Study Protocol

Study Title: Research on the Application of Federated Learning in the Construction of Deep Learning Models for Renal Tumor Imaging

NCT number: XXX

Lead Institution:

The First Affiliated Hospital of Guangxi Medical University

Participating Institutions:

Nanning Second People's Hospital

Guilin People's Hospital

Liuzhou People's Hospital

Guigang People's Hospital

Wuzhou Red Cross Hospital

Yulin Red Cross Hospital

Beiliu People's Hospital

Principal Investigator: Jiwen Cheng

July 1, 2026

1. Research Background

1.1 Clinical Challenges of Renal Tumors

Renal tumors rank among the most common urological malignancies worldwide. Widespread adoption of modern imaging modalities has led to a rising detection rate of renal tumors. Nevertheless, a considerable proportion of these incidentally-detected lesions are benign or indolent. Studies indicate that approximately 14%-20% of surgically resected renal tumors are pathologically benign, for which active surveillance or ablation therapy may be more appropriate. Pre-operative non-invasive differentiation between benign and malignant renal lesions remains a major clinical challenge. Although computed tomography (CT) and magnetic resonance imaging (MRI) are routinely used for characterization of renal masses, substantial imaging overlap exists between benign entities such as fat-poor angiomyolipoma and renal cell carcinoma, especially for small-sized tumors. While percutaneous core biopsy yields high diagnostic accuracy, it receives a low-level recommendation (Grade-C evidence), and its clinical benefit for routine management of renal tumors remains unclear. Therefore, a more accurate and non-invasive auxiliary method for differentiating benign and malignant renal tumors is urgently needed.

1.2 Application Status of Artificial Intelligence in Renal Tumor Diagnosis

In recent years, deep learning has demonstrated remarkable potential in medical image analysis. Advances in radiomics have provided novel approaches for precise tumor characterization. Machine-learning models built on CT images have achieved promising performance (AUC 0.80-0.90) for distinguishing benign from malignant renal tumors. However, most existing studies lack large-scale multi-center clinical validation. Reported effectiveness and safety are constrained by limited sample sizes, homogeneous pathological cohorts, selection bias, and insufficient external validation. Major limitations include:

Lack of model interpretability, which hinders clinical adoption by physicians;

Absence of effective quality-control and rejection mechanisms for low-quality and non-standard imaging data;

Incomplete evidence chains with non-traceable diagnostic outputs;

Lack of systematic traceable architecture for engineering-oriented clinical deployment.

To address these gaps, we developed the CT-based Renal Tumor Intelligent Detection System (CascadeDiagnose Renal Tumor CT), an auxiliary analytical system designed for CT-based renal-tumor AI competitions and engineering demonstration. Built upon a cascaded deep-learning architecture, the system performs fully-automated analysis on multi-phase contrast-enhanced CT images (non-contrast, corticomedullary, nephrographic, and excretory phases). Its core functions include image-quality review, lesion detection and segmentation, benign-malignant differentiation, and risk-stratified reporting with traceable and explainable outputs.

Compared with existing systems, the core innovation of CascadeDiagnose Renal Tumor CT lies in its evidence-chain-first design philosophy. Both lesion identification and classification deliver traceable evidence including sequence information, slice location, spatial coordinates, confidence scores, heatmaps, and segmentation masks. Equipped with an integrated image-quality-review module, the system outputs indeterminate or rejected results for low-quality, non-standard, suspected-tampered, cross-patient mixed, or non-interpretable scans instead of forcing a diagnostic conclusion. Full-lifecycle tracking covers all data versions, model iterations, threshold configurations, inference logs, and manual-review records.

