From 3D CNNs to Clinical AI Agents

The evolution of volumetric medical-image analysis

3D Scan

3D Volume

MAIN RESEARCH QUESTION

How has medical AI evolved from recognising local voxel patterns to reasoning across complete clinical workflows?

Local Feature Learning	Inter-Slice Context	Global Attention	Data Generation	Prompted Multimodal Reasoning	Agent-Assisted Clinical Workflow
CNN	**RNN**	**Transformer**	**GAN | VAE | Diffusion**	**Foundation Model**	**Vision-Language Model**	**AI Agent → Clinician Review**
Convolutions learn local 3D voxel patterns	Recurrent models capture sequential inter-slice context	Self-attention models learn long-range dependencies	Generative models enable synthesis, reconstruction and augmentation	Pretrained models encode broad medical knowledge and context	Multimodal reasoning with natural language prompts and grounded outputs	AI agent supports decisions within the clinical workflow

Model Evolution Timeline

Early 2010s — Mid–2020s and Beyond

- 3 major paradigm families
- 6 core clinical tasks
- CT | MRI | PET | Ultrasound | OCT
- 159 included studies

The field is moving from task-specific prediction toward multimodal, adaptive and clinically supervised AI.

HIGHLIGHTS OF THE PAPER

01 — UNIFIED EVOLUTIONARY TAXONOMY

The review connects classical CNN, RNN and transformer methods with generative, foundation and agentic AI systems.

02 — COMPLETE CLINICAL-TASK VIEW

It analyses detection, segmentation, classification, registration, reconstruction and report generation.

03 — SYSTEMATIC REVIEW METHODOLOGY

From 2,312 initial records, 159 studies were included through a PRISMA-inspired selection process.

04 — CLINICAL-DEPLOYMENT ROADMAP

The review links algorithmic progress to generalisation, interpretability, uncertainty, privacy, safety and human–AI collaboration.

Generalisation | Interpretability | Uncertainty Quantification | Privacy & Security | Safety & Reliability | Human–AI Collaboration

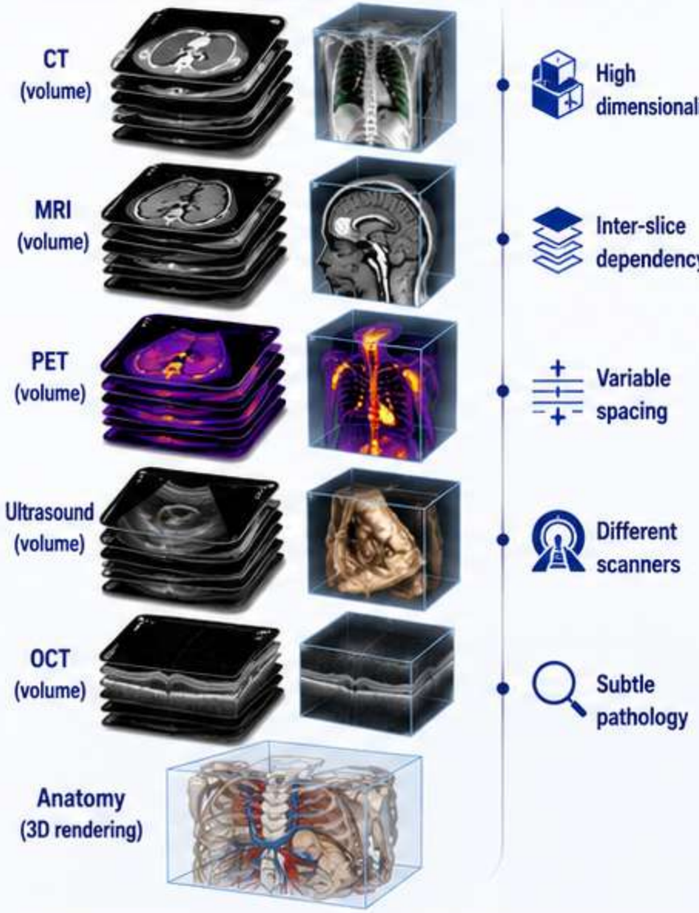
Archives of Computational Methods in Engineering, 2026

