Research protocol

Project Title: Generative model-based system for survival prediction and treatment decision-making in unresectable hepatocellular carcinoma receiving TACE combined with immunotherapy plus targeted therapy

Departments: Interventional and Vascular Surgery

Unit address/zip code: No. 87, Dingjiaqiao, Gulou District, Nanjing City, Jiangsu Province / 210009

Project Leader: Lu Jian

Phone/E-mail： 15850654644/lujian43307131@126.com

Competent authority: Interventional Vascular Surgery

Version number: V1.0

Version date: May 1, 2025

Zhongda Hospital Affiliated to Southeast University

Research project design document

Project justification

1. Current status and challenges in prognosis prediction of unresectable liver cancer

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related deaths worldwide, and the treatment of patients with intermediate and advanced stages is particularly challenging. The burden of liver cancer in China is heavy, accounting for about half of the world's new cases and deaths every year, and about 70% of patients have lost the opportunity for surgery when they are diagnosed. In recent years, transcatheter arterial chemoembolization (TACE) combined with targeted and immunotherapy has become a key treatment modality for unresectable HCC. This strategy exerts dual anti-tumor effects through a synergistic mechanism: TACE induces avascular necrosis and releases high concentrations of chemotherapeutic drugs by local embolization of the tumor-supplying arteries; Immune checkpoint inhibitors such as PD-1/PD-L1 activate systemic anti-tumor immune responses by reversing the immunosuppressive state of the tumor microenvironment. Two phase III randomized controlled trials, EMERALD-1 and LEAP-012, showed that TACE in combination with durvalumab and bevacizumab or pembrolizumab and lenvatinib significantly extended median progression-free survival (PFS) to 15.0 months and 14.6 months, significantly better than 8.2 months and 10.0 months with conventional TACE therapy.

However, there is significant clinical heterogeneity in this combination, with studies showing that only about 30% to 40% of patients achieve a significant response, and nearly 60% of patients fail to achieve substantial survival benefit. This difference in efficacy is closely related to tumor molecular heterogeneity, immune microenvironment characteristics, and host liver function status. Imaging evaluation is the main method of efficacy evaluation, but the widely used clinical efficacy evaluation criteria for solid tumors (RECIST v1.1) are mainly based on tumor diameter changes, which has limitations in evaluating the dynamic efficacy of TACE combined with target-immune therapy. This criterion is a static "time slice" analysis, which is difficult to capture the dynamic evolution of the tumor microenvironment during treatment.

Radiomics and deep learning technologies provide a new way to predict the survival of liver cancer. Radiomics constructs a quantitative prediction model by extracting the texture, morphology and functional characteristics of CT/MRI images through high-throughput. Deep learning models, such as convolutional neural networks (CNNs), use end-to-end learning to mine deep features directly from raw images, avoiding the limitations of artificial feature design. However, existing deep learning models still face three challenges: (1) poor interpretability; (2) Insufficient vertical information integration, mostly relying on single-point data;

(3) The models are mostly constructed based on the baseline status of patients, ignoring the dynamic changes of tumor burden and immune landscape in the treatment cycle, which is difficult to guide treatment adjustment in real time and does not meet the needs of precision medicine. Although the traditional longitudinal prediction model can incorporate dynamic changes, it needs to systematically collect follow-up data for multiple visits. In the real world, patient follow-up is often irregular, and there are medical referrals and referrals in other places, resulting in a large number of missing or uneven longitudinal imaging data, and it is difficult to align the information obtained at different time points and different devices. In addition, such models wait for patients to respond to treatment, which has a prediction lag. The LILAC system (Learning-based Inference of Longitudinal imAge Changes) represents an important breakthrough in this direction, which uses a self-supervised time-series deep learning algorithm to analyze small changes in a series of MRI scans to achieve dynamic monitoring of disease progression. The proof-of-concept study showed that LILAC was significantly superior to the traditional method in the prediction of embryonic development time sequence and geriatric cognitive score. In pediatric glioma studies, the time-series deep learning model improved the F1-score of 1-year recurrence prediction by 58.5% by analyzing serial postoperative MRI images, and the prediction performance increased with the increase of historical scans (plateau after 3-6 scans). However, the clinical translation of longitudinal models faces fundamental difficulties: patients with advanced HCC are often unable to complete the full process of imaging follow-up due to disease progression or adverse reactions, resulting in a lack of high-quality time series data; Data heterogeneity is caused by differences in image acquisition protocols in different centers. Ethics and privacy concerns limit multi-center data sharing. Therefore, there is an urgent need to develop new prediction paradigms that do not rely on real post-treatment images. Breakthroughs in Generative Artificial Intelligence provide a revolutionary solution to this.