1.3 Layout of Distributed Trustworthy Medical Computing Network

This study is implemented on a Guangxi-wide distributed trustworthy medical-sharing computing network with the core principle of “move computing power instead of raw data”. Independent edge-computing nodes are deployed locally at the six participating hospitals. Original patient CT images, electronic medical records, and pathological data remain stored within local hospital servers without external export. Each edge node connects to the regional computing-power convergence node via national-cryptography-encrypted dedicated lines. A U-shaped split federated-learning architecture enables cross-hospital collaborative model training, joint inference, and multi-center coordinated evaluation. This establishes a distributed trustworthy

medical-computing network characterized by “raw data stays in-hospital while models collaborate across sites”. While fully safeguarding institutional data sovereignty and patient-privacy compliance, the network supports end-to-end multi-center clinical-research collaboration and resolves challenges related to data-sharing compliance, information security, and practical engineering deployment.

2. Research Objectives

2.1 Primary Objectives

Effectiveness Evaluation: To assess the diagnostic accuracy of CascadeDiagnose Renal Tumor CT for differentiating benign versus malignant renal tumors. Primary metrics include area under the receiver-operating-characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Safety Evaluation: To evaluate clinical safety in real-world use, including false-negative rate (malignancy missed diagnosis), false-positive rate (benign lesions misclassified as malignant), system failure rate, and potential risks to clinical decision-making.

2.2 Secondary Objectives

a) Compare diagnostic performance of CascadeDiagnose Renal Tumor CT against radiologists with different levels of seniority for benign-malignant differentiation of renal tumors.
b) Evaluate system performance for small renal tumors (≤ 3 cm).
c) Assess generalization and robustness across different CT scanner vendors, scanning protocols, and clinical centers.
d) Quantify changes in diagnostic accuracy and reading efficiency for junior radiologists with AI assistance.
e) Evaluate the potential clinical value of the system for reducing unnecessary renal surgeries.

f) Validate the reliability of built-in image-quality review, evidence-chain tracing, and case-rejection mechanisms.

3. Study Design

3.1 Study Type

This is a multi-center diagnostic clinical trial combining retrospective and prospective cohorts. It complies with Guidelines for Clinical Evaluation and Registration Review of Artificial-Intelligence-Assisted Detection Medical Devices (Software) (NMPA Announcement No. 38, 2023) and the draft guidance document for AI-assisted-diagnosis medical devices released by the National Medical Products Administration. This AI-assisted diagnostic software is classified as Class-III medical device with classification code 21-04-02.

A multi-reader multi-case cross-over (MRMC) design is adopted, which represents the preferred trial paradigm for evaluating medical-imaging artificial-intelligence products.

All multi-center data coordination, collaborative model training, and cross-site AI inference are executed on the “move computing power instead of raw data” distributed trustworthy-computing network. Original imaging and clinical records remain stored and processed locally within each hospital. Only encrypted intermediate features and gradient parameters are transmitted to the convergence node over dedicated encrypted lines. No raw patient-privacy-related data is transferred externally, satisfying medical-data-compliance and privacy-protection requirements.

3.2 Study Phases

Phase 1 (Retrospective Validation): Retrospective diagnostic-performance validation using archived multi-center CT datasets with definitive pathological reference standards.

Phase 2 (Prospective Validation): Real-world performance and safety assessment using prospectively enrolled patients with multi-phase contrast-enhanced CT scans collected under routine clinical workflows.

3.3 Study Population

At least 3000 eligible patients will be enrolled (approximately 2600 retrospective subjects and 400 prospective subjects), with reference to sample sizes reported in published literature.

4. Participating Centers

This multi-center trial involves six tertiary-A hospitals in Guangxi.

Table 1 Participating Institutions

No	Center Name	Role
1	The First Affiliated Hospital of Guangxi Medical University	Lead Center
2	Nanning Second People's Hospital	Participating Center
3	Guilin People's Hospital	Participating Center
4	Liuzhou People's Hospital	Participating Center
5	Guigang People's Hospital	Participating Center
6	Wuzhou Red Cross Hospital	Participating Center

Computing-resource Deployment Note: Independent edge-computing nodes are deployed in all six hospitals and connected to the regional computing-power convergence node, forming a "1 convergence node + 6 edge nodes" distributed network. The core principle "move computing power instead of raw data" is strictly enforced: all local-center data is managed within hospital boundaries; cross-institutional collaboration relies solely on encrypted intermediate computational outputs.

