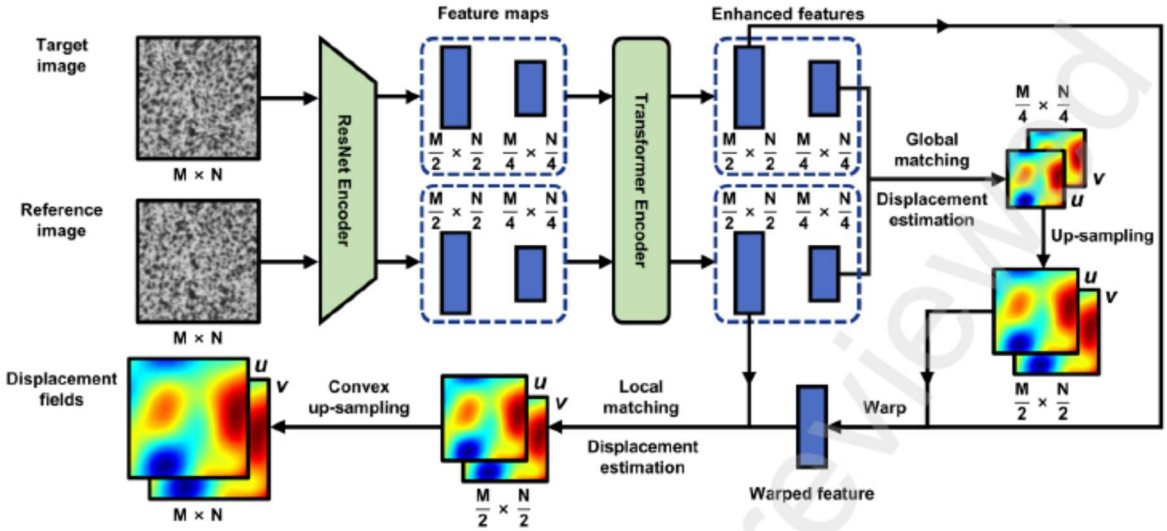


Figure 1. Architecture of DICTr.

Transformer Net	2025-34	

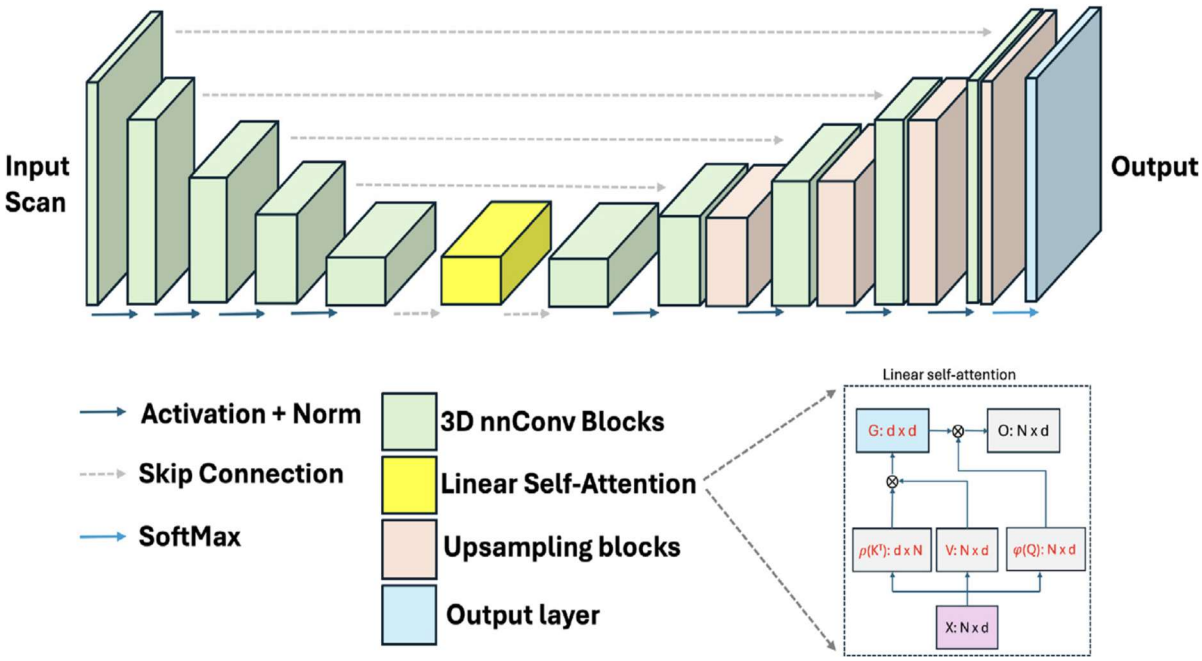
Table 1
Existing CT (light gray) and MRI-based (gray) pancreas segmentation methods

Methods	Approach	Dataset	Performance/Dice
Attention U-Net: Learning Where toLook for the Pancreas (Oktay et al.,2018)	Attention mechanisms (layers) areintegrated within the U-Net to focus on thepancreas region to avoid false positives.	NIH (8)	83.1 ± 3.8
Fully automated pancreassegmentation with two-stage 3Dconvolutional neural networks (Zhaoet al.,2019)	A two-stage 3D model is designed with thefirst stage for coarse pancreas segmentationand the second stage for refinedsegmentation.	NIH (8)	86.0 ± 4.5

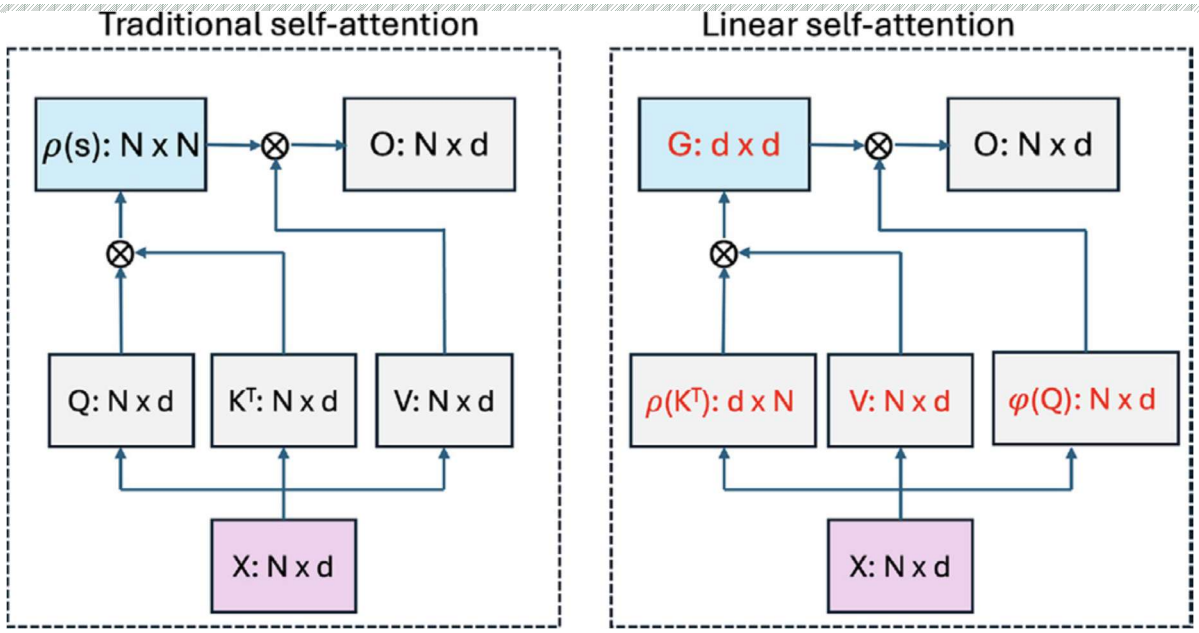
Automated pancreas segmentationand volumetry using deep neuralnetwork on computedtomography (Lim et al.,2022)	This paper performs four individualthree-dimensional pancreas segmentationnetworks on 1006 participants.	1006 in-house CTscans	84.2
Automated pancreas segmentationusing recurrent adversariallearning (Ning et al.,2018)	A recurrent adversarial learning frameworkis developed to enhance the pancreassegmentation robustness.	NIH (8)	88.72 ± 3.23
Deep Q-learning-driven CT pancreassegmentation with geometry-awareU-Net (Man et al.,2019)	A combination of deep Q-network andgeometry-aware U-Net introducereinforcement learning to improve thepancreas segmentation performance further.	NIH (8)	86.9 ± 4.9
Pancreas segmentation in MRI usinggraph-based decision fusion onconvolutional neural networks (Caiet al.,2016)	The paper conducts pancreatic detectionwith spatial intensity context and pancreassegmentation by graph-based decisionfusion.	78 in-house T1 MRIscans	76.1 ± 8.7
Hierarchical 3D Feature Learning forPancreasSegmentation (Proietto Salanitriet al.,2021)	A multiheaded decoder structure is designedto predict intermediate segmentation maps,and the final segmentation result comesfrom the aggregation of each levelprediction.	40 In-house T2scans	77.5 ± 8.6
Improving deep pancreassegmentation in CT and MRI imagesvia recurrent neural contextuallearning and direct loss function (Caiet al.,2017)	The paper proposes recurrent neuralcontextual learning and a direct lossfunction and involves training the networkto learn contextual information fromneighboring pixels in the image.	79 in-house T1 MRIscans	80.5 ± 6.7

2025-34

***PanSegNet* (volumetric pancreas segmentation network)**



PanSegNet is based on a combination of *nnUnet* with linear self-attention mechanism. Linear self-attention is obtained by converting the self-attention mechanism with linearization operation, as described below. The architecture accepts volumetric input, therefore appreciating the full anatomy details compared to pseudo-3D approaches

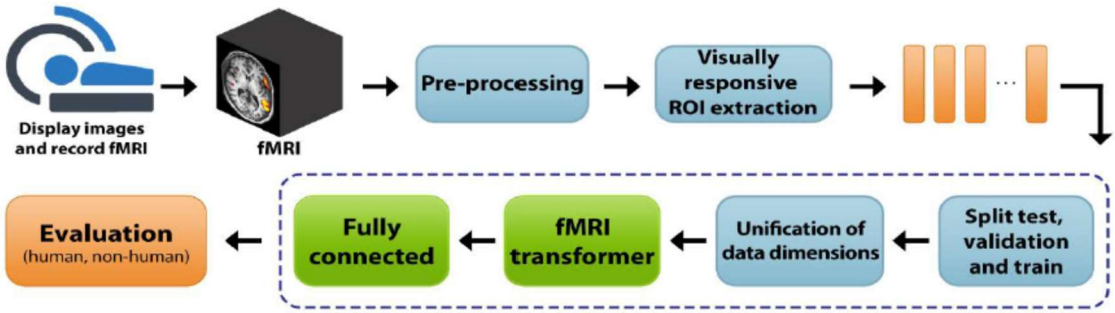


Comparison of traditional self-attention mechanism (left) vs. linear self-attention mechanism (right). X is input, O is output. Red fonts show the specific changes we apply to self-attention to linearize

Table 7
Quantitative segmentation performance evaluation and Comparison with diverse state-of-art methods across multi-center T1W and T2W MRIs.

Multi-center T1W MRI Pancreas segmentation							
Methods	Modality	Dice (%)	Jaccard (%)	Precision (%)	Recall (%)	HD 95 (mm)	ASSD (mm)
nnUNet	2D	80.19	80.01	80.19	81.01	20.99	1.75
SSformer	2D	78.81	77.01	76.67	75.89	23.09	2.54
SwinUNETR	2D	76.01	75.21	73.21	74.11	27.78	2.98
MedSegDiff	2D	83.75	82.11	81.78	80.99	18.97	1.56
SynergyNet	2D	85.78	84.37	84.09	84.44	17.88	0.95
VNet	3D	73.15	74.01	72.11	84.92	71.47	2.99
TransBTS	3D	75.89	74.18	74.87	73.92	26.44	3.01
MedNext	3D	80.05	79.99	83.33	80.02	17.77	1.67
nnFormer	3D	82.11	83.28	83.23	81.11	18.45	1.98
nnUNet	3D	80.09	81.29	83.87	81.98	18.12	1.79
nnUnet-Res	3D	83.02	84.01	82.91	82.54	17.92	1.52
PanSegNet	3D	**86.02**	**85.78**	**84.18**	**84.76**	**17.47**	**0.92**

Transformer Net	**2025-40**	



Block diagram illustrating the step-by-step process for constructing the developed model

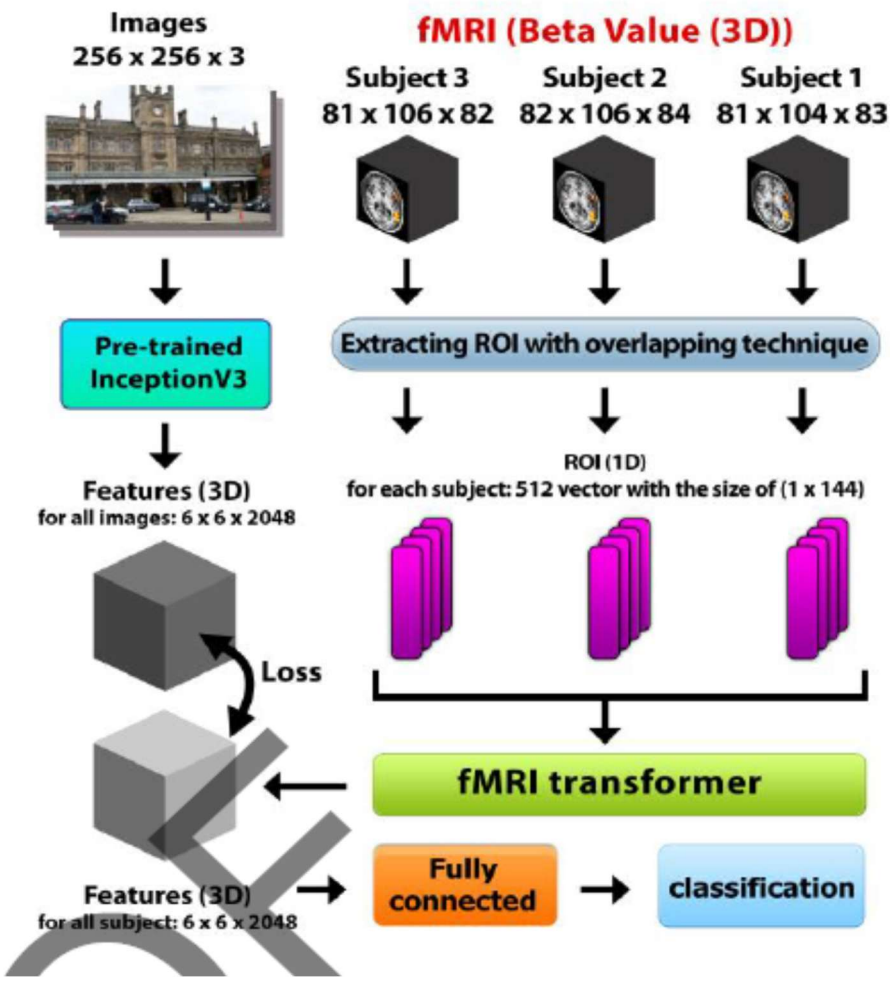
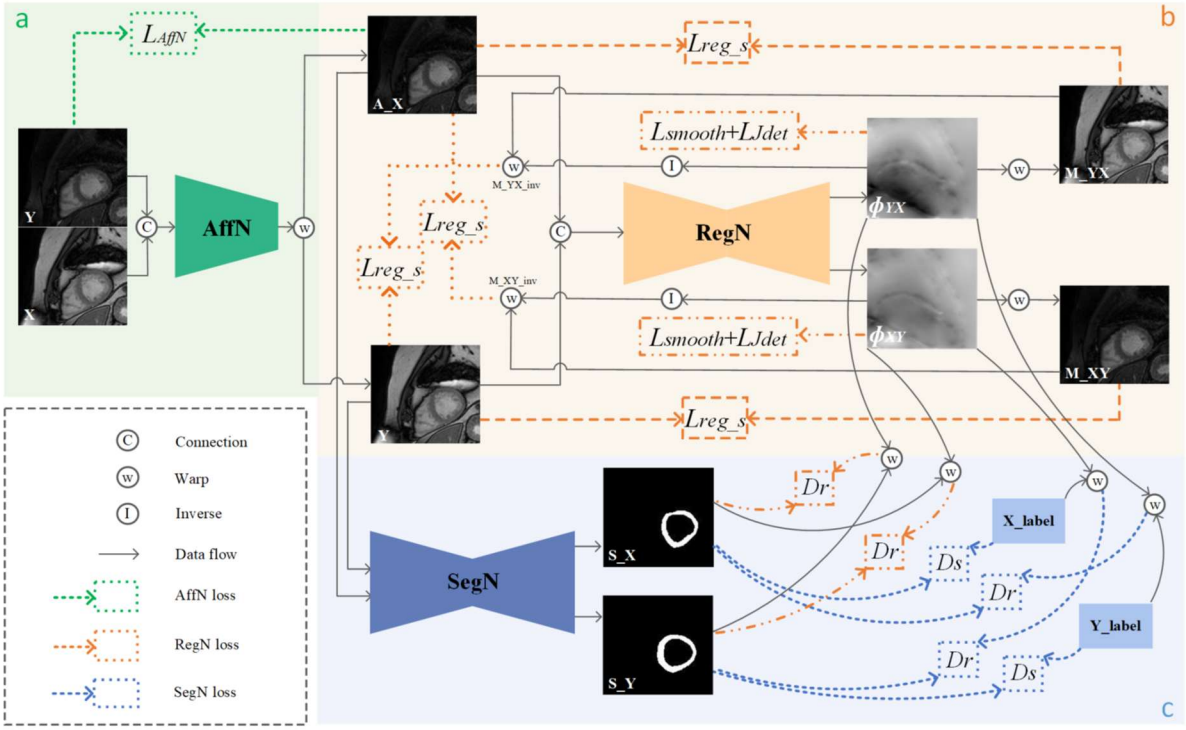