The core strength of generative models is their ability to learn the latent distribution of data and generate high-fidelity new samples. This field has achieved technological evolution from generative adversarial networks (GANs), variational autoencoders (VAEs) to diffusion models. In recent years, the proposed diffusion model has demonstrated better image generation quality than the previous model in a number of studies through the iterative process of forward noise addition and reverse noise reconstruction. In medicine, generative models were initially used primarily for data amplification and image synthesis (e.g., to generate images of rare diseases to alleviate sample imbalances), but recent advances have expanded into the areas of treatment response simulation and disease course prediction. Compared with traditional prediction models, generative models have shown unique potential in predicting curative effects: they can capture the spatiotemporal dynamics of tumor growth through the hidden spatial representation learning ability of

diffusion models. Based on this, the model can generate virtual images that simulate the optimal treatment response according to the initial state of the patient's tumor and the characteristics of the treatment plan in the absence of data. This provides clinicians with an intuitive tool for predicting efficacy, which helps to improve the robustness of survival prediction and provides a potential reference for subsequent treatment decisions.

2. The entry point and significance of this study

The entry point of this study is to propose "generative longitudinal prediction": only pre-treatment images are needed to generate high-fidelity post-treatment image predictions, which effectively avoids the clinical dilemma of obtaining real longitudinal follow-up data. This paradigm shift not only solves the problem of longitudinal data scarcity, but also develops a new approach to treatment simulation based on digital twins. Clinicians can intuitively evaluate the possible efficacy of different regimens through the virtual multi-time point images generated by the model before treatment, providing a visual basis for individualized treatment decisions. This study integrates generative AI and dynamic risk models to achieve: 1. From static assessment to dynamic deduction; 2. Survival prediction is moved forward; 3. Individualized optimization of treatment plan. It is expected that we can provide more accurate and individualized treatment decision support for patients with advanced liver cancer without relying on longitudinal data, and ultimately improve the survival prognosis and quality of life.

Research content

1. Research Objectives

Main objectives: 1) To construct a generative image model based on preoperative CT/MRI; 2) the treatment response was evaluated in the training cohort based on the virtual image images generated by the model, including: disease control rate (DCR) and objective response rate (ORR), and compared with the best treatment response images after real treatment; 3) predict the prognosis of patients based on real pre-treatment imaging images (CT/MRI/DSA) and virtual post-treatment images, including progression-free survival (PFS) and overall survival (OS);

Secondary objectives: 1) Integrate text information such as patient demographic data, laboratory examination data, and image reports/surgical records to achieve cross-modal fusion and further improve the performance of the model; 2) Adopt machine learning methods based on causal inference to realize the recommendation of diagnosis and treatment plans.

2. Research content detail

(1) Selection of research site:

Zhongda Hospital Affiliated to Southeast University, General Hospital of Chinese People's Liberation Army (301), Third Affiliated Hospital of Naval Medical University (Oriental Hepatobiliary Surgery Hospital), First Affiliated Hospital of Zhengzhou University, Cancer Hospital Affiliated to Chinese Academy of Medical Sciences, First Affiliated Hospital of University of Science and Technology of China (Anhui Provincial Hospital), Jiangsu Cancer Hospital, Changzhou First People's Hospital, etc.