5. Eligibility Criteria

5.1 Inclusion Criteria

a) Patients aged ≥ 18 years who underwent multi-phase contrast-enhanced renal CT.

b) CT examinations contain two or more of the following phases: non-contrast, corticomedullary, nephrographic, and excretory phases.

c) CT demonstrates a solitary renal-parenchymal mass lesion.

d) Definitive post-surgical histopathological diagnosis available, classified according to WHO and International Society of Urological Pathology (ISUP) grading systems.

e) Complete clinical records including age, sex, clinical manifestations, and laboratory findings.

f) Informed consent obtained (written informed consent required for prospective-phase participants).

5.2 Exclusion Criteria

a) Severe motion or metal artifacts substantially impairing image interpretation, confirmed as unqualified by the built-in image-quality-review module.

b) Non-definitive postoperative pathology; lesions representing macroscopic-fat-containing angiomyolipoma or simple cysts (Bosniak Category I).

c) Incomplete CT phase acquisition or incomplete lesion coverage.

d) Prior history of prostate cancer or other malignant tumors.

e) Pre-operative interventions including renal-tumor biopsy, ablation, targeted therapy, or radiotherapy.

f) Multiple or bilateral renal masses.

g) Critical missing baseline clinical or pathological variables.

6. Detailed Description of CascadeDiagnose Renal Tumor CT

6.1 System Architecture

The system adopts a cascaded deep-learning framework composed of six core modules:

Image-Quality Review Module: Automated assessment of artifact severity, scan completeness, slice continuity, and spatial resolution. Scans failing quality thresholds return “non-interpretable (indeterminate)” or “rejected” outputs with documented root causes. This module also validates dataset integrity and flags suspected data tampering or cross-patient mixing artifacts.

Kidney and Tumor Segmentation Module: Built on 3D U-Net architecture for automatic segmentation of renal parenchyma and renal tumors from corticomedullary-phase CT volumes. Segmentation outputs undergo automatic quality control; segmentation failures are flagged for manual human review.

Multi-Phase Feature-Extraction Module: Multi-view convolutional neural networks extract and fuse imaging features across registered CT phases (non-contrast, corticomedullary, nephrographic, excretory). A continuous benign-malignant probability score ranging from 0 to 1 is generated per lesion.

Explainable-Output Module: Structured evidence-chain reports are automatically generated for each case, containing:

a. Sequence-and-slice metadata: exact CT sequence and slice index supporting the diagnostic conclusion (e.g., corticomedullary phase, slice 42).

b. Spatial localization: three-dimensional coordinates and anatomical location of the lesion within the kidney (upper/mid/lower pole; ventral/dorsal).

c. Confidence score: model-derived diagnostic confidence (0-100%).

d. Grad-CAM heatmap: attention-weighted overlay visualizing image regions driving model decisions.

e. Segmentation mask: precise pixel-wise boundary delineation of the target lesion.

Risk-Stratification Module: Provides imaging-based aggressiveness-risk recommendations for lesions classified as malignant.

Distributed Cross-Hospital Collaborative Computing-Support Module:

Multi-center joint training and synchronized inference run on the distributed trustworthy-medical-computing network using a U-shaped split federated-learning framework:

a. Edge side (local hospital premises): Lightweight bottom sub-models are deployed locally. Raw CT images, electronic health records, and pathology data stay on-site. Image quality control, segmentation, and feature extraction are performed locally without exporting raw patient data.

b. Transmission channel: Only desensitized and SM4-encrypted intermediate features and gradient parameters are transmitted over dedicated encrypted lines to the convergence node. No identifiable original patient data leaves the hospital.

c. Convergence-node side: Computationally-intensive top sub-models execute multi-center feature aggregation and global-model updating. Updated lightweight sub-model parameters are encrypted and sent back to each local edge node.

d. Inference workflow: AI inference is executed independently at each participating hospital. Only desensitized aggregated statistical metrics are submitted to the convergence node for multi-center performance analysis. The architecture realizes “raw data remains in-hospital while models collaborate across institutions”, reconciling multi-center-research needs with healthcare-data-privacy regulations.