Figure 4. Model architecture; pre-trained image encoder extracts features from images. The transformer network extracts features from fMRI signals and tries to construct an fMRI space similar to visual space. Then object categorization is done using the FC network

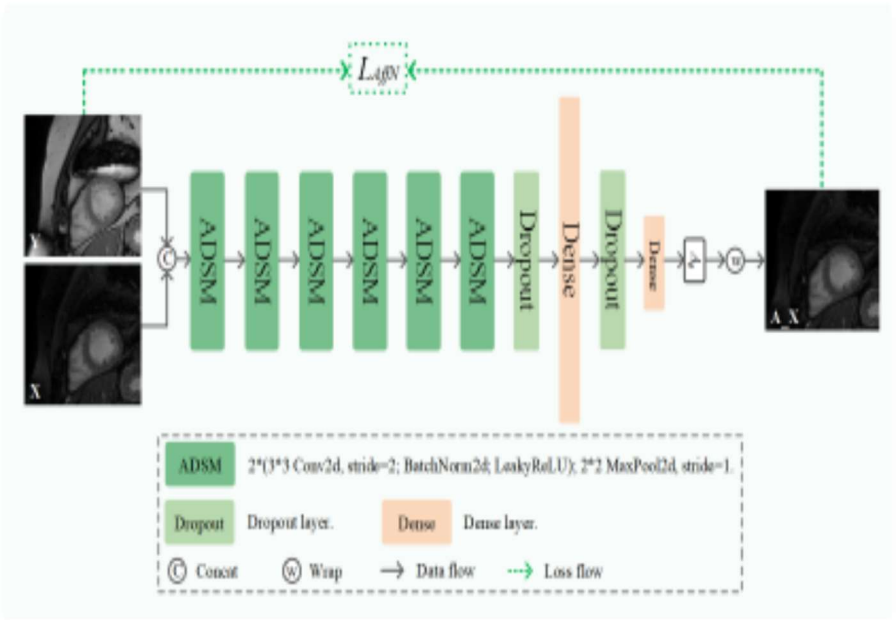
True category: human
Predicted category: human

True category: human
Predicted category: non-human

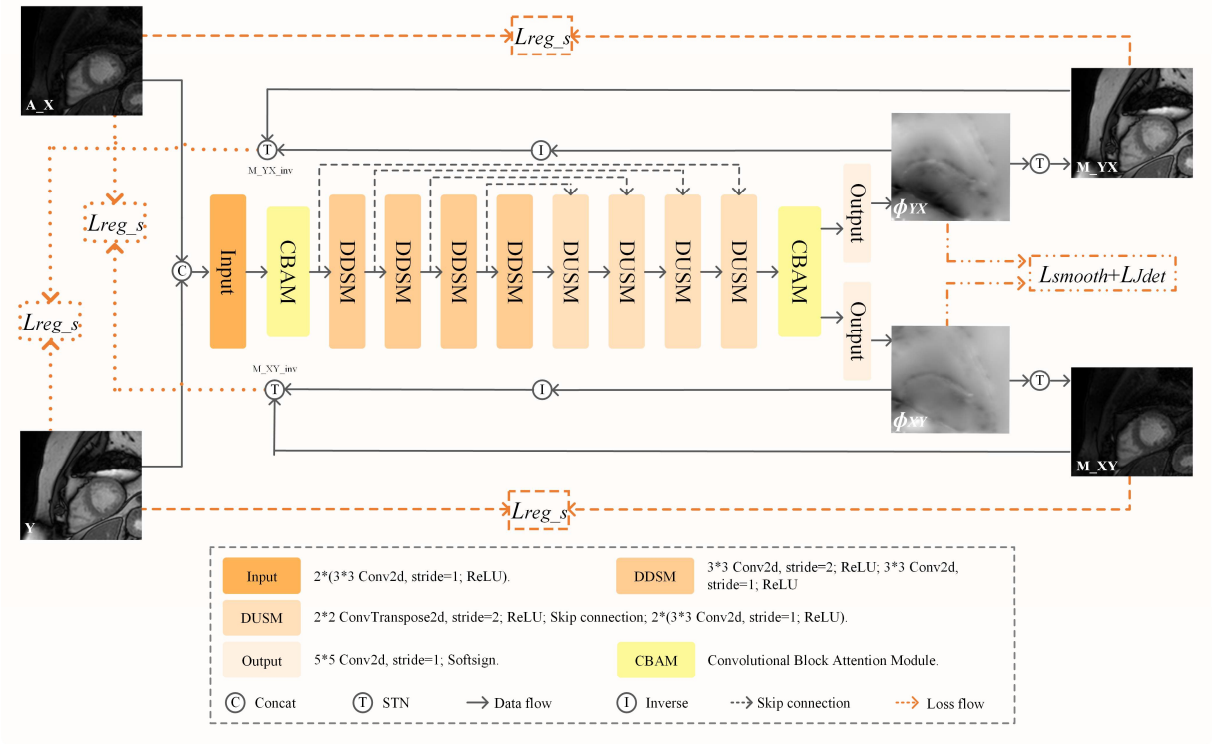
True category: non-human
Predicted category: non-human



workflow diagram of RS-MOCO



Implementation details of the network architecture for AffN.



Implementation details of the network architecture for RegN.

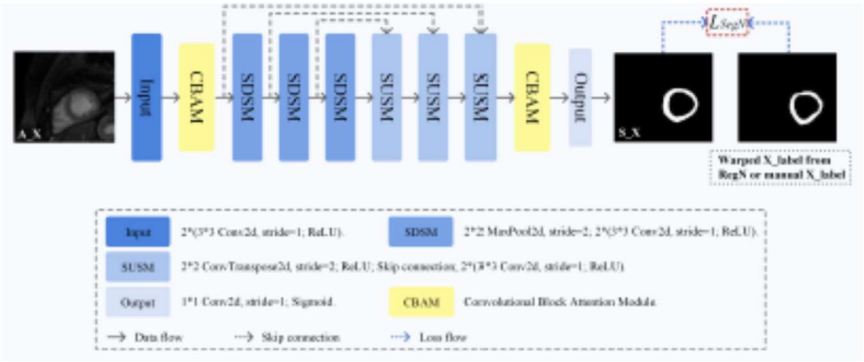


Fig. 5. Implementation details of the network architecture for SegN.

Transformer Net	**2025-70**	

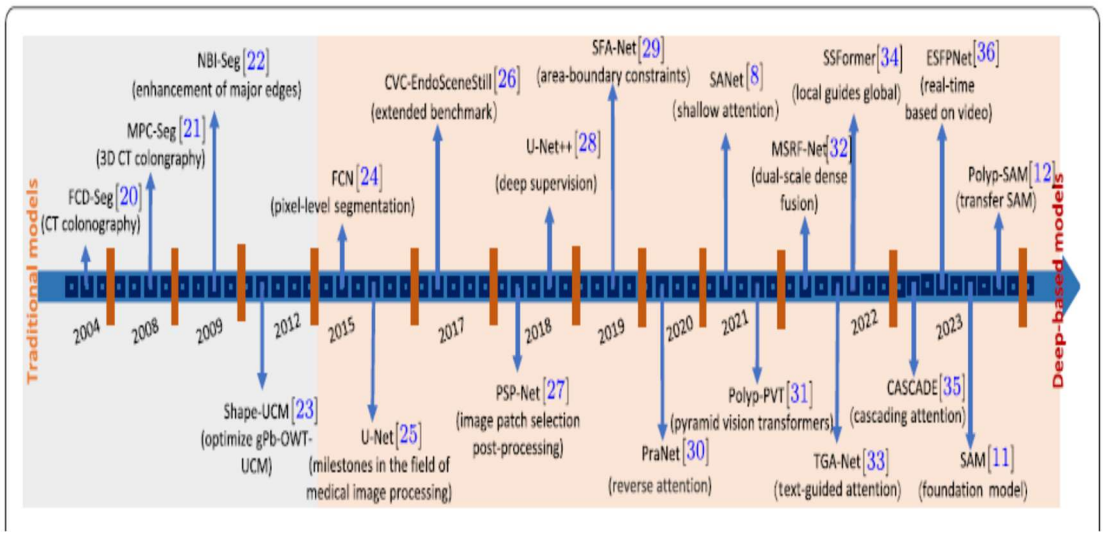
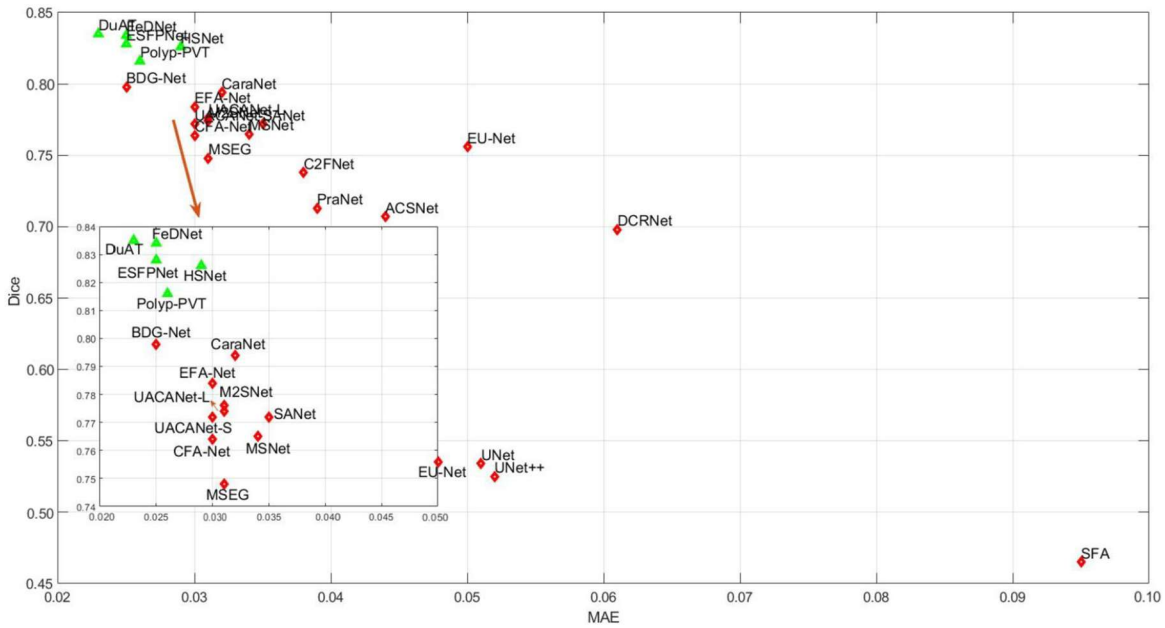
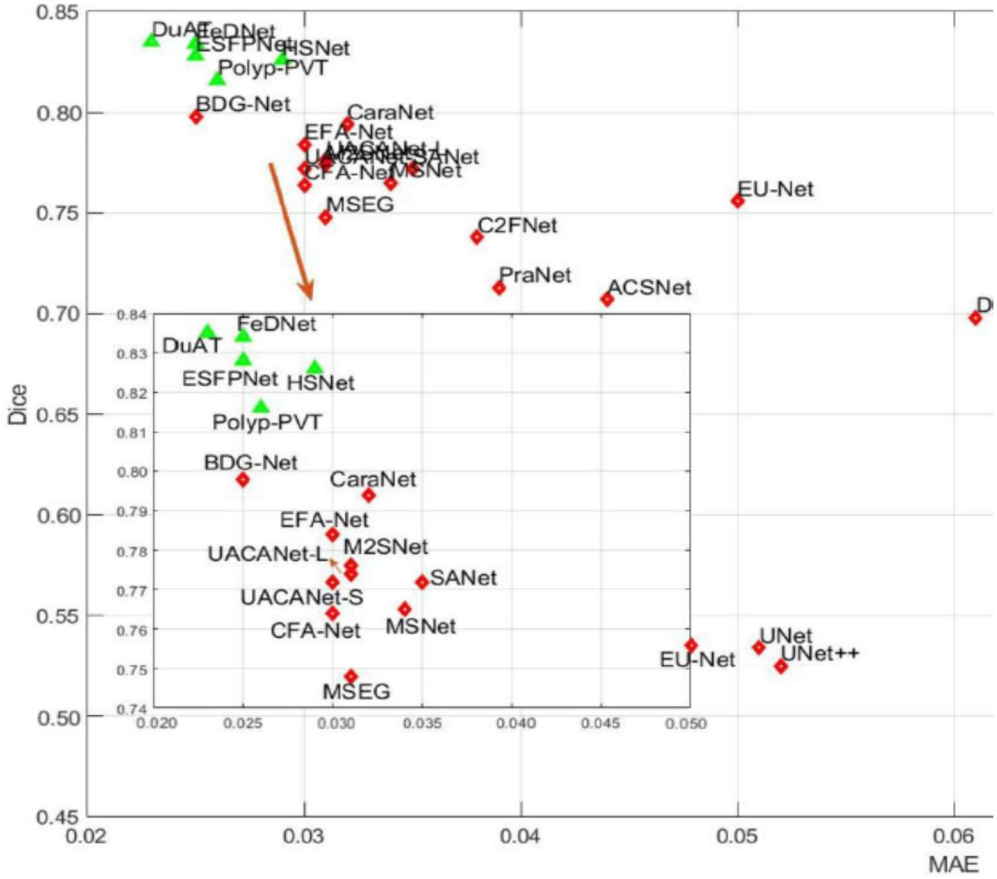
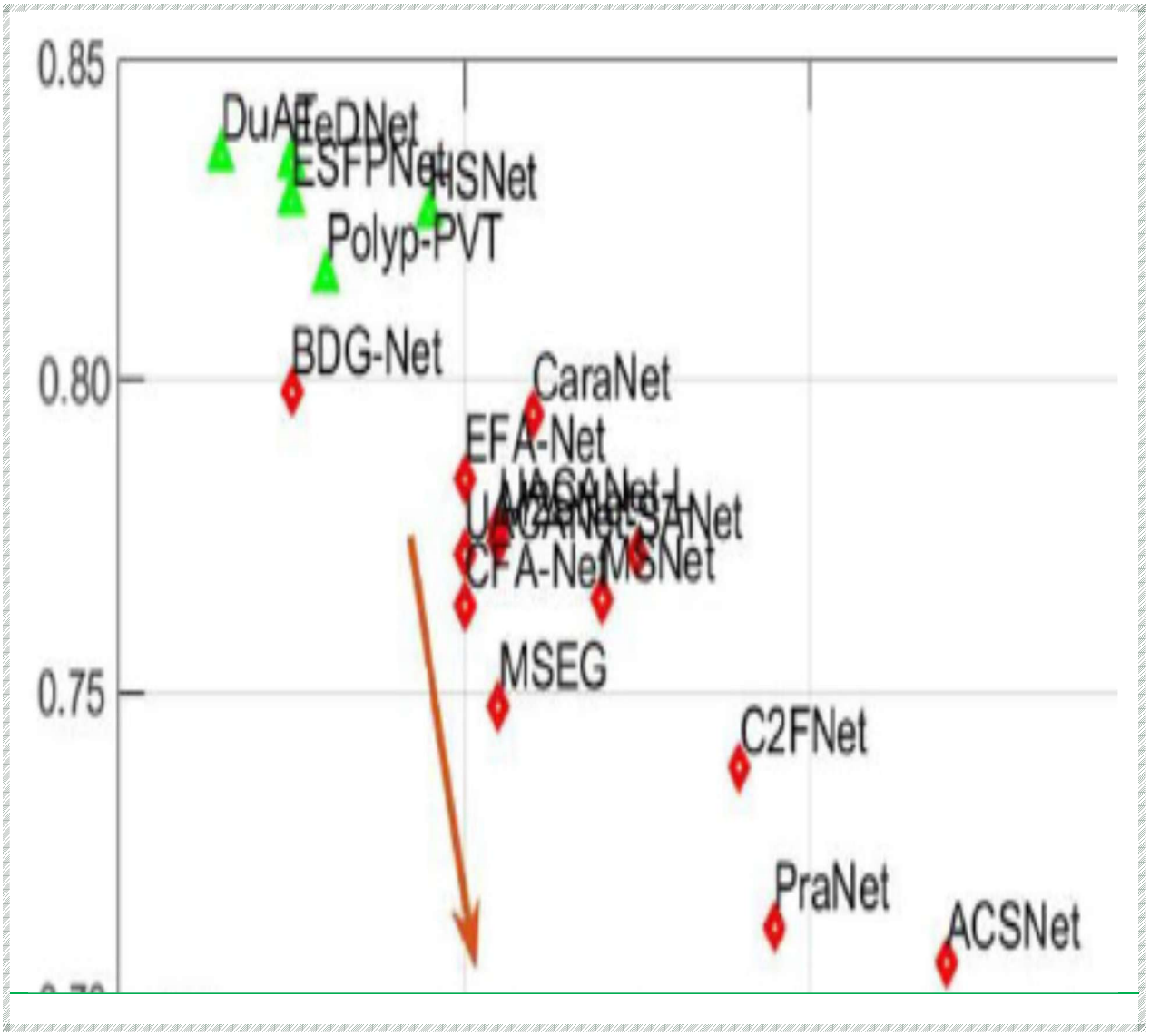


Figure 1 A brief chronology of polyp segmentation.