DOI: 10.1007/s11831-026-10738-w

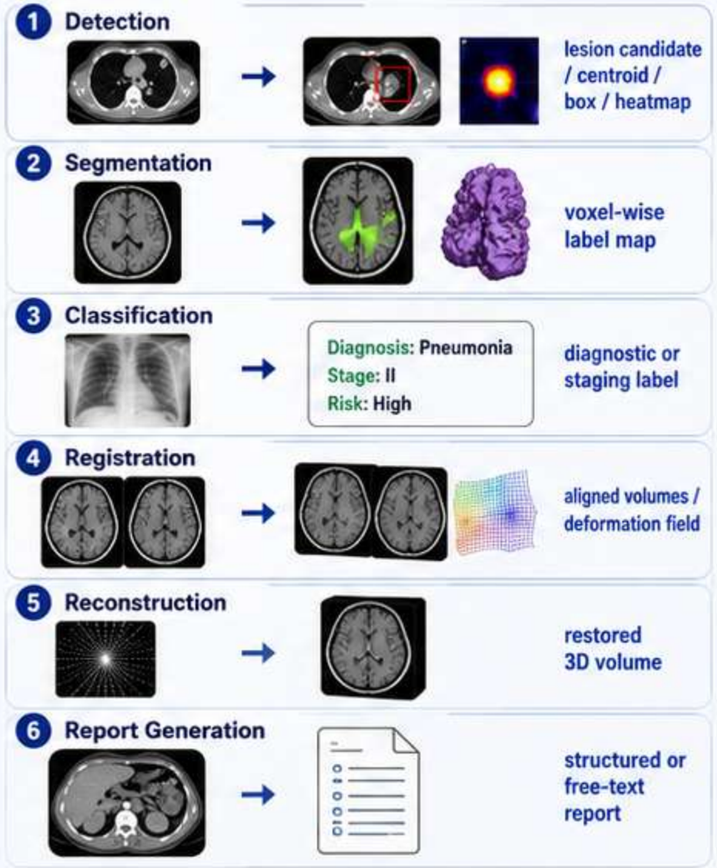
Why 3D Medical Images Are Harder Than Ordinary Images

Large volumes, interdependent slices and clinically variable data

3D Volume



Six Core Clinical Tasks (Input → Output)



Data and Annotation Challenges

Review Methodology (PRISMA-Inspired)

Representative Datasets and Benchmarks

A model score is meaningful only in the context of its dataset, modality, annotations and validation design.

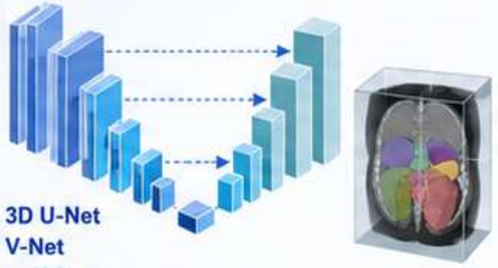
The First Wave: CNNs, RNNs and Transformers

From local voxel patterns to global volumetric context

3D Volume
($H \times W \times D$)

1 CNN

3D Volume → convolution blocks → multiscale features → segmentation / classification



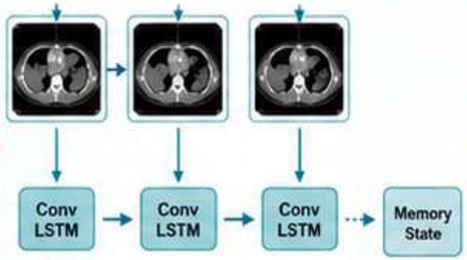
- 3D U-Net
- V-Net
- nnU-Net
- 3D ResNet

STRENGTH
Strong local detail and dense prediction

LIMITATION
Limited long-range context

2 RNN / ConvLSTM

Slice 1 → Slice 2 → Slice 3 ··· memory state



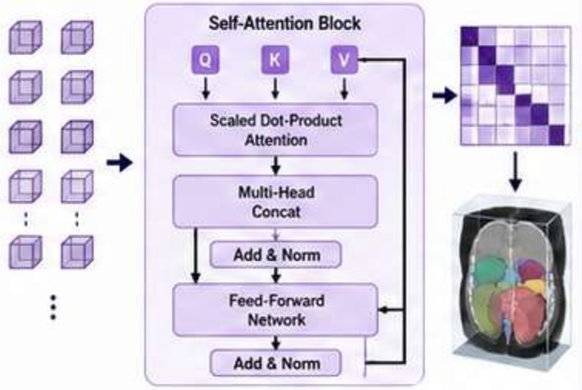
- RNN
- LSTM
- GRU
- ConvLSTM

STRENGTH
Models inter-slice or longitudinal relationships

LIMITATION
Sequential and less parallelisable

3 Transformer

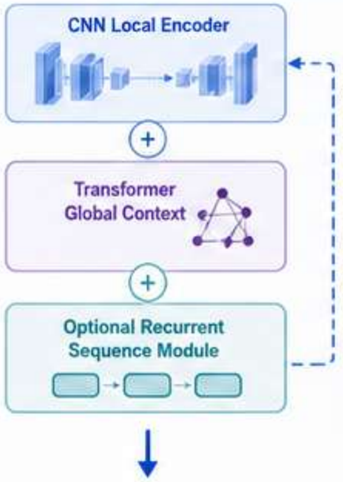
3D patches → Q K V attention → global relationship map → multi-organ output



STRENGTH
Captures global context and distant relationships

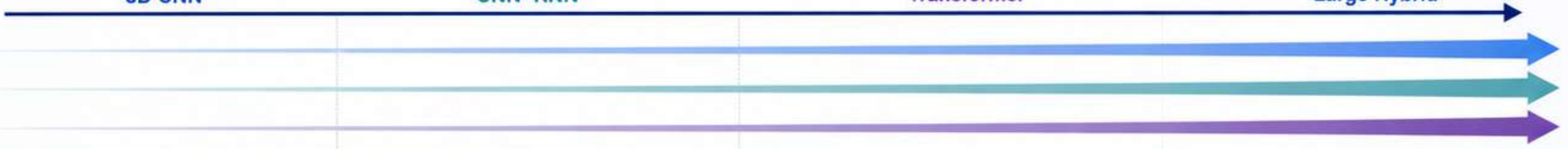
LIMITATION
High data and memory requirements

4 Hybrid Model



Hybrid systems combine complementary inductive biases.