6.2 Full-Chain Traceability Design

Comprehensive audit trails are implemented:

a. Data-version traceability: Record source, acquisition timestamp, de-identification version identifier, and cryptographic hash for every input dataset.

b. Model-version traceability: Log model-release version, training-dataset iteration, and parameter hash value for each inference run.

c. Threshold-version traceability: Document decision-threshold versions and historical modification logs.

d. Inference-log traceability: Preserve complete inference logs including input metadata, module-level outputs, timestamps, and exception events.

e. Manual-review traceability: Capture operator identity, timestamp, review comments, and revision history for every manual-override event.

Additional cross-node collaborative logs generated by the distributed computing network record timestamps of encrypted intermediate-data uploads, global-model-update events, and dedicated-line transmission receipts to enable end-to-end auditability of multi-party collaborative computation.

6.3 Rejection and Indeterminate-Result Mechanism

The system outputs indeterminate or rejected conclusions (instead of forced diagnoses) under the following circumstances:

Substandard image quality (severe artifacts, low signal-to-noise ratio, insufficient resolution).

Incomplete scanning phases (missing required contrast phases).

Poor lesion conspicuity (lesion < 5 mm, indistinct margins, prominent partial-volume effect).

Data anomalies (suspected tampering, cross-patient contamination, file corruption).

Model confidence falls below configurable preset threshold (default < 70%).

7. Evaluation Metrics

7.1 Effectiveness Metrics

Primary Effectiveness Metrics

a. Area under the receiver-operating-characteristic curve (AUC): overall discriminative performance separating benign and malignant renal tumors.

b. Sensitivity: capacity for correctly identifying malignant renal tumors.

c. Specificity: capacity for correctly identifying benign renal tumors.

d. Positive predictive value (PPV) and negative predictive value (NPV).

e. Sensitivity achieved at fixed specificity of 90%: detection performance under high-specificity constraints, consistent with published benchmarks.

Secondary Effectiveness Metrics

a. Head-to-head diagnostic-performance comparison: AUC, sensitivity, and specificity comparisons between the AI system and physicians of varying seniority (radiologists and urologists).

b. Changes in diagnostic performance under AI assistance: differences in diagnostic accuracy and reading time for junior radiologists before versus after AI support.

c. Subgroup analyses:

Tumor-size stratification: ≤ 3 cm (including sub-centimeter lesions) versus > 3 cm;

Pathological-subtype stratification: clear-cell renal-cell carcinoma, papillary renal-cell carcinoma, chromophobe renal-cell carcinoma, fat-poor angiomyolipoma, oncocytoma and other entities;

Stratification by CT scanner vendor and acquisition parameters;

Stratification by participating clinical center.

d. Decision-curve analysis (DCA): clinical net benefit across varying decision thresholds.

e. Rejection-rate and indeterminate-case rate: performance of the image-quality-review and rejection workflow.

f. Evidence-chain completeness: integrity and accuracy of generated evidence-chain reports in clinical validation.

g. Distributed-network operational metrics: local-edge-node processing completion rate, encrypted intermediate-data transmission success rate, cross-site model-synchronization latency, and network-failure incidence. These metrics quantify operational performance of the “move computing power instead of raw data” architecture for multi-center research.

7.2 Safety Metrics

a. False-negative rate: proportion of malignant renal tumors misclassified as benign; a core safety endpoint because false negatives may cause delayed oncological treatment.

b. False-positive rate: proportion of benign renal tumors misclassified as malignant; false positives risk triggering unnecessary surgery and procedural complications.

c. Adverse events (AE): all adverse events potentially related to system usage will be documented.

d. System failure rate: frequency of technical malfunctions including startup failure, inference time-out, software crash, or invalid outputs.

e. Clinical-decision impact assessment: evaluate real-world influence of system outputs on physician decision-making and risk for clinical misguidance.

f. Distributed-network-security risk indicators: count alerts for suspected cross-node data leakage, unauthorized access attempts, and detected-data-tampering events, validating privacy-protection performance of the “move computing power instead of raw data” framework in multi-center research.