- ✓ Before 2015, methods relied on hand-crafted features combined with machine learning algorithms.
- ✓ Since 2015, U-Net [20] and FCN [21] have significantly advanced the development of deep learning techniques in polyp segmentation.




Dice
MAE
DuAT
FeDNet
ESFPNet
HSNet
Polyp-PVT
BDG-Net
CaraNet
EFA-Net
CFA-Net
MSNet
MSEG
EU-Net
C2FNet
PraNet
ACSNet
UACANet-L
M2SNet
SANet
UACANet-S
UNet
UNet++


0.85
0.80
0.75
DuAtt
ESFPNet
HSNet
Polyp-PVT
BDG-Net
CaraNet
EFA-Net
MSEG
C2FNet
PraNet
ACSNet

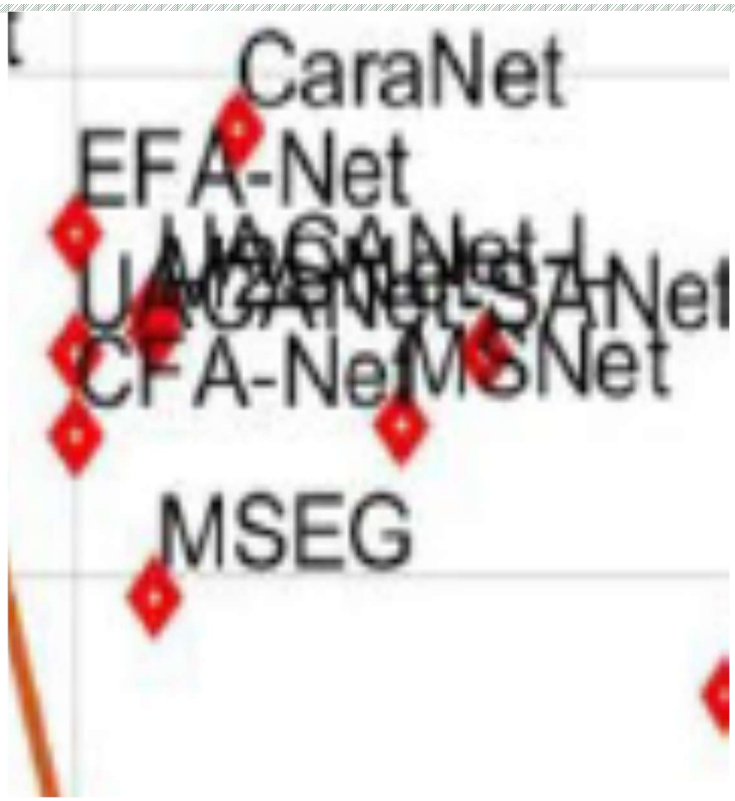


Figure 3 A comprehensive evaluation is conducted on 23 representative deep-learning models, including U-Net [20], UNet++ [101], SFA [22], PraNet [26], ACSNet [25], MSEG [30], EU-Net [31], SANet [8], MSNet [9], UACANet-S [34], UACANet-L [34], C2FNet [35], DCRNet [53], BDG-Net [50], CaraNet [44], EFA-Net [102], CFA-Net [75], M2SNet [103], Polyp-PVT [33], HSNet [46], DuAT [62], ESFPNet [72], and FeDNet [70], with SAM [104] excluded. We report the average Dice and MAE values for each model across five datasets (i.e., ETIS-LaribPolypDB [91], CVC-ColonDB [5], CVC-ClinicDB [7], CVC-300 [92], and Kvasir-SEG [93]). To Note that the models represented in the top left corner are better, i.e., they have larger Dice scores and smaller MAE values. In this context, the green triangles represent Transformer-based models, while the red diamonds signify CNN-based models

Transformer Net	**2025-70**	

Table 1 Summary of polyp segmentation methods (published from 2019 to 2021)

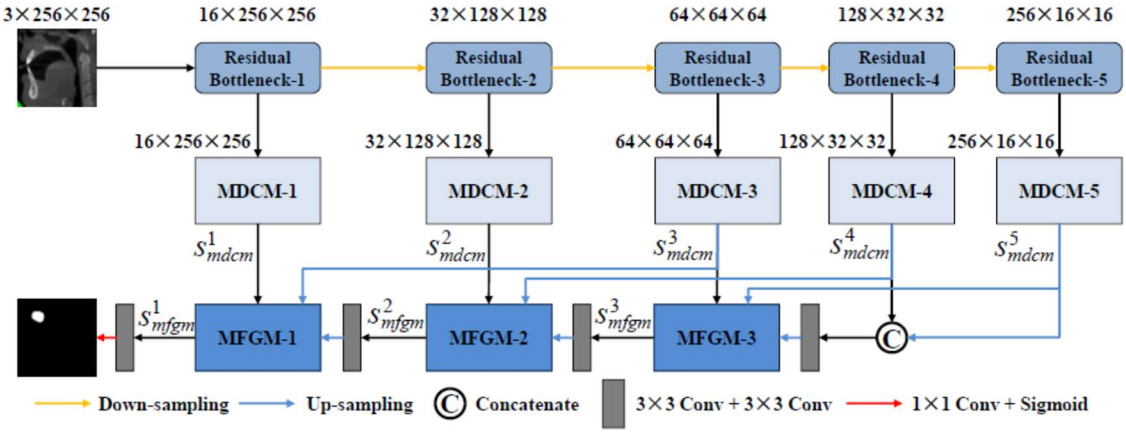
#	Year	Method	Pub.	Backbone	Description	Code
1	2019	SFA [22]	MICCAI	light UNet	Boundary-sensitive loss; selective feature aggregation	N/A
2	2019	ResUNet++ [23]	ISM	ResUNet	Squeeze and excitation blocks; atrous spatial pyramid pooling (ASPP); attention blocks	https://github.com/DebeshJha/ResUNetPlusPlus
3	2020	PolypSeg [24]	MICCAI	U-Net	Improved attention mechanism; separable convolution	N/A
4	2020	ACSNet [25]	MICCAI	ResNet34	Adaptively select; aggregate context features through channel attention	https://github.com/ReaFly/ACSNet
5	2020	PraNet [26]	MICCAI	Res2Net	Parallel partial decoders; reverse attention	https://github.com/DengPingFan/PraNet

					convolution	
44	2023	DuAT [62]	PRCV	PVT	Dual-aggregation Transformer; global-to-local spatial aggregation; selective boundary aggregation	https://github.com/Barrett-python/DuAT
45	2023	PolypSeg+ [63]	TCYB	ResNet50	Adaptive scale context module; lightweight attention mechanism	https://github.com/szuzzb/polypsegplus
46	2023	APCNet [64]	TIM	ResNet50	Attention-guided multi-level aggregation strategy; complementary information from different layers	N/A
47	2023	RA-DENet [65]	CBM	Res2Net	Improved reverse attention; distraction elimination	N/A
48	2023	EFB-Seg [66]	Neurocom-puting	ConvNet	Boundary Embedding; semantic offset field learned	N/A
49	2023	PPNet [67]	CBM	P2T	Channel attention; pyramid feature fusion	N/A
50	2023	Fu-TransHNet [68]	arxiv	HardNet68	CNN and Transformer; multi-view learning	N/A

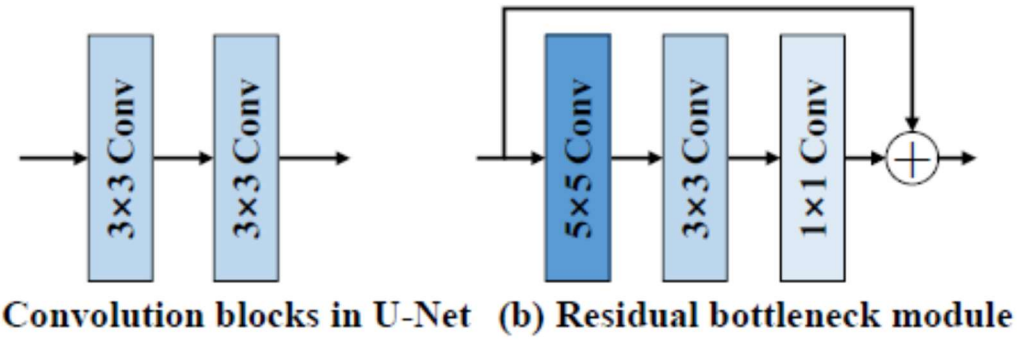
Table 4 Benchmark results of 24 representative polyp segmentation models (18 CNN-based and 6 Transformer-based models) on five commonly used datasets in terms of SPE and SEN. The top three results are displayed in bold, italic, and underlined fonts

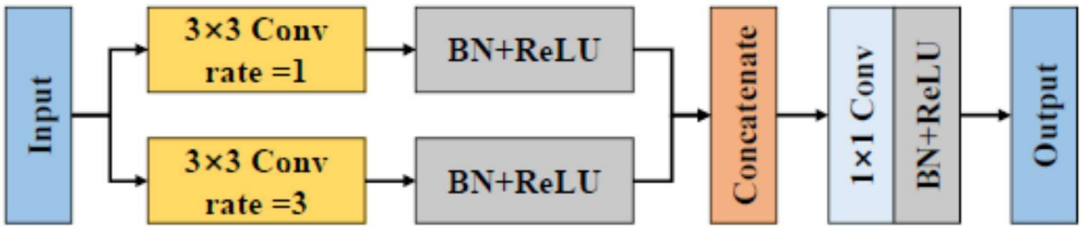
Method	Pub.	ETIS-Larib		CVC-ColonDB		CVC-ClinicDB		CVC-300		Kvasir	
		SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN
U-Net [20]	MICCAI 2015	0.703	0.484	0.798	0.525	0.947	0.835	0.965	0.768	0.949	0.857
UNet++ [101]	MICCAI 2018	0.727	0.415	0.828	0.497	0.927	0.795	0.957	0.738	*0.986*	0.807
SFA [22]	MICCAI 2018	0.781	0.633	0.861	0.703	0.919	0.802	0.934	0.889	0.965	0.799
PraNet [26]	MICCAI 2020	0.805	0.688	0.874	0.740	0.990	0.911	0.988	0.941	0.978	0.912
ACSNet [25]	MICCAI 2020	0.775	0.738	0.873	0.760	0.956	0.909	0.984	0.959	0.973	0.907
MSEG [30]	arxiv 2021	0.844	0.740	0.912	0.753	**0.992**	0.924	0.989	0.934	0.985	0.900
EU-Net [31]	CRV 2021	0.871	0.872	0.939	*0.851*	0.986	**0.960**	0.982	*0.969*	0.974	**0.934**
SANet [8]	MICCAI 2021	0.943	*0.904*	0.952	0.811	0.989	0.952	0.989	**0.971**	*0.986*	0.915
MSNet [9]	MICCAI 2021	0.893	0.796	0.931	0.775	0.975	0.933	0.988	0.931	0.981	0.911
UACANet-S [34]	ACM MM 2021	0.887	0.833	0.958	0.801	*0.991*	0.942	*0.992*	0.959	0.976	0.911
UACANet-L [34]	ACM MM 2021	0.932	0.813	0.953	0.754	**0.992**	0.943	**0.993**	0.940	0.983	0.923
C2FNet [35]	IJCAI 2021	0.902	0.745	0.894	0.752	0.973	0.941	0.988	0.952	0.974	0.904
DCRNet [53]	ISBI 2022	0.756	0.747	0.884	0.777	0.959	0.913	0.972	0.945	0.973	0.903
BDG-Net [50]	SPIE MI 2022	0.879	0.820	0.949	0.827	0.990	0.942	*0.992*	0.957	0.984	0.918
CaraNet [44]	SPIE MI 2022	0.910	0.812	0.947	**0.858**	*0.991*	0.955	0.976	0.927	0.982	0.912
EFA-Net [102]	arxiv 2023	0.918	0.866	0.940	0.820	0.975	0.934	0.988	0.950	**0.987**	0.914
CFANet [75]	PR 2023	0.910	0.804	0.953	0.761	*0.991*	**0.960**	0.990	0.952	0.985	0.926
M2SNet [103]	arxiv 2023	0.893	0.796	0.931	0.775	0.975	0.933	0.988	0.931	0.981	0.911
Polyp-PVT [33]	arxiv 2021	**0.962**	0.902	*0.965*	0.829	**0.992**	*0.959*	**0.993**	0.943	**0.987**	0.928
HSNet [46]	CBM 2022	0.955	0.868	*0.965*	0.821	**0.992**	0.949	0.991	0.947	*0.986*	0.913
DuAT [62]	PRCV 2023	0.941	0.891	0.962	0.841	**0.992**	0.956	0.991	0.956	0.984	*0.933*
ESFPNet [72]	MI 2023	*0.961*	**0.917**	0.961	0.837	**0.992**	0.940	0.991	0.967	0.985	0.910
FeDNet [70]	BSPC 2023	0.945	0.893	**0.966**	0.845	*0.991*	0.954	*0.992*	0.950	**0.987**	0.924
SAM-B [104]	arxiv 2023	0.717	0.415	0.621	0.246	0.681	0.309	0.730	0.412	0.904	0.510
SAM-H [104]	arxiv 2023	0.768	0.525	0.811	0.480	0.877	0.547	0.873	0.685	0.934	0.769
SAM-L [104]	arxiv 2023	0.810	0.567	0.813	0.500	0.834	0.623	0.904	0.756	0.935	0.774

Transformer Net	**2025-71**	



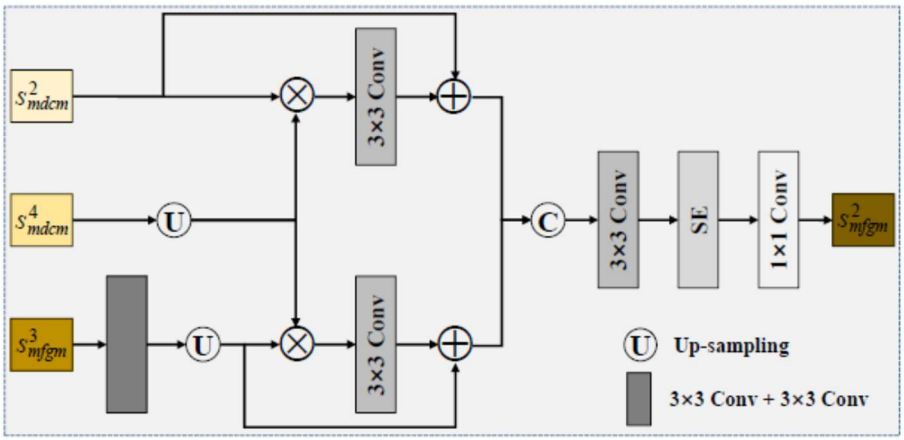
The proposed RBMDC-Net architecture





Structure of multiscale dilated convolution module.