COMPUTATIONAL PROFILE



MODEL-SELECTION RULE

Local boundary detail → CNN | Inter-slice sequence → RNN or ConvLSTM | Global anatomical context → Transformer or hybrid

RESEARCH TAKEAWAY

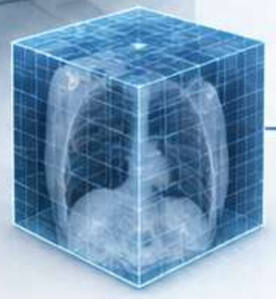
There is no universal winner: model choice depends on the task, data and computational budget.

SLIDE 4

The Second Wave: From Image Generation to Clinical Reasoning

Generative models, foundation models and emerging AI agents

1 GENERATIVE MODELS



Input 3D Volume

VAE → Latent Representation

GAN → Generator → Discriminator

Diffusion → Noise-to-Volume Reconstruction

Applications

- Data augmentation
- Reconstruction
- Denoising
- Synthesis
- Modality translation
- Segmentation refinement

Synthetic realism is not proof of clinical correctness.

Key Risks

- Hallucinated anatomy
- Hallucinated pathology
- Measurement distortion
- Weak physics consistency

2 FOUNDATION MODELS

Text Prompt

"Clinical question + 3D image → VLM / medical foundation model"

- BioMedCLIP
- MedBLIP
- Med-Flamingo
- LLaVA-Med
- UniMed-CLIP
- Med3DVLM
- MedPaLM

Visual Prompt

"Point | Box | Mask + 3D volume → prompted segmentation"

- MedSAM
- SAM-Med3D
- MedSAM2

Many current systems still process volumes as 2D slices.

3 AI-AGENT WORKFLOW

3D CT (volume)

Clinical history

Planning agent → Segmentation tool → Measurement tool → Knowledge retrieval

Report draft → Consistency check → Clinician review

Safety Controls

- Image grounding
- Uncertainty flag
- Audit log
- Human review
- Clinician override

Agentic volumetric imaging remains an emerging field.

The direction of travel is from producing one prediction toward coordinating an explainable clinical workflow.

What the Review Concludes

Better models alone are not enough for trustworthy clinical deployment

KEY CHALLENGES

- Scarce expert annotations
- Class imbalance
- Cross-site domain shift
- Scanner and protocol variation
- 3D memory and computation
- Missing modalities
- Weak external validation
- Poor uncertainty calibration
- Limited interpretability
- Privacy and cybersecurity
- Hallucination
- Regulatory and audit requirements
- Reproducibility of closed models

MODEL-SELECTION MAP

1. CNN

Best suited for: local features; dense segmentation; efficient baselines.

Primary weakness: limited global context.

2. RNN / ConvLSTM

Best suited for: slice sequences; temporal or longitudinal data.

Primary weakness: sequential processing cost.

3. Trunsformer / Hybrid

Best suited for: global context; multi-organ reasoning.

Primary weakness: data and memory demand.

4. Generative Model

Best suited for: augmentation; reconstruction; synthesis.

Primary weakness: hallucination and measurement fidelity.

5. Foundation / VLM

Best suited for: prompting; multimodal transfer; low-label settings.

Primary weakness: domain adaptation and grounding.

6. AI Agent

Best suited for: multi-step workflow coordination; tool use; interactive clinical assistance.

Primary weakness: planning errors, safety and accountability.

Choose the model according to the clinical task, data condition, supervision regime, compute budget and deployment risk.

CLINICAL-TRANSLATION CHAIN

Diverse Data → Robust Model → External Validation → Calibrated Uncertainty → Explainable Evidence → Clinician Oversight → Safe Clinical Integration

Clinician review and override at any point

FUTURE DIRECTIONS

- Self-supervised volume learning
- Masked-volume modelling
- Image–report alignment
- Multimodal clinical fusion
- Domain generalisation
- Test-time adaptation
- Continual learning
- Federated and decentralised learning
- Synthetic rare-condition data
- Lightweight deployment-ready models
- Uncertainty quantification
- Fairness and robustness audits
- Prospective clinical evaluation
- Human-in-the-loop agents
- Governed model updates

The review provides a shared repository of public 3D datasets and recent model implementations.

Volumetric medical AI is evolving from task-specific networks toward multimodal and agentic systems, but clinical impact depends on reliability, interpretability, validation and human supervision.

Understand the volume. Validate the model. Keep the clinician in control.