8. Data Collection and Pre-Processing

8.1 CT Image Acquisition

All participating centers adhere to standardized multi-phase renal-contrast-enhanced CT acquisition protocols adapted from Dai et al. Recommended scanning parameters:

a. Slice thickness: ≤ 5 mm; thin-slice reconstruction (1-3 mm) preferred.

b. Phases: non-contrast phase; corticomedullary phase (30-40 s post-contrast-media injection); nephrographic phase (90-120 s post-injection); excretory phase (3-5 min post-injection).

c. Reconstruction kernel: standard soft-tissue algorithm.

d. Matrix size: 512×512.

Distributed-computing-acquisition specification: all CT data are stored in DICOM format locally on hospital-based edge-node servers. No raw images are exported to portable media or transmitted cross-institutionally. Only locally-desensitized, quality-controlled, feature-extracted encrypted intermediate outputs participate in cross-site collaborative workflows.

8.2 Image Registration and Pre-Processing

To compensate for respiratory-motion-induced positional shifts across contrast phases, the corticomedullary phase serves as the reference volume. Non-contrast, nephrographic, and excretory-phase volumes undergo rigid registration against the reference using 3D Slicer (version 5.10.0). Automatic image-quality assessment is performed by the built-in review module; scans flagged as poor quality receive secondary manual review by senior on-site radiologists.

Computational-resource allocation: all registration and pre-processing operations run locally on hospital edge nodes without consuming resources of the central convergence node, implementing the “local data processed locally” principle.

8.3 Histopathological Gold Standard

Surgical histopathology constitutes the reference gold standard. Pathological specimens from each site are interpreted by local-center pathologists following WHO-ISUP classification criteria. Challenging ambiguous cases are submitted for secondary review by the pathology department of the lead center.

Benign entities include fat-poor angiomyolipoma, oncocytoma, complex renal cysts, and other benign lesions.

Malignant entities include clear-cell renal-cell carcinoma, papillary renal-cell carcinoma, chromophobe renal-cell carcinoma, and other malignant renal neoplasms.

8.4 Clinical-Data Collection

Collected clinical variables: age, sex, clinical symptoms, prior medical history, laboratory tests (complete blood count, renal-and-hepatic biochemistry etc.), maximum tumor diameter measured on CT, tumor location (left/right kidney;

upper/mid/lower pole), and CT imaging features (solid/cystic appearance, margin characteristics, enhancement patterns).

9. Observer Study

9.1 Panel of Readers

Table 2 Composition and Qualifications of Image Readers

Category	Number of Readers	Experience Requirement
Senior radiologists	2 名	≥ 10 years genitourinary-imaging experience
Mid-/junior-level radiologists	2 名	3-10 years radiological practice
Urologists	2 名	≥ 5 years renal-tumor clinical-management experience

9.2 Reading Workflow

a) Training phase: A genitourinary-imaging expert (lead center, > 5 000 renal-tumor-case reading experience) delivers standardized training for all readers, covering imaging-feature interpretation for 50 illustrative representative cases and standardized scoring frameworks (e.g., likelihood score for clear-cell carcinoma, Bosniak classification).
b) Independent-reading phase: Each reader interprets every CT dataset in isolation and blinded to clinical and pathological reference data, assigning a continuous 0-100% probability estimate of benign histology.
c) AI-assisted-reading phase: Readers re-evaluate identical cases with outputs from CascadeDiagnose Renal Tumor CT visible. Diagnostic conclusions and reading time are recorded.
d) Computing-workflow note: Reading terminals connect exclusively to the local hospital edge node. AI inference executes on-site. Under the “move computing power instead of raw data” architecture, readers never access raw case data originating from other hospitals. Only desensitized aggregated summary statistics are forwarded to the multi-center analysis platform.