MULTI-LEVEL FEATURE GUIDANCE MODULE



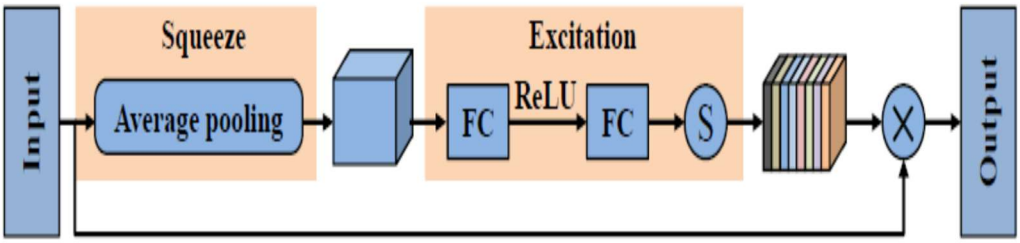


FIGURE 5. Structure of SE module.

TABLE 4. Ablation experiment on the original dataset.

Method	F1 (%)	Mcc (%)	Jaccard (%)
U-Net	0.8738	0.8772	0.7786
U-Net+RBM	0.9183	0.9178	0.8496
U-Net+MFGM	0.9190	0.9186	0.8512
U-Net+MDCM	0.9245	0.9241	0.8603
U-Net+RBM+MFGM	0.9290	0.9285	0.8678
U-Net+MFGM+MDCM	0.9251	0.9247	0.8612
U-Net+RBM+MDCM	0.9263	0.9259	0.8631
RBMDC-Net	**0.9313**	**0.9308**	**0.8717**

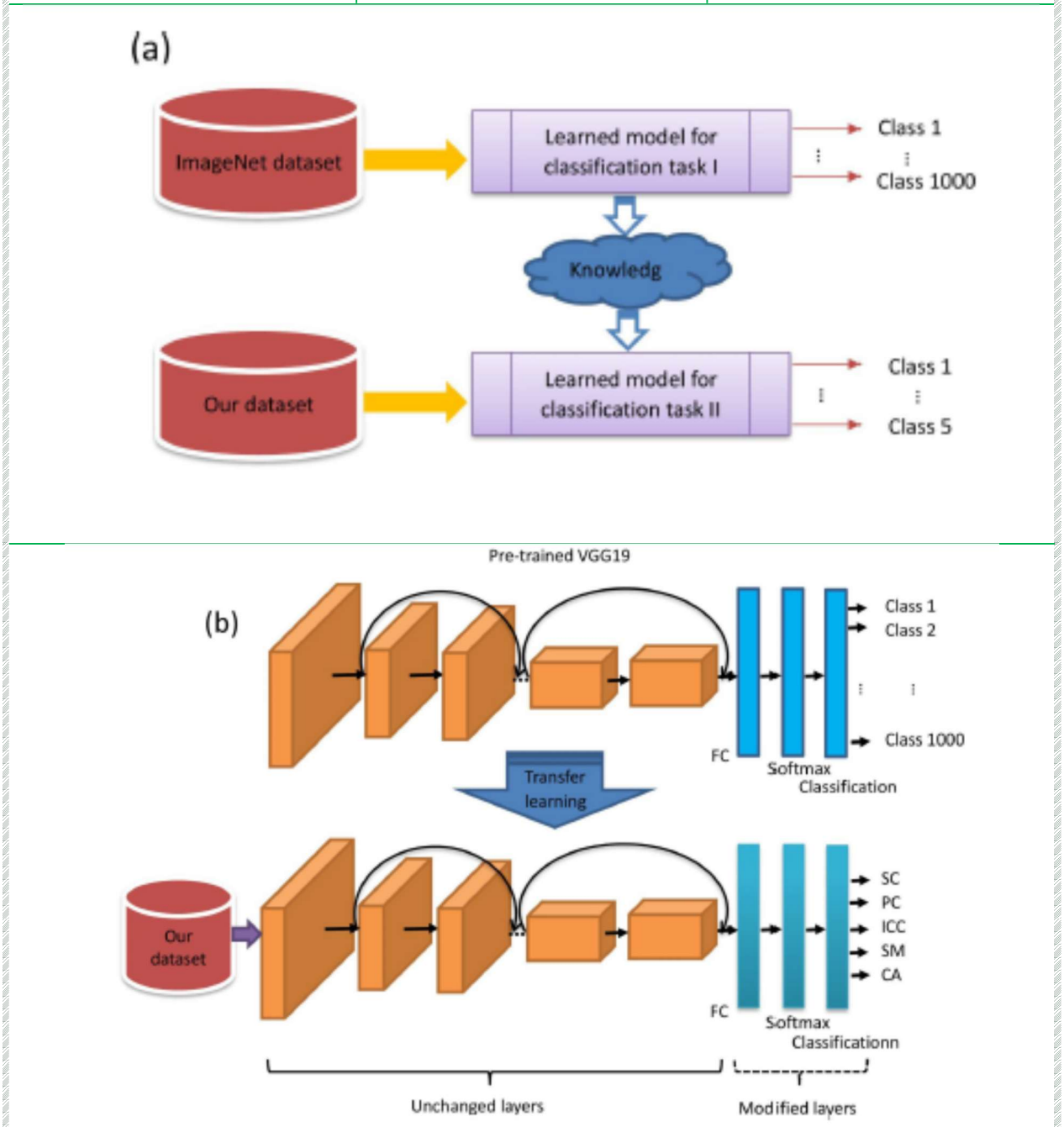


Fig. 3. Deep learning framework. (a) A transfer learning framework (b) Modifying VGG19 for the application

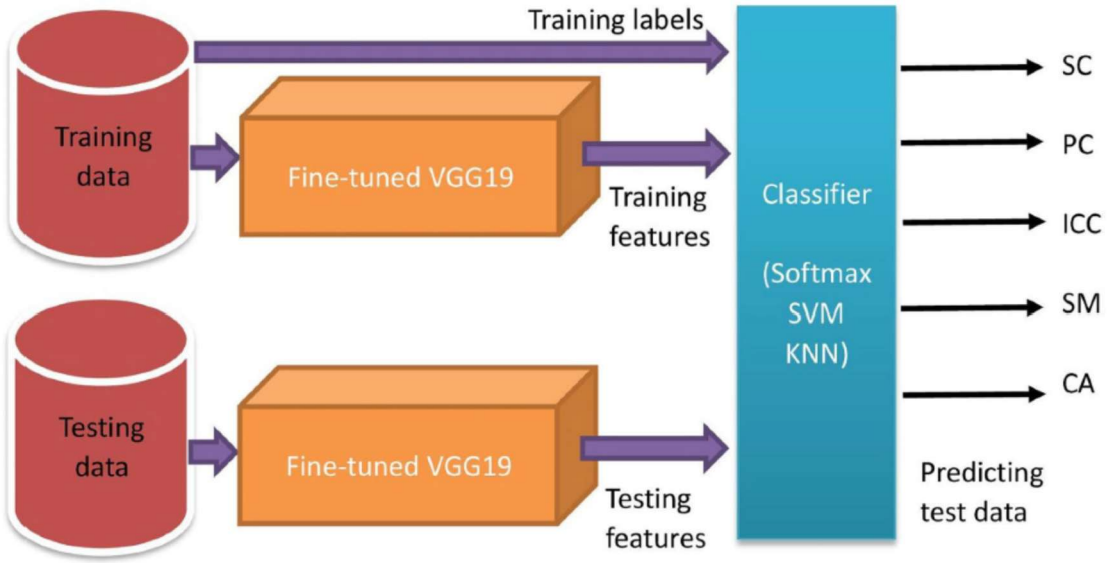
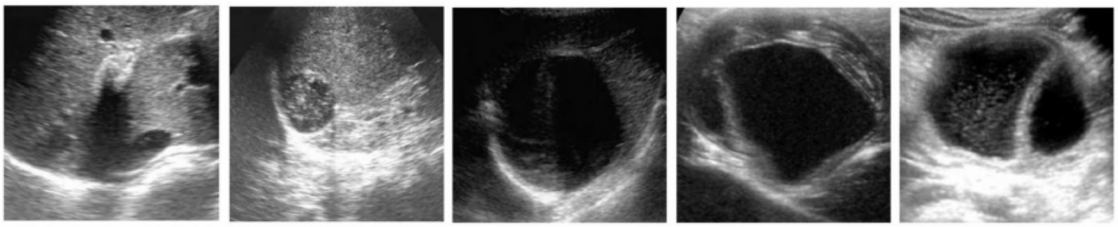


Fig. 5. Overall framework for HCE classification.



(a) (b) (c) (d) (e)

Fig. 9. Misclassified samples.
(a) CA type misclassified as ICC by softmax-based classifier, but correctly classified by SVM.
(b) SM type misclassified as polycystic by softmax-based classifier, but correctly classified by SVM.
(c) SC type misclassified as ICC for both softmax-based classifier and SVM.
(d) PC type misclassified as ICC for both softmax-based classifier and SVM.
(e) ICC type missclassified as PC for both softmax-based classifier and SVM.

Transformer Net	**2025-91**	

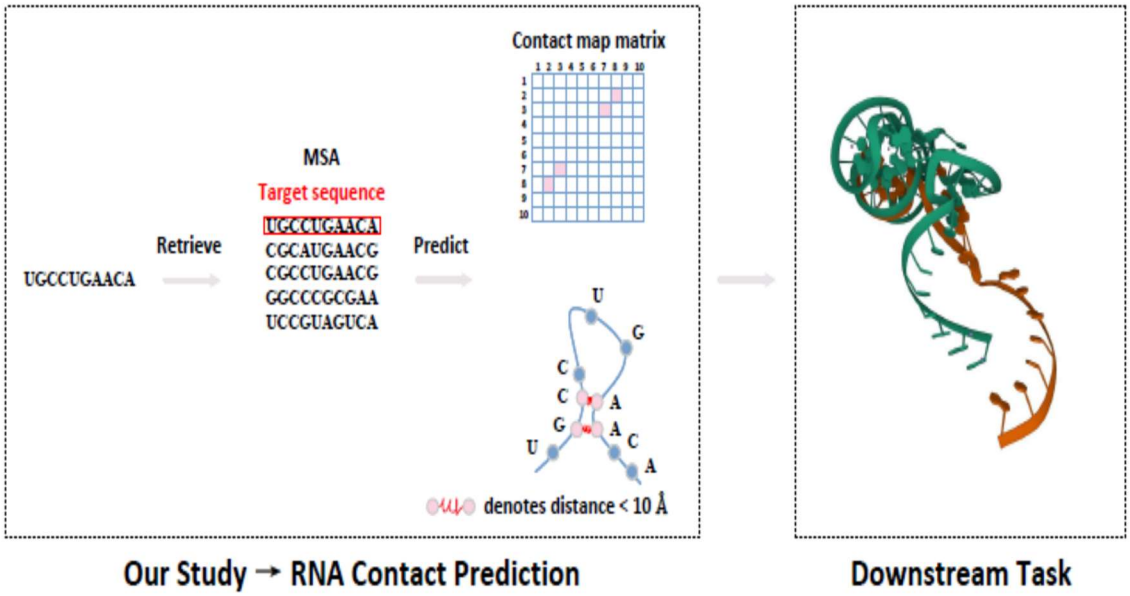


Figure 1. Our study is focused on RNA contact prediction, i.e., predicting the contact map matrix for an RNA sequence.
The contact map indicates the proximity between each nucleotide, with those closer than a threshold (10 A°) being deemed in contact.
Correct predictions of the contact map can benefit downstream tasks, e.g., by acting as constraints for filtering 3D RNA structure predictions

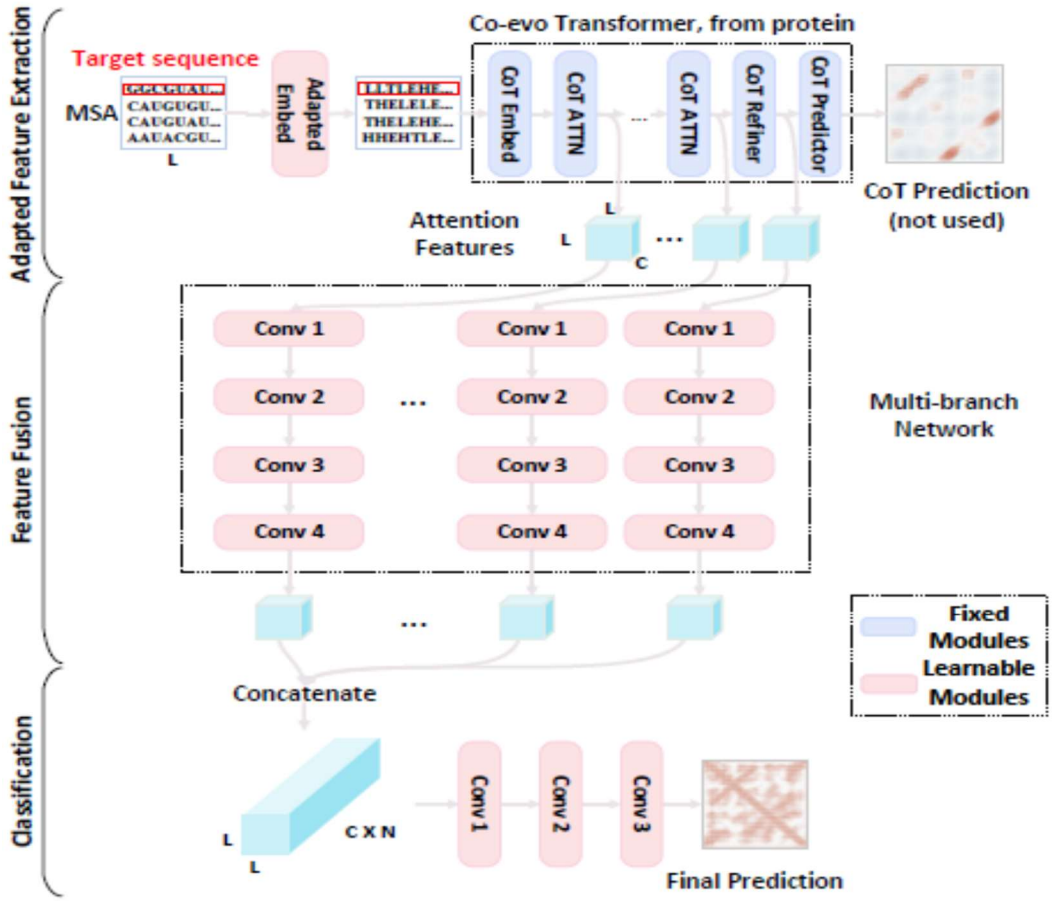


Figure 2. Overview of our three-stage method (from top to bottom).
Adapted Feature Extraction: First, a projection layer is used to translate the RNA MSA sequences into protein language (e.g., from nucleotide "AUCG" to amino acids "HETL"). Then, we leverage a fixed large-scale pre-trained protein contact prediction transformer model (called Co-evolution Transformer model (CoT)) to extract attentive (i.e., contribution) features at different layers. Feature Fusion: Features from different layers are processed by separate convolution blocks before being concatenated. Classification: The aggregated features are sent into a standard Convolutional Network (ConvNet) classifier with three layers of convolution