(2) Selection of research subjects:

This project is a real-world study, including retrospective training, validation and testing cohorts and prospective testing cohorts, all registries are derived from real cases in clinical practice, and the collected case data should include the following requirements:

1. Unresectable hepatocellular carcinoma confirmed by histopathological diagnosis and/or clinical diagnosis (typical imaging features, clinical manifestations, laboratory tests, etc.) (Annex: Guidelines for the diagnosis and treatment of primary liver cancer (2024 edition);
2. Patients who receive TACE combined with targeted immunotherapy on the basis of the diagnosis of unresectable hepatocellular carcinoma;
3. Liver function grade Child-Pugh A or B;
4. Over 18 years old, regardless of gender;
5. Expected survival time≥ 3 months;
6. ECOG PS score ≤2;

7. Meet the following laboratory examination parameters;

   1) Hematologic function: absolute neutrophil count ≥1.0×10⁹/L; Platelet count ≥ 50×10⁹ /L; hemoglobin ≥90 g/L; The international normalized ratio is less than 1.7 or the prothrombin time is prolonged by no more than 4 seconds,

   2) Liver function: alanine aminotransferase/aspartate aminotransferase does not exceed 5 times the upper limit of normal; total bilirubin ≤ 210 micromol/L [≤2.38 mg/dL]; Albumin ≥ 28g/L,

   3) Renal function: serum creatinine does not exceed 1.5 times the upper limit of normal.

Exclusion Criteria:

This project is a real-world study, including retrospective training, validation, and testing cohorts and prospective testing cohorts. All registries are derived from real cases in clinical practice, and the collected case data should include the following requirements:

1. At the same time, other malignant tumors except HCC were combined;
2. Moderate to severe ascites (ascites on a Child-Pugh score of 3);
3. Received other first-line or second- and third-line systemic therapy (including any regimen of systemic therapy, etc.) or any local therapy (including transcatheter interventional therapy, ablation therapy, internal/external radiotherapy, etc.) and surgical resection or Chinese herbal medicine within 4 weeks prior to TACE combined with targeted immunotherapy.
4. Incomplete data, such as incomplete baseline laboratory examination data, missing or poor quality imaging data, no prognosis information, etc.;
5. Severe liver dysfunction: liver diseases such as decompensated cirrhosis that seriously affect bilirubin levels;
6. Severe comorbidities: such as refractory hypertension (blood pressure is still higher than 150/100 mm Hg after optimal drug therapy), sustained arrhythmia (CTCAE standard grade 2 and above), atrial fibrillation of any degree, Qtc interval prolongation (more than 450 msec in men and more than 470 msec in women), renal insufficiency, etc.;
7. Concomitant human immunodeficiency virus infection (HIV) or acquired immunodeficiency disorder syndrome;
8. pregnant or lactating women;
9. Accompanied by acute or chronic mental disorders (including mental disorders that affect the enrollment, treatment intervention, and follow-up of subjects).

(3) Sample size:

Sample size estimation based on the core principles of predictive model development (TRIPOD statement). The core strategy follows an outcome event-driven approach, focusing on the number of target events (e.g., disease positivity). Using the method recommended by Riley et al., the minimum sample size was calculated by the R language pmsampsize package. The calculation parameters include: (1) expected model discrimination (e.g., AUC is preset to 0.8);(2) the estimated number of predictors in the final model (k according to the literature and pre-experiments); and (3) the expected effect strength of the predictor (conservative estimates). At the same time, considering the high-dimensional characteristics of radiomics, the total sample size meets the following empirical criteria: (1) the number of events ≥ (10 times k (EPV principle); (2) Total sample size ≥ (10 times k). The final sample size was increased by 10-20% to compensate for missing data and heterogeneity. Multicenter image heterogeneity was included in the sample size calculation, and the sample size adequacy was verified by Bootstrap sampling.

(4) Methods of extraction and allocation of research objects:

All subjects in this study were derived from real cases in clinical practice.

(5) Informed consent of the study subject:

In this study, the cases obtained from the previous process of diagnosis and treatment were used without the application for informed consent.

（6） Evaluation Indicators:

Key research indicators

In this study, the performance of the prediction model was comprehensively evaluated by the following five types of indicators, and all indicators reported point estimates and 95% confidence intervals (1000 times repeated sampling by the Bootstrap method): 1) discrimination: the C-index (Concordance Index) was used to evaluate the consistency of the model's ranking of individual survival time (two patients were randomly selected to predict the probability that the one with longer survival time would actually survive longer); 2) Time-dependent AUC (time-dependent AUC) was used to evaluate the model's ability to distinguish whether an event occurred or not at a preset time point; 3) calibration, the difference between the predicted probability and the actual outcome and the probability distribution were visualized by calibration plot and density plot, and the mean square error of the Brier Score was calculated. 4) Decision Curve Analysis (DCA) was used to evaluate clinical practicability; 5) The proportion of variation explained by the Nagelkerke R^2 assessment model was used.