10. Sample-Size Calculation

Sample-size assumptions are derived from published clinical-AI studies:

a. Expected system AUC = 0.85

b. Minimal clinically meaningful difference = 0.05

c. Two-sided $\alpha = 0.05$

d. Statistical power $(1-\beta) = 0.80$

e. Benign-to-malignant case ratio 1:1, reflecting real-world surgical-cohort distributions.

Calculations indicate approximately 1500 benign and 1500 malignant subjects are required. Accounting for multi-center heterogeneity and anticipated protocol violations/exclusions, the target total sample size is no fewer than 3000 participants (2600 retrospective, 400 prospective).

11. Statistical Analysis

11.1 Analysis Datasets

a) Full Analysis Set (FAS): All enrolled participants for baseline-characteristic descriptive statistics.

b) Per-Protocol Set (PPS): Subjects completing the full study workflow without major-protocol deviations; primary-efficacy analyses rely on this dataset.

c) Safety Set (SS): All patients receiving system assessment for safety-endpoint analysis.

d) Rejection Set (RS): Cases marked indeterminate or rejected by the system; clinical features and outcomes for these patients will be analyzed separately.

e) Distributed-computing-dataset specification: Desensitized feature subsets are generated locally on each edge node. Only encrypted aggregated summary statistics are transmitted to the convergence node for pooled analysis. Individual-level raw datasets are permanently retained within each participating hospital and are not centrally pooled.

11.2 Statistical Methods

a) Descriptive statistics: Continuous variables reported as mean ± standard-deviation or median (inter-quartile range). Categorical variables presented as frequency (percentage).

b) Diagnostic-performance comparisons: DeLong test for inter-group AUC comparisons. Analyses performed in R (version 4.5.1) and SPSS (version 26.0).
c) Sensitivity-and-specificity pairwise comparisons: McNemar's test.
d) Multi-center heterogeneity assessment: Random-effects modelling or Cochran's Q-statistic.
e) Subgroup analyses: Logistic-regression models incorporating interaction terms.
f) Rejected-case analysis: Descriptive characterization of rejection-reason distribution and associations with clinical outcomes.
g) All hypothesis tests are two-tailed. A P-value < 0.05 denotes statistical significance.

12. Quality Management and Traceability

12.1 Data-Quality Management

Standard operating procedures (SOPs) titled CT-Image-Acquisition and Data-Management Protocol for Renal-Tumor Research are enforced uniformly at all centers.

Local-center data managers oversee completeness, accuracy, and image-quality auditing.

De-identification is applied to all patient datasets prior to system ingestion in compliance with the Personal Information Protection Law of the People's Republic of China.

Specialized distributed-network quality control: Cross-node transmission integrity checks and anti-tampering receipts are implemented. Transfer of raw imaging or medical-record data across edge nodes is strictly prohibited. Periodic access-permission audits are performed on local edge-node storage to enforce the "move computing power instead of raw data" mandate.

12.2 System-Traceability Governance

Version control: Unique version identifiers, release timestamps, and change logs are maintained for datasets, AI-model iterations, and decision-threshold configurations.

Log management: The system automatically logs every inference workflow including inputs, outputs, timestamps, and exception events.

Review-audit trails: Manual-review events permanently store operator identity, timestamp, and review comments.

13. Ethical Considerations

a) Ethical-committee approval: The research protocol must receive ethical-review-committee clearance from each participating institution. The Ethics Committee of The First Affiliated Hospital of Guangxi Medical University serves as the primary reviewing body.

b) Informed consent: Waiver of written informed consent for retrospective data may be granted per local-ethical-committee determinations. Written informed consent must be obtained for all prospective participants.

c) Data protection: Patient datasets undergo anonymization complying with the Personal Information Protection Law and the Declaration of Helsinki. Built-in system safeguards mitigate data-leakage and unauthorized-access risks.

d) Clinical-trial registration: Trial registration shall be completed at the Chinese Clinical Trial Registry (ChiCTR) prior to subject enrollment.

e) Distributed-trustworthy-computing-network-specific ethical compliance: The multi-center collaborative framework adheres strictly to the “move computing power instead of raw data” principle:

① Data sovereignty: Original CT, electronic-medical-record, and pathology data remain stored and processed locally in-hospital. Each institution retains full sovereignty over its patient datasets without raw-data export or centralized aggregation.