Transformer Net	2025-99	

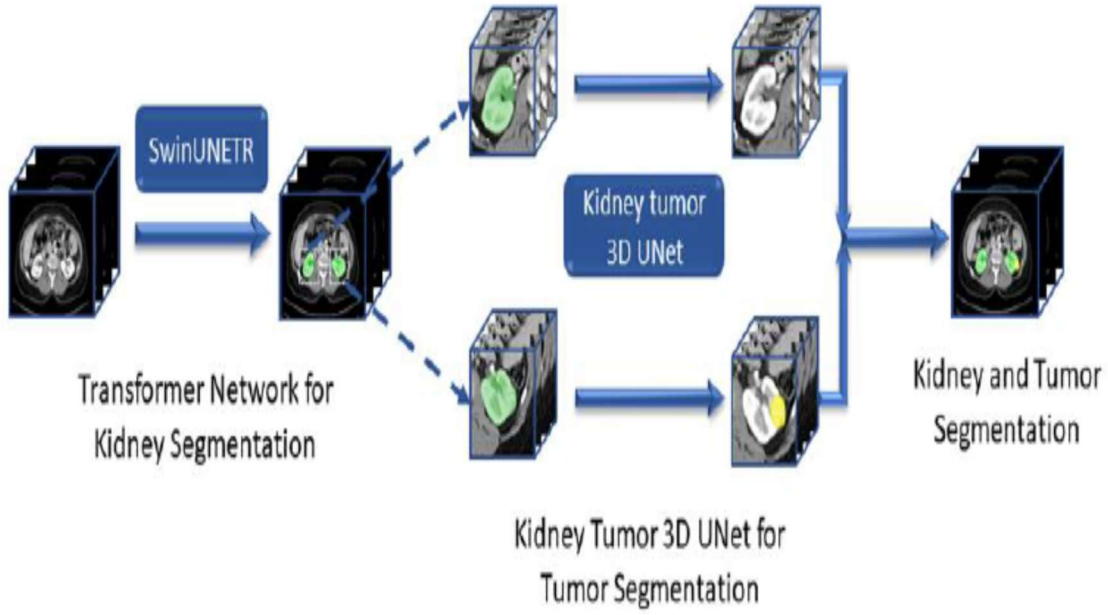


Figure 1. Proposed dual-stage kidney and tumor segmentation framework

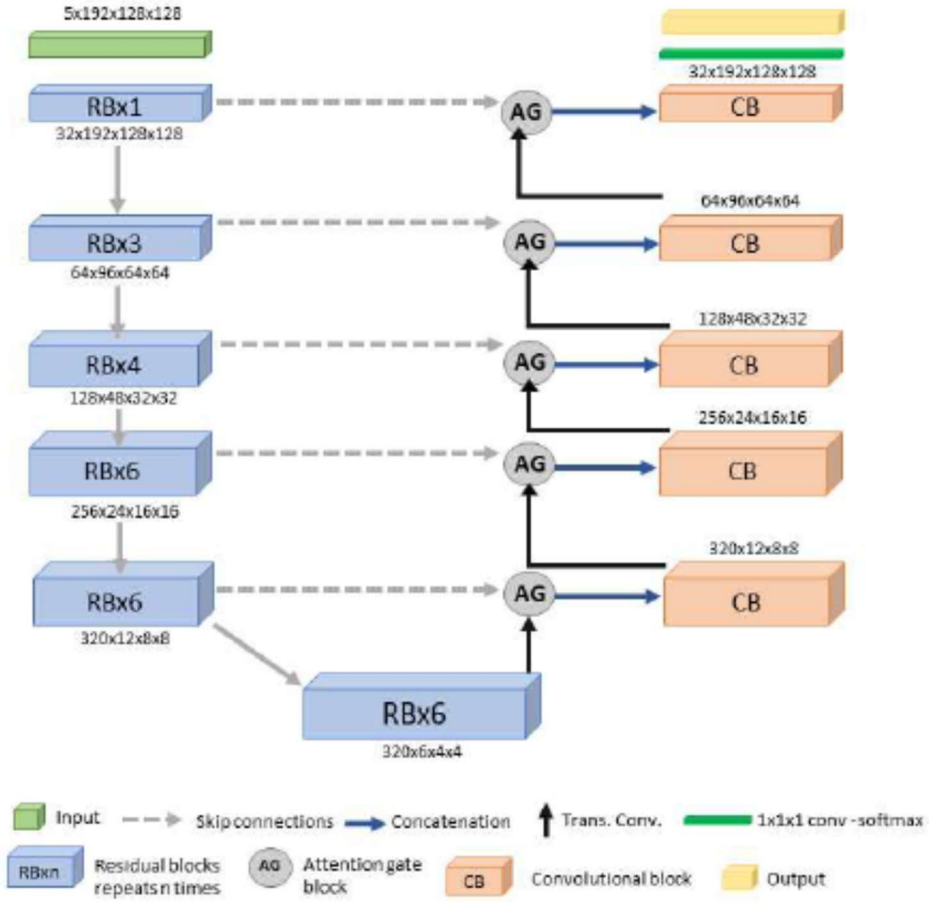
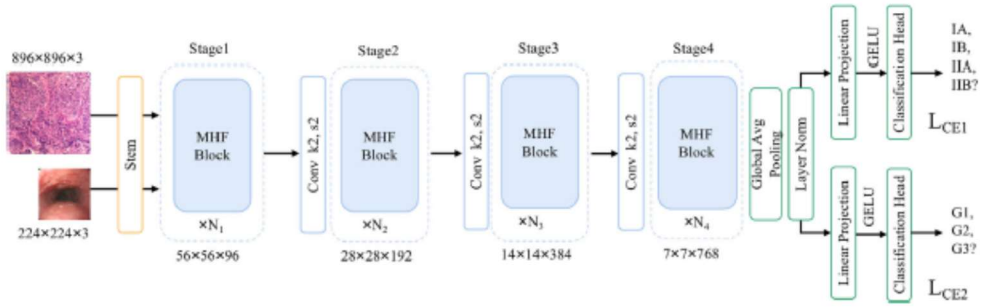
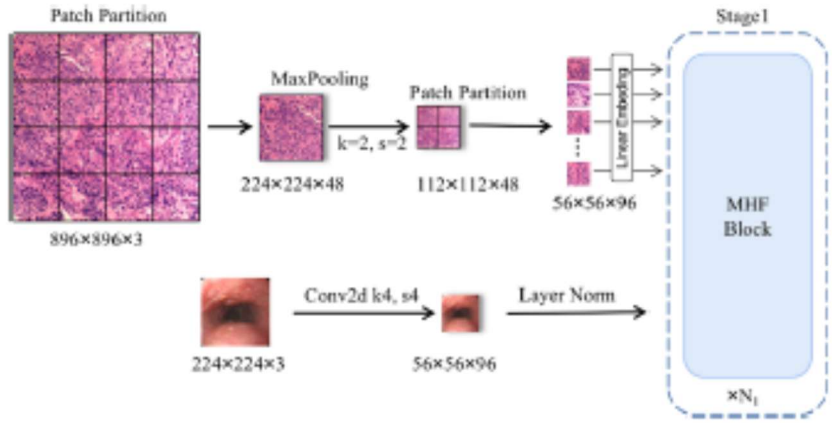


Figure 2. Proposed Kidney Tumor 3D UNet network architecture

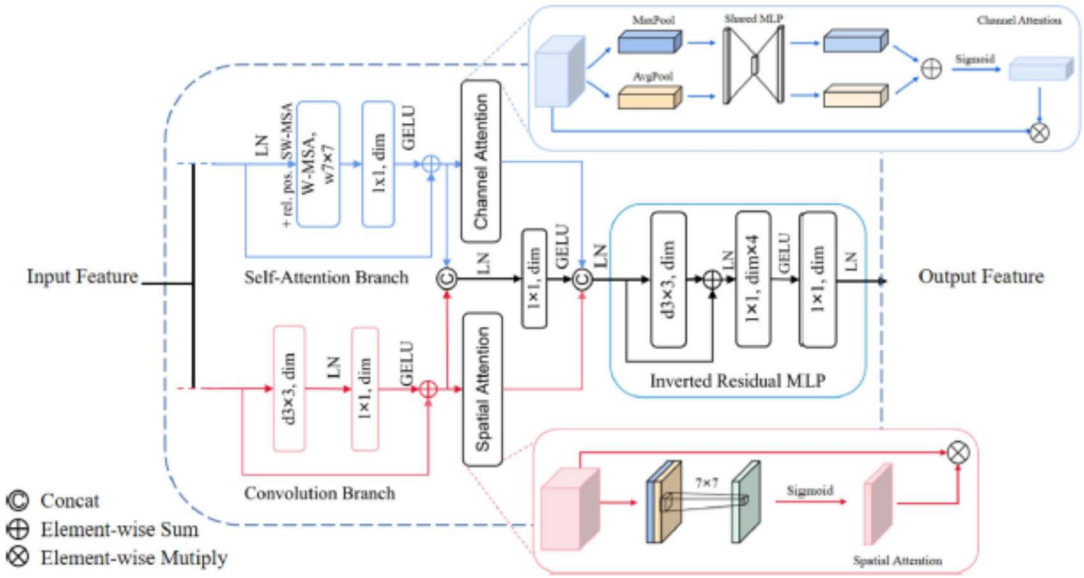
Transformer Net	**2025-102**	



Overall structure of the MM-HiFuse model. Stem is used to process multimodal images of different sizes input in the early stage, convert
them into suitable tensors, and input them into the backbone. Then, features are extracted through the backbone and output to two task branches

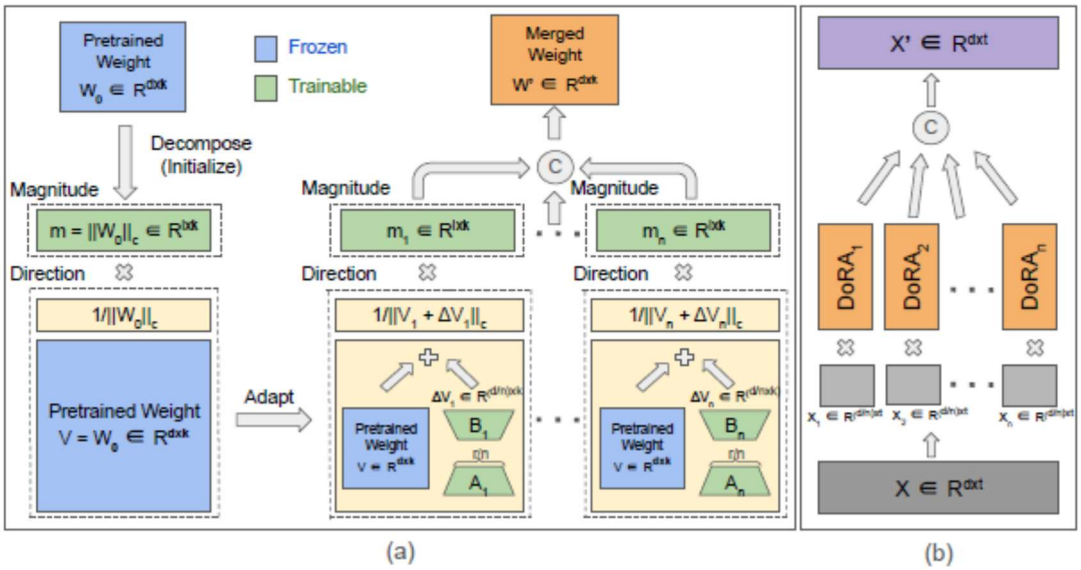


Multi-modal stem detail display

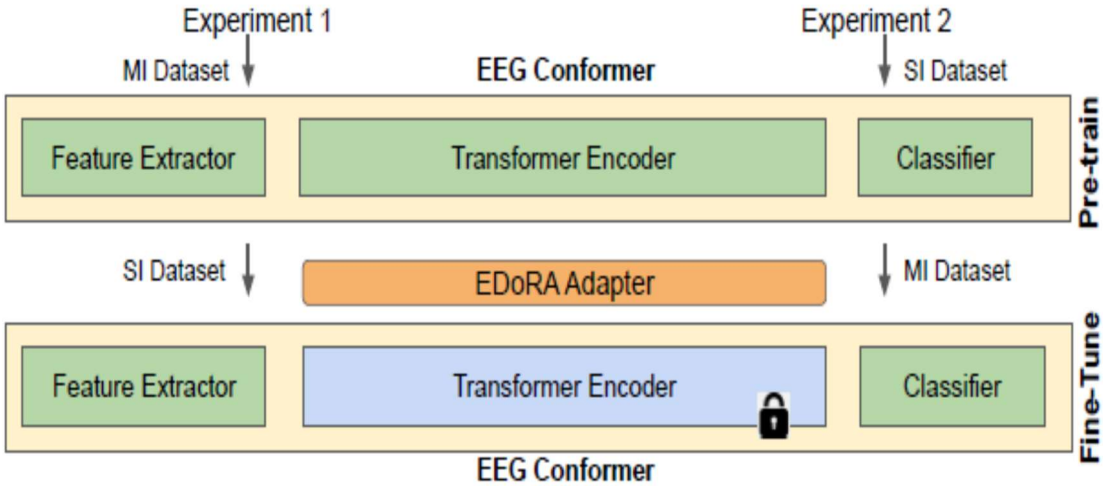


MHF block detail display. The MHF block in the backbone consists of a self-attention branch and a convolution branch.
After a series of complex transformation processes (including channel attention, spatial attention, and IRMLP), the fusion feature is finally output

Transformer Net	2025-107	

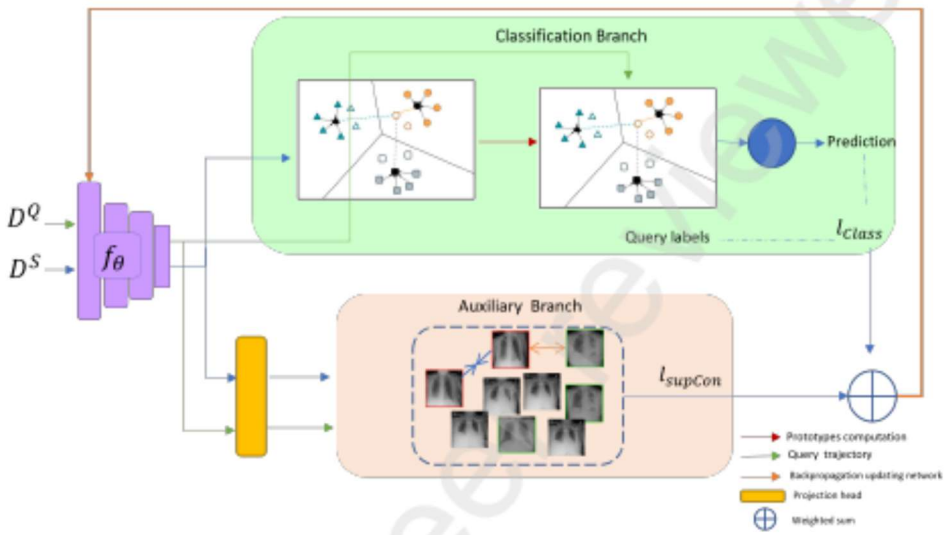


(a) Overview of proposed EDoRA paramter-efficient fine-tuning approach. It depicts the overall parameter updation process of EDoRA, before and after fine-tuning (b) Feature updation via EDoRA adaptation. [© symbol represents concatenation, × symbol represents product, X represents input features, X′ represents output features.]

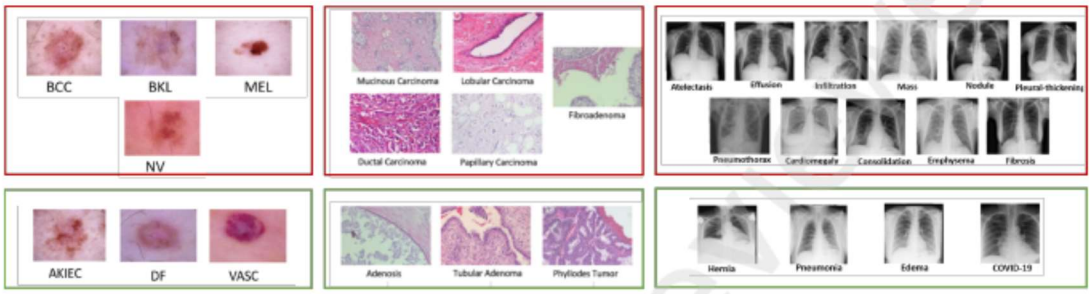


Framework of the proposed method. Two experiments are performed in this work, and in these experiments EEG Conformer model is pre-trained on one dataset, and then fine-tuned on other dataset with only EDoRA adapter on each operation of transformer encoder of EEG Conformer and vice-versa. [Freezed weights are shown with lock]