Secondary Research Indicators

1) Safety: the incidence of adverse reactions and the severity of adverse events. Adverse events were

graded using the National Cancer Institute (NCI) Common Adverse Event Evaluation Criteria (CTCAE) version 5.0, which were classified as 1 to 5 based on the severity of adverse events (AEs):

Grade 1: mild; asymptomatic or mild; Seen clinically or diagnostically only; No treatment is required.

Grade 2: moderate; Requires minor, topical, or non-invasive treatments; Age-appropriate instrumental limitations in activities of daily living (e.g., cooking, buying clothes, using the phone, managing money, etc.).

Grade 3: Severe or medically significant but not immediately life-threatening; resulting in hospitalization or prolongation of hospitalization; Disability; Limitation of self-directed activities of daily living (e.g., bathing, dressing, undressing, eating, washing, taking medication, etc., and not being bedridden).

Grade 4: life-threatening; Urgent treatment is required.

Grade 5: AE-related deaths.

2) Validity**:**

Overall survival (OS) is defined as the time from the start of chemotherapy combined with immunotherapy to the patient's death from any cause, measured in "months". For subjects who were still alive at the time of data analysis, OS was calculated based on the date on which the subjects were last known to be alive**;**

Progression-free survival (PFS): defined as the time from the start of chemotherapy combined with immunotherapy to the first disease progression or death from any cause.

To assess response to treatment in patients with intrahepatic lesions:

a. Objective response rate (ORR): according to RECIST 1.1 criteria, objective response rate = (CR + PR / CR + PR + SD+PD); Participants with CR + PR efficacy should be retested 4 weeks after the first evaluation.

b. Disease control rate (DCR): according to RECIST 1.1 criteria, lesion control rate = (CR + PR + SD/ CR + PR + SD + PD); Subjects with CR + PR + SD efficacy should be retested 4 weeks after the first evaluation.

(7) Intervention methods:

The basic information of patients and related information of treatment plans are collected through the electronic medical record system, including treatment plan, dosing cycle, dosing date, dosage, frequency,

number of cycles, efficacy, adverse reactions, etc. Ensure that the data collected in the study can be traced back to the original medical records.

3. Research Methodology

This study is based on a multicenter retrospective cohort including a prospective test cohort to include patients with unresectable hepatocellular carcinoma who have received TACE in combination with target-immune therapy. The data comes from the clinical diagnosis and treatment records of 8 tertiary hospitals from 2016 to 2025, and strictly follows the preset inclusion standards.

Generative model construction: diffusion model-based treatment response simulator: input data: enhanced CT/MRI images at the patient's baseline, real post-treatment enhanced CT/MRI images and clinical features (Child-Pugh score, AFP level, treatment plan, etc.). Model architecture: The conditional diffusion model was used to dynamically simulate the virtual follow-up images of the best response after treatment with the treatment protocol (TACE embolic agent dose + target immune drug combination) as the generation conditions. Training mechanism: The model is optimized by comparing the differences in deep visual features between the generated images and the real follow-up images to ensure that the generated results conform to the laws of biological evolution. Quantitatively evaluate the visual fidelity of the generated images (FID index <25); The clinical rationality was scored by 3 interventional radiology experts by blind (on a 5-point scale≥ 4 points were considered qualified).

Dynamic survival prediction framework: Capture 256-dimensional semantic feature vectors from the generated virtual images to characterize the dynamic changes of tumor microenvironment. The depth features extracted by the generative model are fused with the patient's clinical variables, and the key predictors are automatically identified through the gated attention mechanism. Survival risk modeling: Construct a Transformer-based time-series survival model to output individualized survival curves and survival probability predictions at key time points (12/24/36 months).

Treatment decision support systems: causal effects deduction: quantifying the expected survival benefit of different treatment regimens (eg, TACE + lenvatinib + pembrolizumab versus TACE + bevacizumab + atezolizumab) using a counterfactual analysis framework. Generate a heat map of treatment regimen comparisons to visualize simulated tumor regression differences. Decision rule: When the median survival benefit of a regimen is expected to be more than 3 months and the confidence probability is greater than 80%, the system preferentially recommends the regimen.