② Privacy protection: Cross-institutional transmission is limited to desensitized, cryptographically-secured intermediate features and gradient parameters. The convergence node cannot reconstruct original patient images or personally-identifiable information, achieving the “data is usable but not observable” paradigm.

③ Transparent traceability: Tamper-resistant audit trails capture every cross-node computing-resource scheduling and encrypted-data-transfer event. Ethics committees may inspect these logs to verify data-use compliance.

④ Risk isolation: Physical and logical network isolation separates edge-node infrastructure from the convergence node. The convergence node cannot access local raw datasets, architecturally mitigating patient-privacy-breach risks inherent to conventional multi-center clinical research.

14. Study Timeline

Total study duration: August 2026 – July 2028.

Subject-recruitment windows: retrospective-phase recruitment October 2026 – December 2027; prospective-phase enrollment January 2027 – December 2027.

Table 3 Study-Phase Timeline

Phase	Time Window	Key Deliverables
Preparation	2026-08-01 – 2026-09-30	Protocol finalization, ethical approval, clinical-trial registration, system deployment, investigator training
Retrospective-data collection	2026-10-01 – 2027-12-30	Retrospective-case curation, quality-control, retrospective-phase system validation
Prospective enrollment	2027-01-01 – 2027-12-30	Prospective-subject enrolment, clinical-and-imaging-data capture, prospective-phase system validation
Statistical analysis	2028-01-01 – 2028-03-30	Data-lock, statistical computation, result interpretation

Study conclusion	2028-04-01 – 2028-07-30	Final-study-report composition, manuscript submission and publication, knowledge-translation and dissemination

15. Expected Outcomes

a) Comprehensive multi-center performance characterization of CascadeDiagnose Renal Tumor CT for differentiating benign-malignant small renal masses ≤ 3 cm, including sub-centimeter (≤ 1 cm) lesions.

b) Empirical validation of generalization and robustness across varying CT scanner hardware, acquisition parameters, and geographically-diverse clinical centers.

c) Empirical validation of image-quality screening, evidence-chain tracing, and case-rejection mechanisms.

d) Quantification of clinical-decision-support utility, especially potential reduction of unnecessary renal-surgical procedures.

e) Generation of high-quality clinical-evidence packages supporting medical-device-regulatory submission.

f) Publication of 2-3 high-impact peer-reviewed clinical-research articles.

g) Real-world demonstration of a “move computing power instead of raw data” distributed trustworthy-medical-computing network for multi-center medical-device clinical trials. A replicable compliance-oriented implementation specification for healthcare-AI multi-center collaborative research will be produced to serve as engineering-and-ethics reference for analogous medical-device-evaluation projects.

16. References

[1] Dai C, Xiong Y, Zhu P, et al. Deep Learning Assessment of Small Renal Masses at Contrast-enhanced Multiphase CT. Radiology. 2024;311(2):e232178.

[2] Center for Medical Device Evaluation, National Medical Products Administration. Guidelines for Clinical Evaluation and Registration Review of Artificial-Intelligence-Assisted Detection Medical Devices (Software).

[3] Center for Medical Device Evaluation, National Medical Products Administration. Guidelines for Clinical Evaluation and Registration Review of Artificial-Intelligence-Assisted-Diagnosis Medical Devices (Draft).

[4] Çi çek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2016: 424-432.

[5] Heller N, Isensee F, Maier-Hein KH, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. Med Image Anal. 2021;67:101821.

[6] SPIRIT-AI Extension: Guidelines for clinical trial protocols for interventions involving artificial intelligence. Equator Network, 2025.

[7] DECIDE-AI: Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence, 2025.

[8] Srigley JR, Delahunt B, Eble JN, et al. The International Society of Urological Pathology (ISUP) Vancouver Classification of Renal Neoplasia. Am J Surg Pathol. 2013;37(10):1469-1489.