Transformer Net	2025-130	



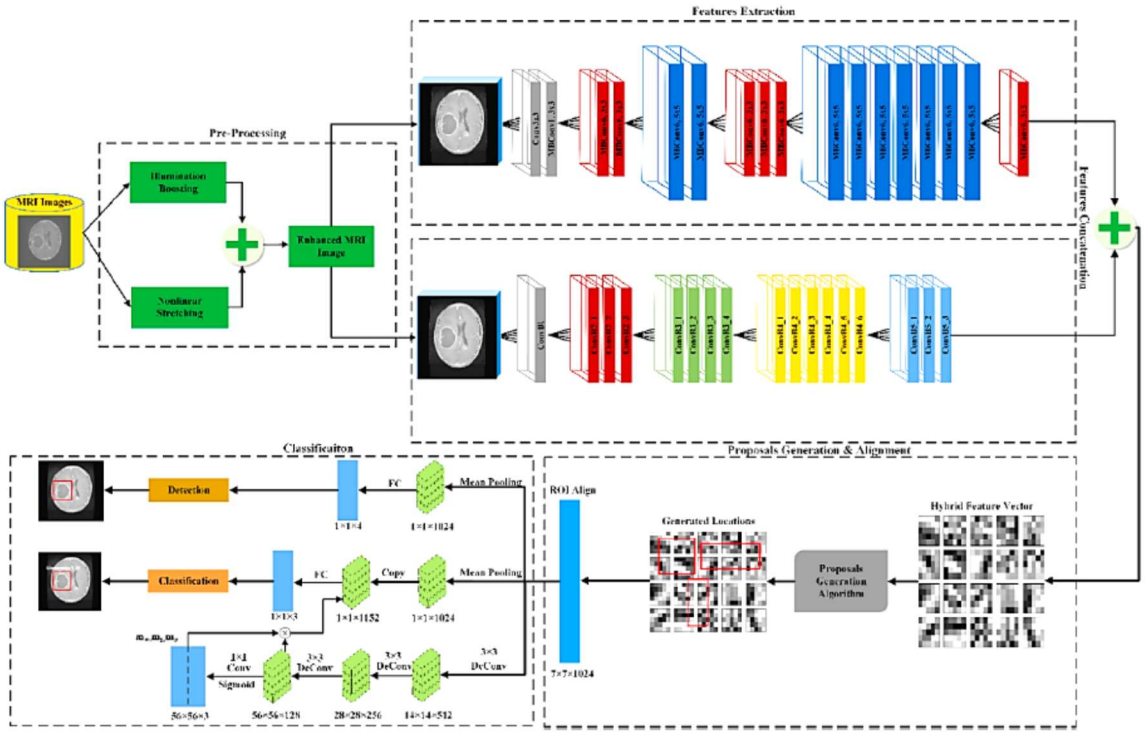
Proposed framework for few shot diagnosis



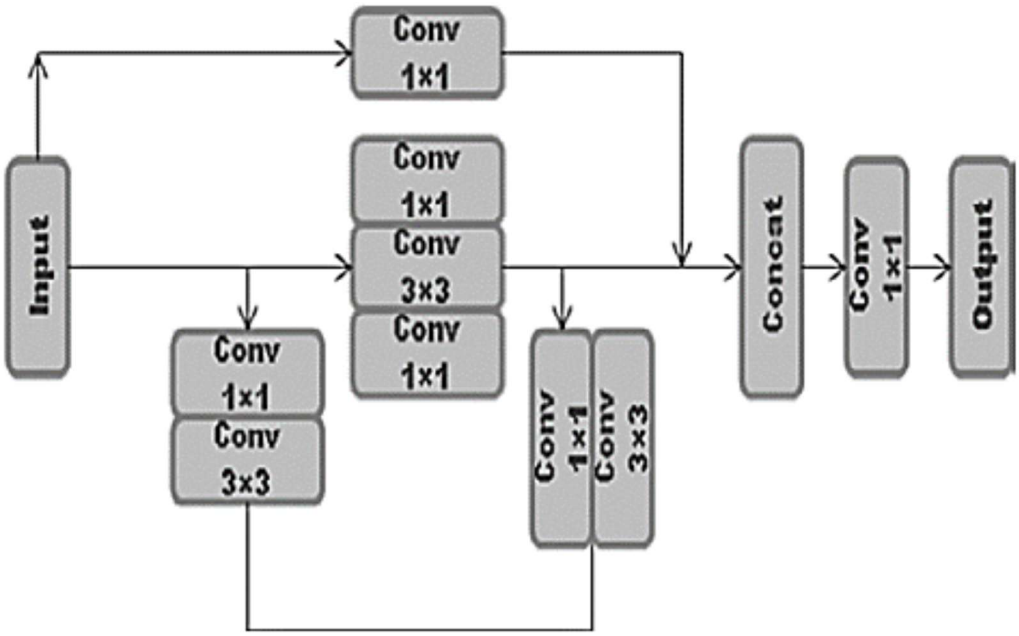
\\\\

Examples of images from the three selected datasets: First row (Red) and second row (Green) present respectively examples from the meta-train and the meta-test datasets

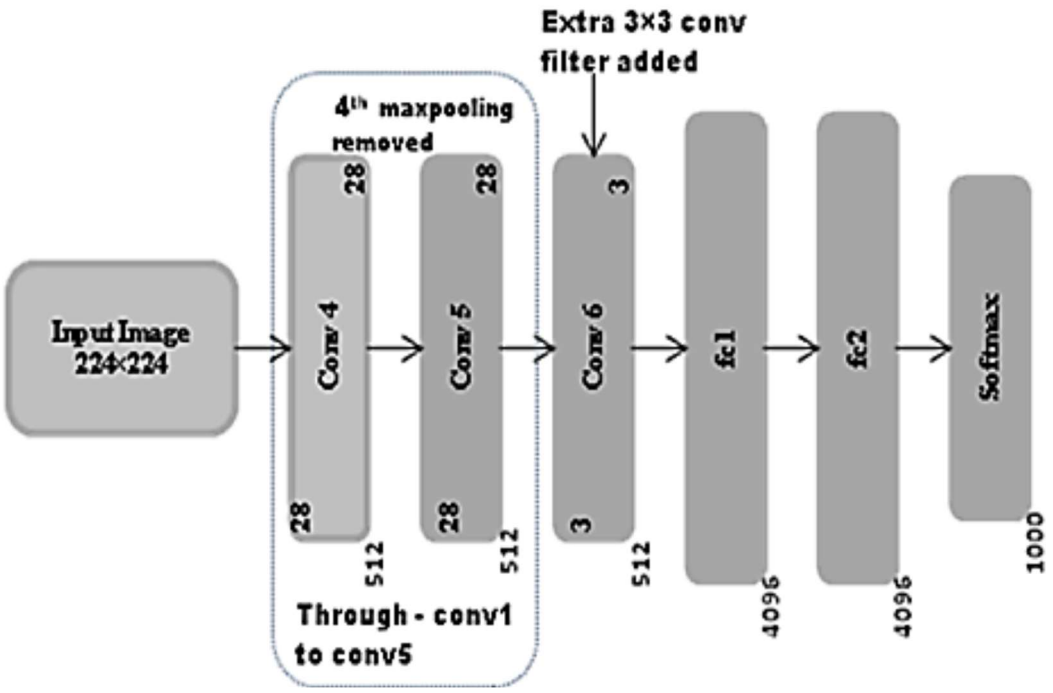
Transformer Net	2025-144	



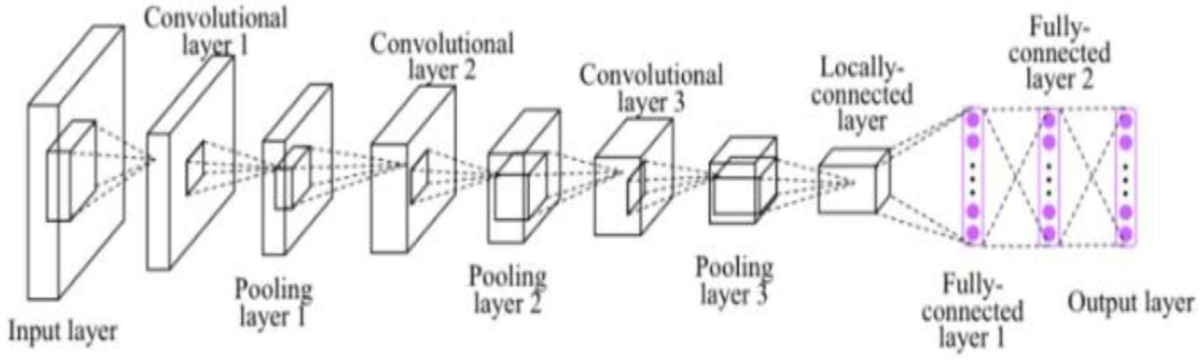
The overall significant steps involved tumor classification[16]



The architecture of ResNet

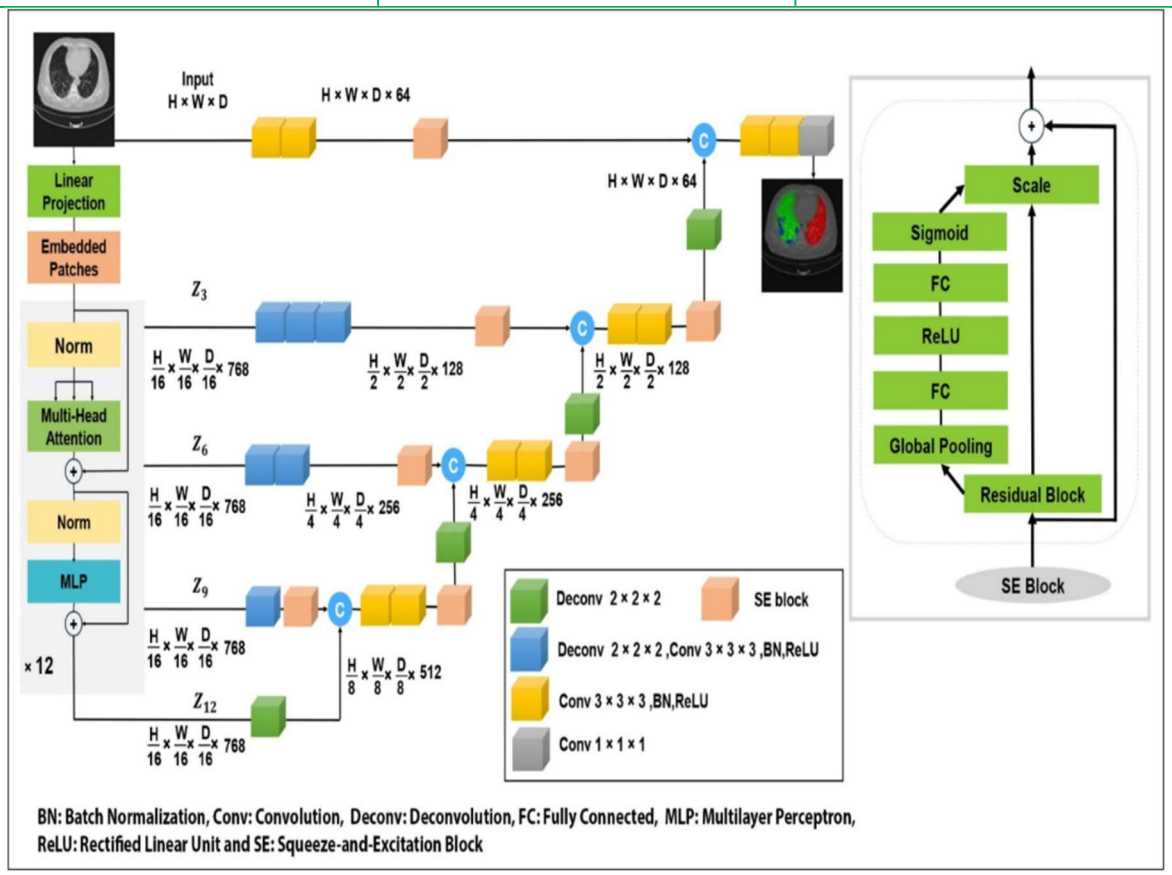


The improved architecture of VGG network

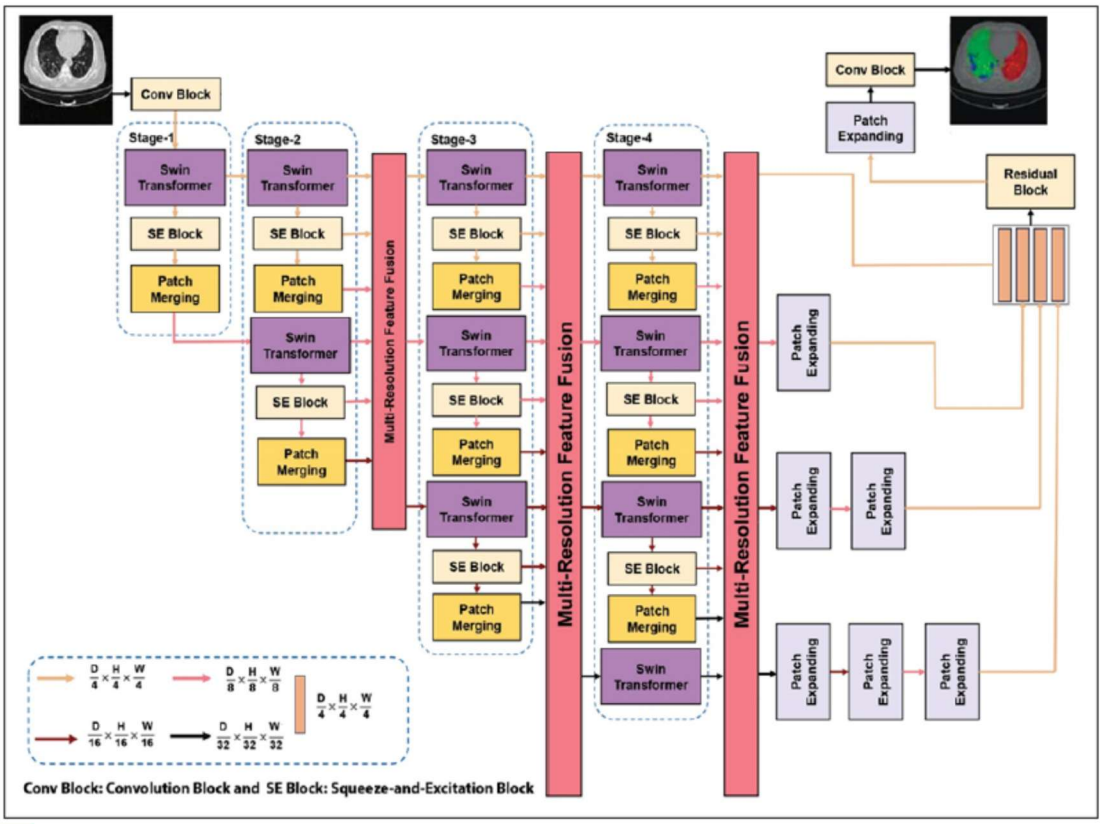


Architecture of CNN

Transformer Net	2025-145	

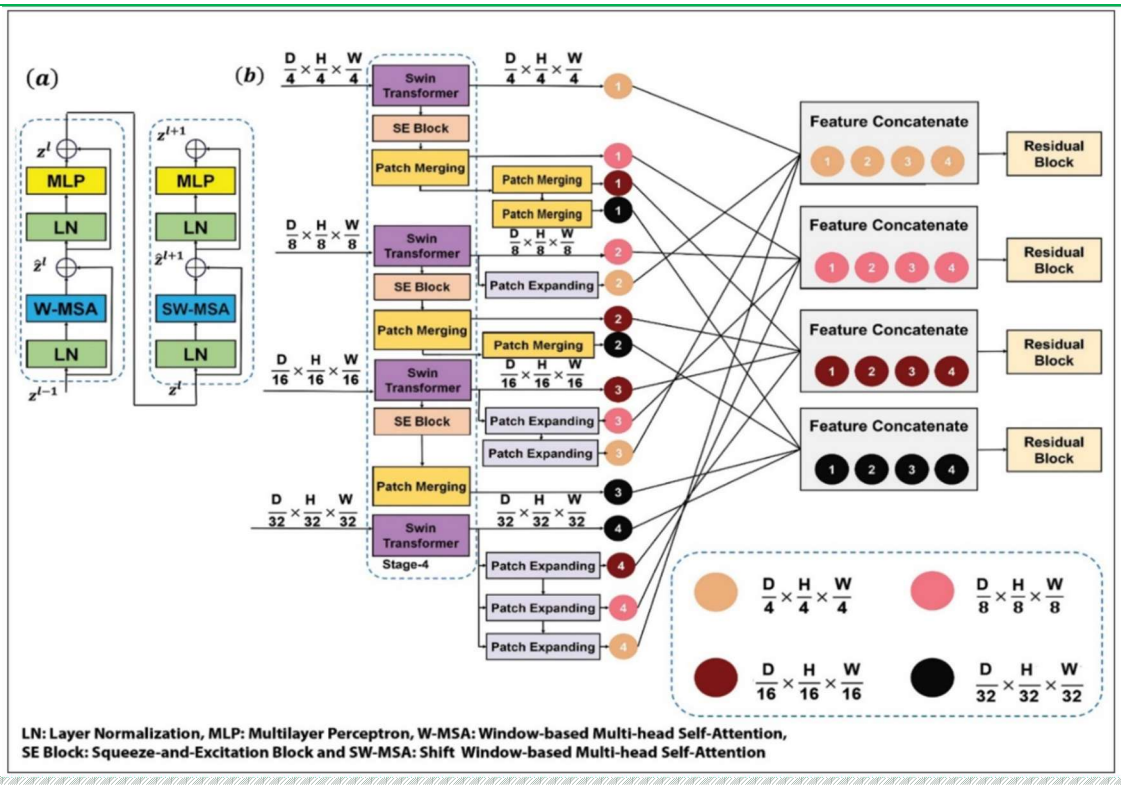


Squeeze and Excitation-based UNet TRansformers (SE-UNETR) architecture.
H:height;W: width; D: depth