Model validation strategy: Stratified by study center, randomly assigned to the training set (70%), validation set (30%), and multiple test sets.

Survival prediction assessment: main indicators: C-index evaluation of overall survival prediction

accuracy; Secondary indicators: 12/24/36 months of time dependence on AUC to assess discriminant ability at a specific time point; Calibration analysis: The Brier score quantifies the prediction error, and the calibration curve evaluates the risk prediction consistency. Comparative experimental design: baseline model (clinical variables only); benchmark model + real follow-up images; Datum model + generate model virtual image.

Reproducibility & Ethics: The model code and preprocessing process are published in the public repository (GitHub). The self-service resampling method (1000 times) was used to calculate the 95% confidence interval.

5. The innovation of the project

This study is a multicenter, retrospective, real-world study of unresectable hepatocellular carcinoma (HCC) undergoing TACE in combination with targeted immunotherapy in the project from January 2016 to June 2025. Baseline imaging data, previous treatments, adverse reactions, clinical outcomes, follow-up imaging data, longitudinal laboratory examination data, etc., were collected to construct a generative model and generate imaging images of the best efficacy after virtual treatment. Combine virtual and real pre-treatment images with other clinical information for accurate prognosis prediction.

6. Research plan and forecast progress

2025/6-2025/12: Complete the collection, cleaning, and statistical analysis of patients' baseline images, efficacy evaluation, and safety data;

2026/01-2026/06: Complete the construction of the generative model, and evaluate and predict the performance of the model;

2025/06-2026/12: Write a research paper and submit it, and complete the final report.

7. Expected research results

(1) Construct a prognosis and treatment decision-making model based on TACE combined with targeted immunotherapy for unresectable liver cancer based on generative model.

(2) Publish 2-3 high-quality original papers in journals indexed by SCI.

(3) Apply for 1 invention patent and participate in academic exchanges at home and abroad.

Bases and conditions of work

1. The accumulation of research work related to the project and the research results obtained;
2. The existing experimental conditions, the missing experimental conditions and the ways to be solved;
3. List of personnel in the main enrollment center;

name	gender	Job title	Units and departments	Division of labor in the project
Lu Jian	man	Deputy Chief Physician	Zhongda Hospital Affiliated to Southeast University	Primary enrollment center
Zhang Xiuping	man	Deputy Chief Physician	General Hospital of the Chinese People's Liberation Army	Primary enrollment center
Zhang Kai	man	Attending	Cancer Hospital Affiliated to Chinese Academy of Medical Sciences	Primary enrollment center
Lu Weifu	man	Chief physician	The First Affiliated Hospital of University of Science and Technology of China	Primary enrollment center
Ge Naijian	man	Chief physician	The Third Affiliated Hospital of the Naval Medical University	Primary enrollment center
Wu Jun	man	Chief physician	The First People's Hospital of Changzhou	Primary enrollment center
Wu Pingping	woman	Chief physician	Jiangsu Provincial Cancer Hospital	Primary enrollment center
Li Zhen	man	Chief physician	The First Affiliated Hospital of Zhengzhou University	Primary enrollment center

4. Research team;

	name	job title	workplace	Division of tasks
Key members of the project team (including project leader)	Lu Jian	Deputy Chief Physician	Interventional & Vascular Surgery	Schematic design
	Chen yang	professor	School of Computer Science, Southeast University	Schematic design
	Xue Cheng	professor	School of Computer Science, Southeast University	Algorithm support

	Mei Junhao	____	Interventional & Vascular Surgery	Statistical analysis
	Du Zheyu	____	School of Computer Science, Southeast University	Code implementation
	Jia Kaizhi	____	Interventional & Vascular Surgery	Data Collection

Budget

The clinical research project initiated by me is based on generative models to construct a survival prediction and treatment decision-making system for unresectable hepatocellular carcinoma with TACE combined with targeted immunotherapy, without funding from any institution, organization or individual.

Miscellaneous

The applicant undertakes that the above content has taken into account all possible risks in the project. Once any risk arises, it will be reported in a timely manner and measures will be taken to control and resolve the risk to prevent the risk from expanding and spreading, and to minimize the risk. In the event of a serious adverse event, I will promptly report it to the Science and Technology Department, the Medical Department and the Hospital Ethics Committee within 24 hours.