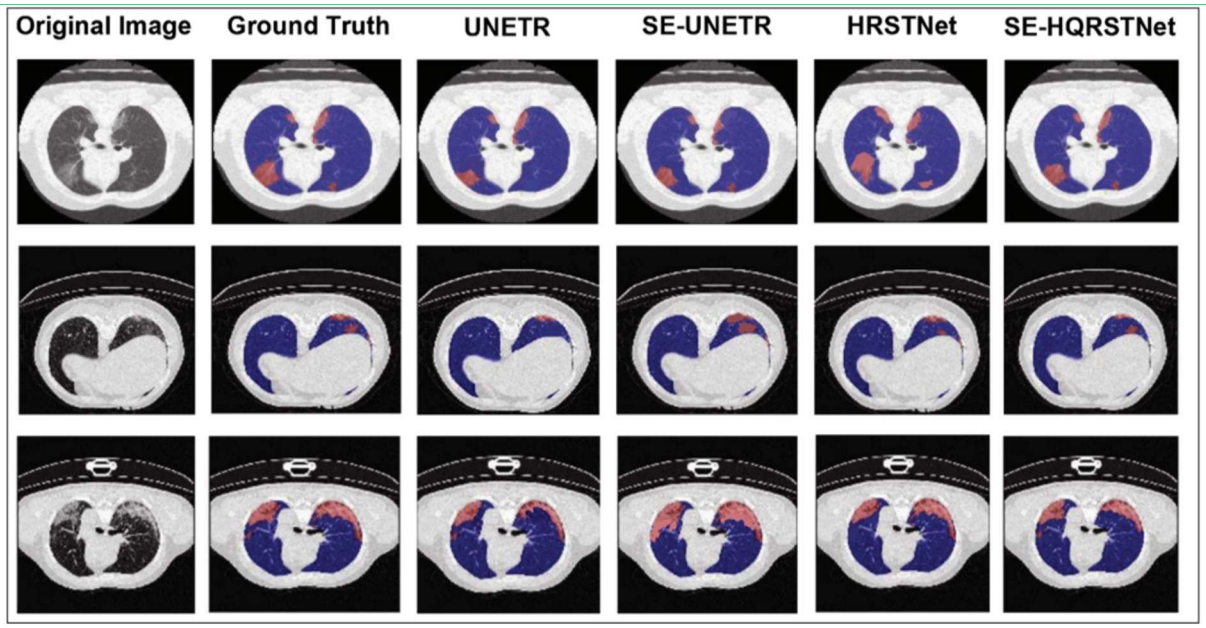


Squeeze and Excitation-based High-Quality Resolution Swin Transformer Network (SE-HQRSTNet) architecture.

H:height;W: width; D: depth

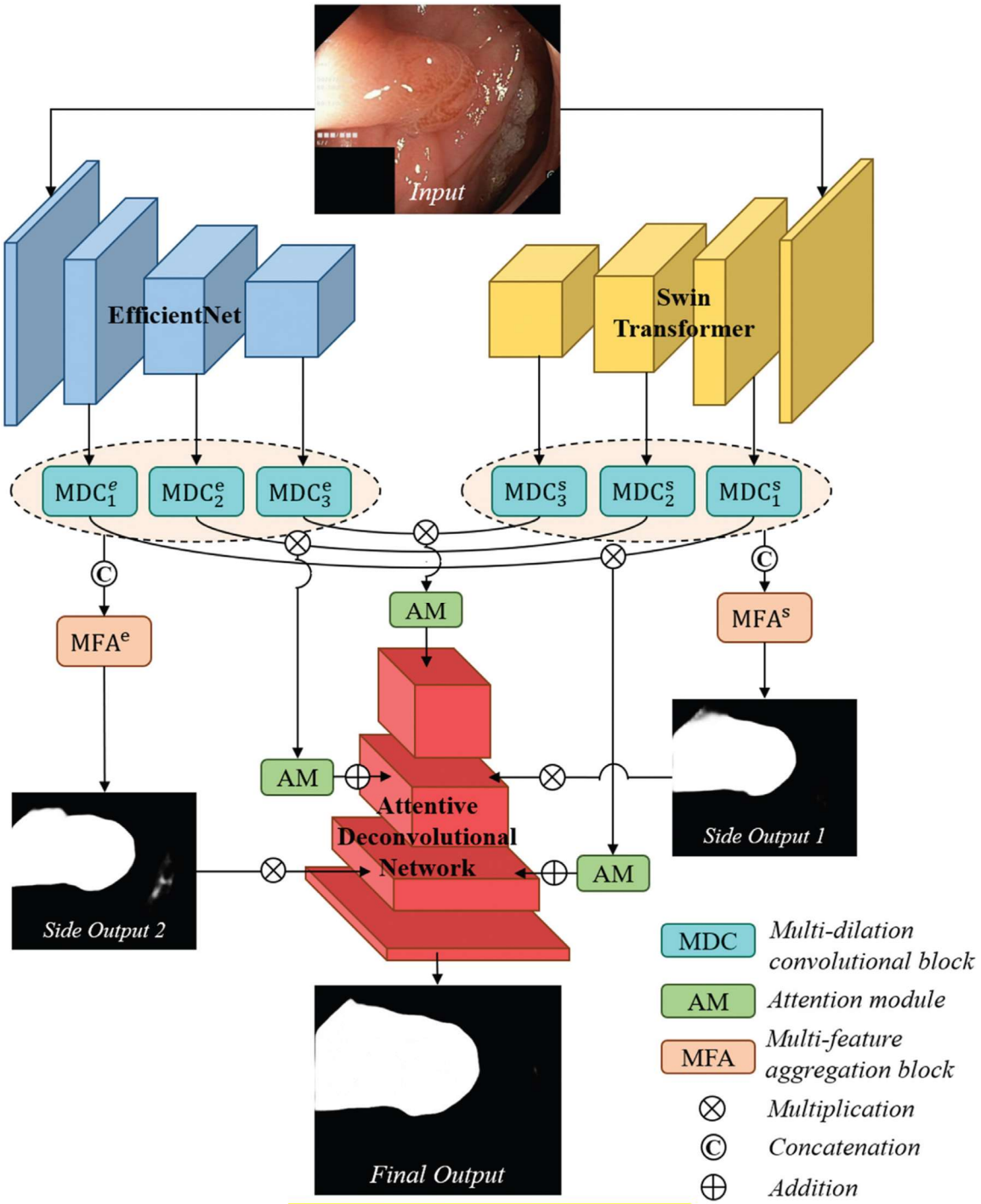


(a) Swin Transformer block, and (b) multi-resolution feature fusion (MRFF) block. H:height;W: width; D: depth

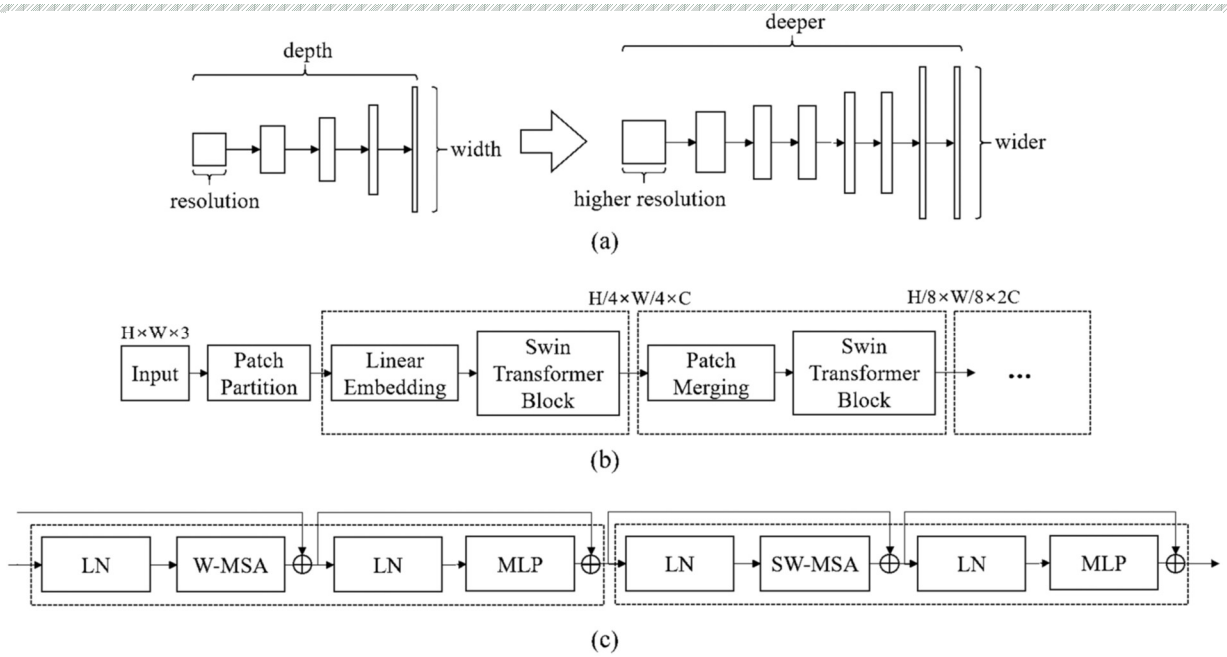


Visual comparison between the ground truth and prediction of the models segmentation for 3 computed tomography (CT) scan samples of COVID-19 patients. SE: Squeeze-and-Excitation, UNETR: UNEt TRansformers, HRSTNet: High-Resolution Swin Transformer Network, HQRSTNet: High-Quality Resolution Swin Transformer Network

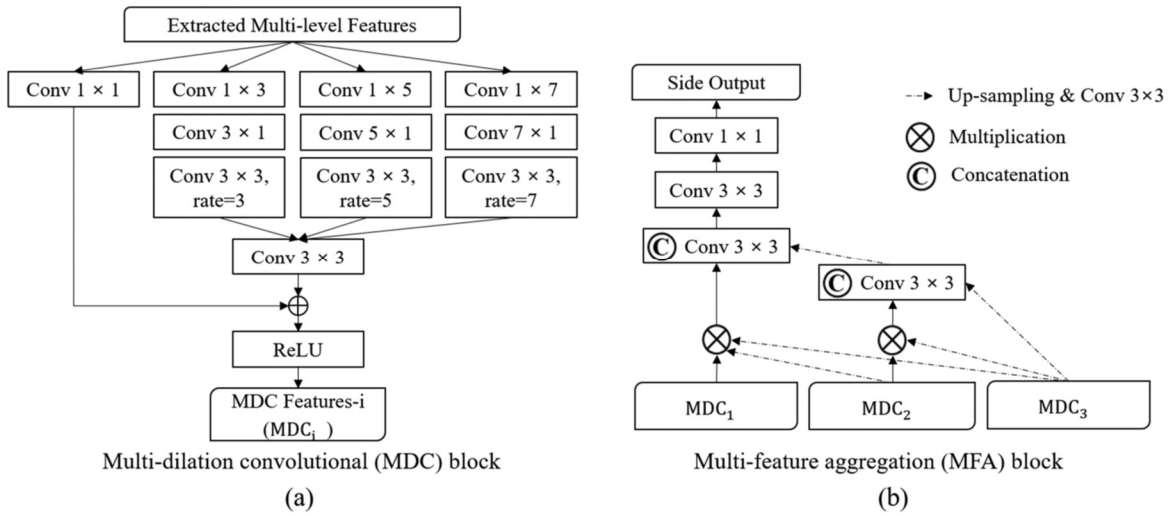
Transformer Net	2025-153	



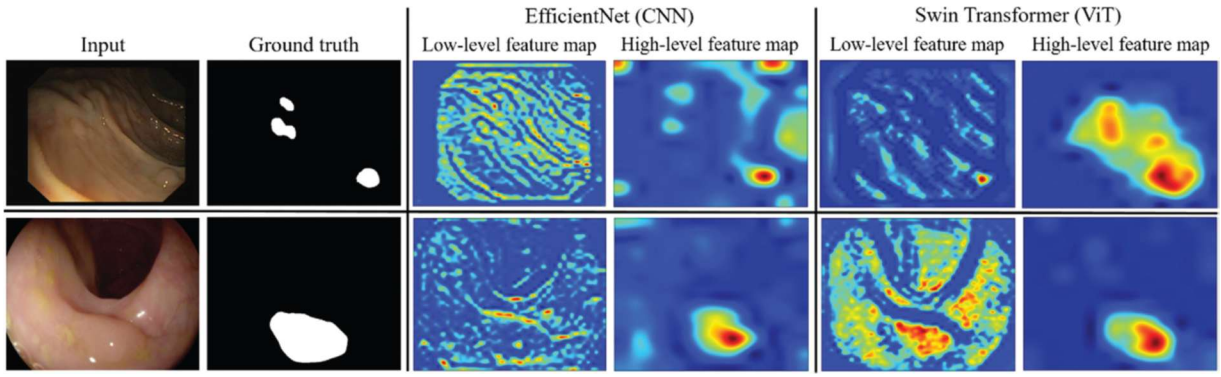
Architecture of the proposed SwinE-Net



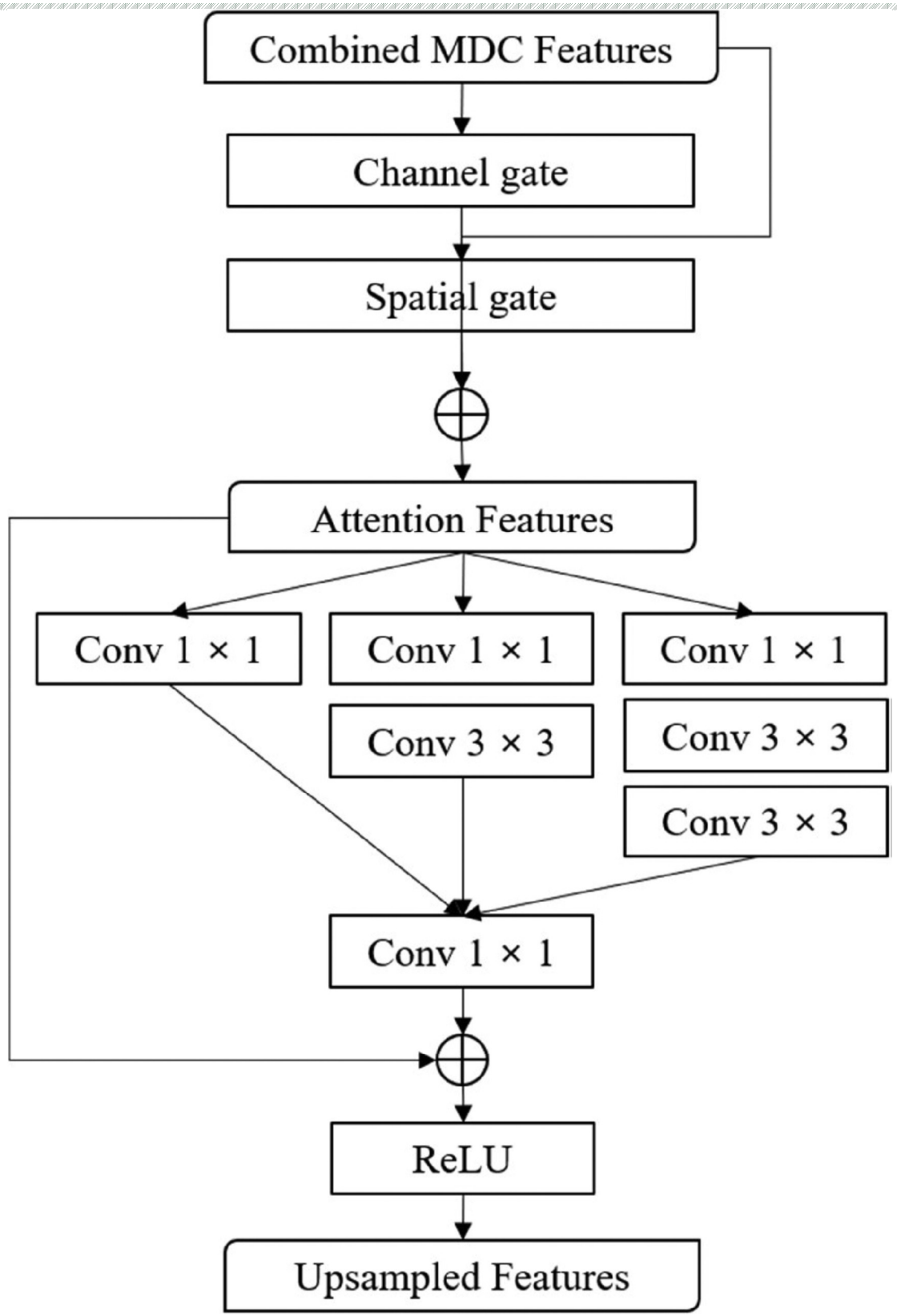
CNN and ViT models: (a) EfficientNet, (b) Swin Transformer, and (c) Swin Transformer block



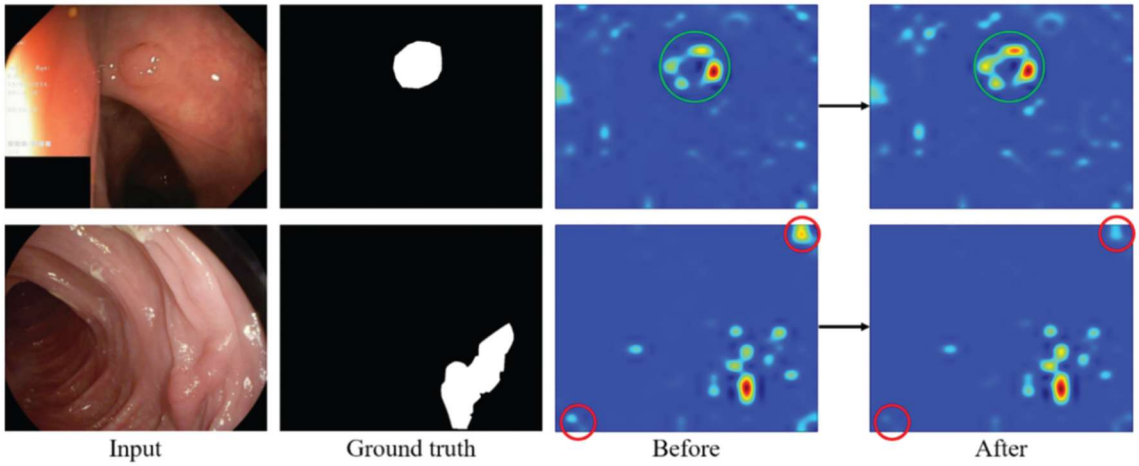
Multidilation convolutional block and multifeature aggregation block



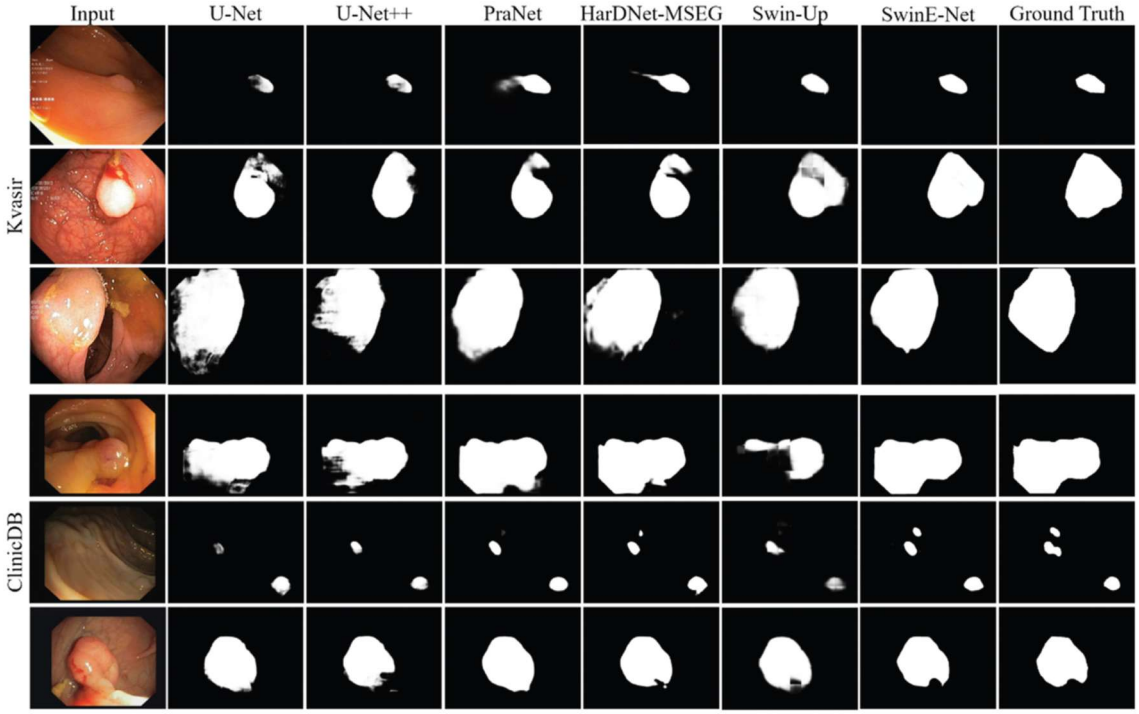
Examples of visualizing low-level and high-level feature maps of EfficientNet and Swin Transformer



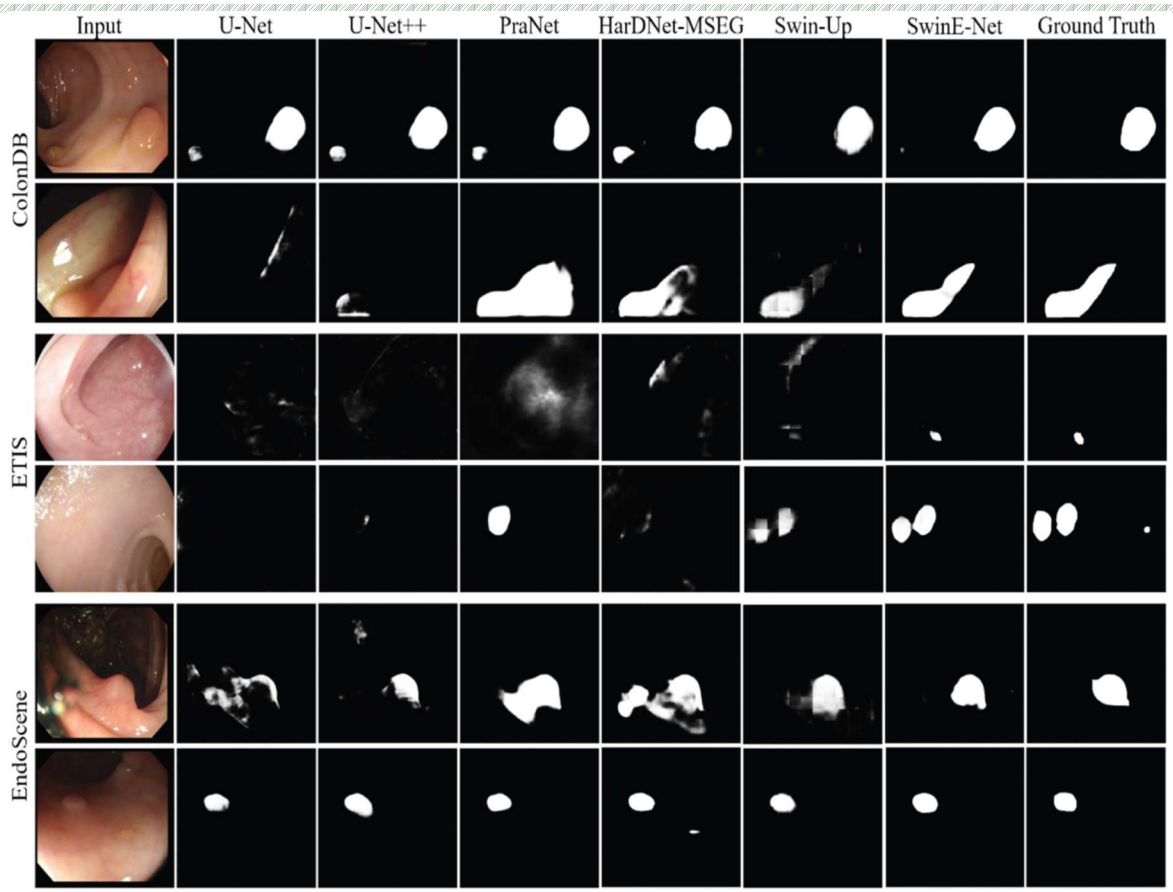
Decoder based on the attentive deconvolutional network



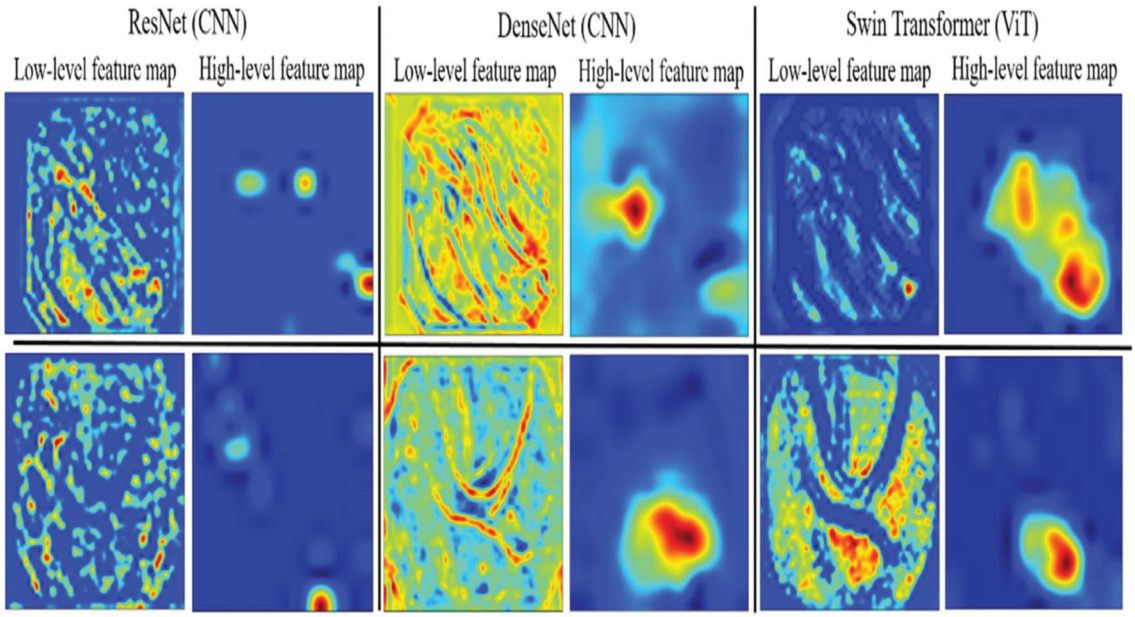
Examples of feature visualization before and after applying the attention module. The green circle of the feature map shows feature refinement, and the red circles show noise reduction



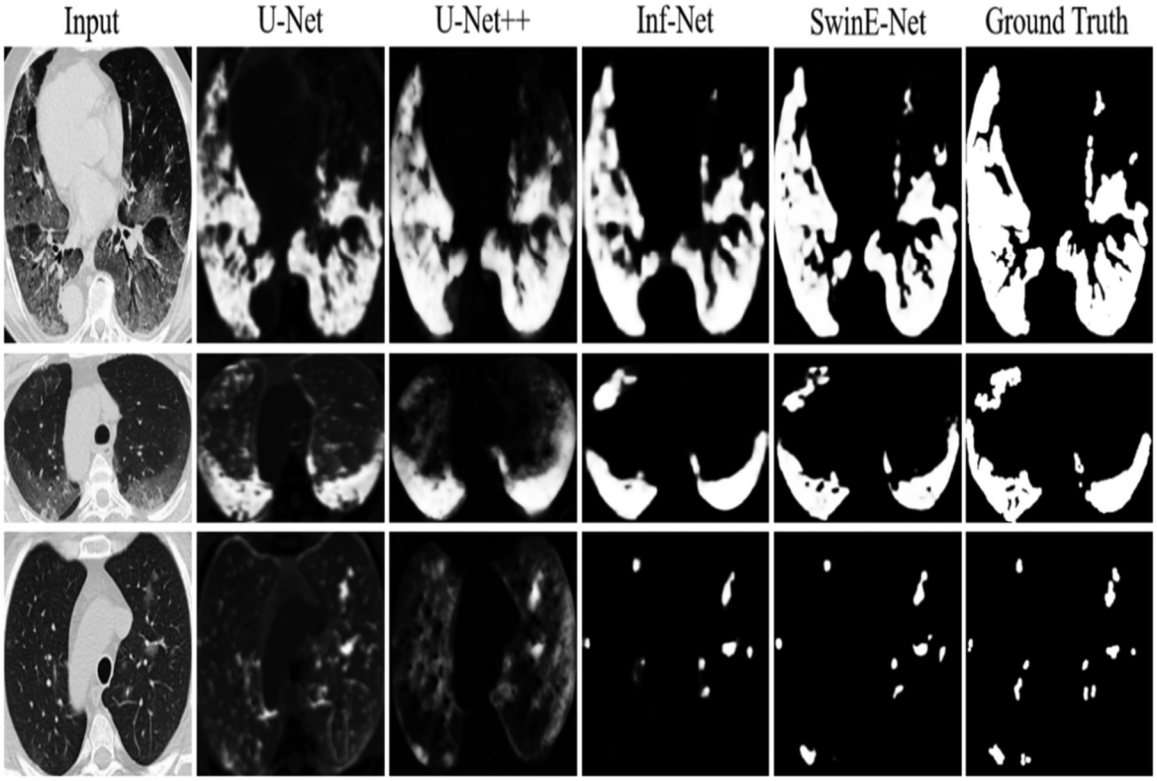
Qualitative evaluation of polyp segmentation results in the seen datasets.



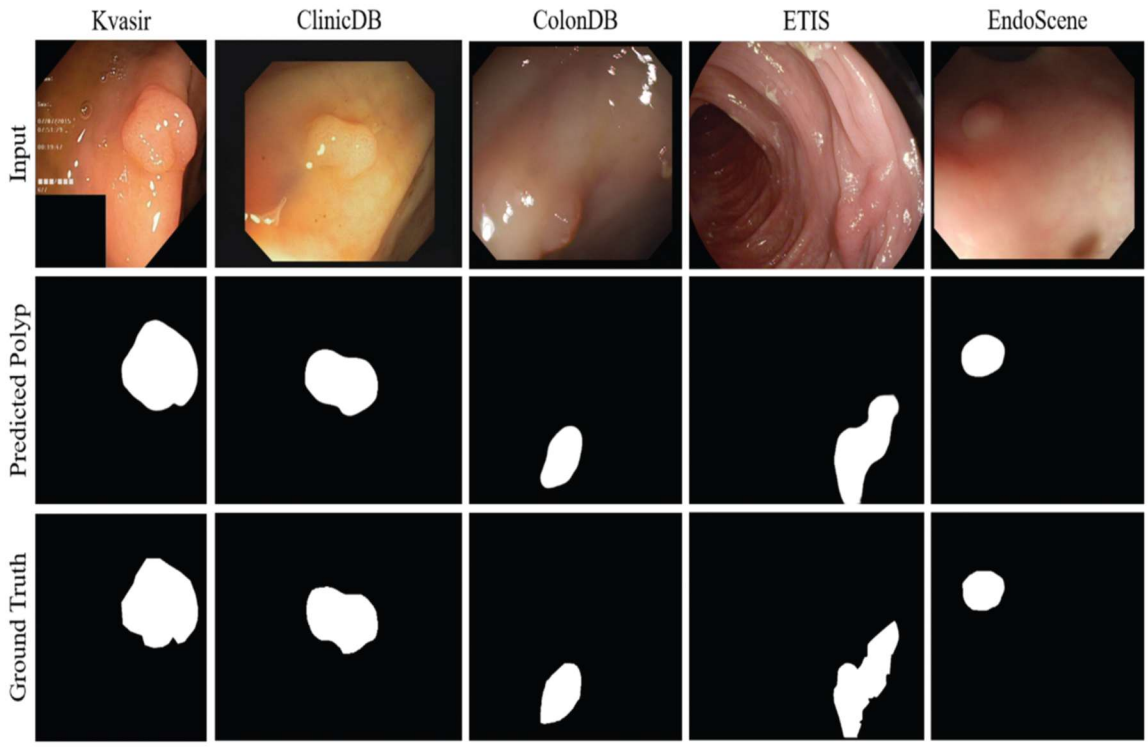
Qualitative evaluation of polyp segmentation results in the unseen datasets



Examples of visualizing low-level and high-level feature maps of ResNet and DenseNet compared with Swin Transformer



Infected region segmentation in the COVID-19 CT segmentation dataset



Examples of the colorectal polyp segmentation using the proposed SwinE-Net.