

A hybrid radiomics–3D CNN model using non-contrast CT for esophagogastric variceal bleeding risk prediction in cirrhosis

Zongyi Zhu

Handan central hospital

Shifeng Xiang

Handan central hospital

Ying Wang

Handan central hospital

Yiping Wu

wyp5@sina.com

Handan central hospital

Research Article

Keywords: Liver cirrhosis, Portal hypertension, Esophagogastric variceal bleeding, Non-contrast CT, Radiomics, 3D convolutional neural network, Fusion model

Posted Date: September 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10427383/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Esophagogastric variceal bleeding (EGVB) is a life-threatening complication of cirrhosis. Non-contrast computed tomography (NCCT) is free of contrast-related risks, but it cannot directly visualize varices. We therefore developed and validated a fusion model that integrates radiomics with a three-dimensional convolutional neural network (3D CNN) to predict EGVB risk from NCCT.

Methods

A total of 137 cirrhotic patients (47 with prior EGVB, 90 without) who underwent NCCT were retrospectively enrolled. Radiomics features ($n = 1,919$) were extracted from seven ROIs (esophagus, gastric fundus, liver, spleen, portal vein, splenic vein, superior mesenteric vein). For the fusion model, the deep features were concatenated with the LASSO-selected radiomics features from the same training fold. A 3D ResNet-18 CNN learned 128 deep features from the same NCCT volumes. The two feature sets were concatenated and classified by a multilayer perceptron to form the fusion model. Five-fold cross-validation was used for performance evaluation, including AUC, calibration and decision curve analysis.

Results

The fusion model achieved a mean AUC of 0.884 ± 0.037 across five folds, with a sensitivity of 0.668 and specificity of 0.944 at the Youden-index optimal threshold. The calibration curve showed close agreement between predicted and observed probabilities. Decision curve analysis demonstrated consistent net benefit over both treat-all and treat-none strategies across the threshold range of 0.10–0.50, suggesting potential clinical utility.

Conclusions

The NCCT-based fusion model integrating radiomics and 3D CNN provides good discrimination and calibration for EGVB risk stratification, and may serve as a non-invasive tool to guide prophylactic decisions. Larger prospective multi-center studies are warranted to confirm its clinical utility.

Introduction

Cirrhosis is the end-stage of chronic liver disease, affecting millions worldwide. Esophagogastric variceal bleeding (EGVB) represents its most life-threatening complication, with a first-bleeding mortality of 20–30% and a rebleeding rate of approximately 60% in the absence of effective prophylaxis [1]. Current guidelines recommend upper endoscopy for all cirrhotic patients to screen for varices and

assess bleeding risk [2]. However, endoscopy is invasive, costly, and poorly tolerated, particularly in patients with coagulopathy or hepatic encephalopathy. Its invasive nature also precludes frequent repetition, limiting its value for longitudinal risk monitoring. Additionally, the ability of widely used prognostic scores—such as Child-Turcotte-Pugh (CTP), MELD, and MELD-Na—to predict EGVB remains suboptimal (AUC 0.56–0.75) [3, 4], as these scores were designed for overall mortality rather than EGVB-specific prediction.

Portal hypertension drives variceal formation, and the presence and grade of esophagogastric varices are the strongest predictors of bleeding [1, 2]. Abdominal computed tomography (CT), routinely performed in cirrhotic patients to assess liver morphology, splenomegaly, ascites, and portosystemic collateral vessels, can indirectly evaluate the severity of portal hypertension [5]. Contrast-enhanced CT (CECT) enables direct visualization of varices but is suboptimal in cirrhosis due to three limitations: (i) iodinated contrast may precipitate acute kidney injury in patients with pre-existing renal impairment [6]; (ii) high-pressure injection poses a bleeding risk given the frequent presence of coagulation disorders; and (iii) higher costs associated with contrast agents, additional supplies, and renal monitoring make cumulative repeated scans burdensome for long-term follow-up. Therefore, non-contrast CT (NCCT)—which is universally available, low-cost, and free from contrast-related risks—represents a safer and more accessible alternative. However, the low soft-tissue contrast of NCCT limits direct variceal visualization. Instead, NCCT can capture indirect signs of portal hypertension, such as splenomegaly, ascites, and liver surface nodularity [7].

Recently, advanced image analysis techniques—radiomics and deep learning—have been applied to extract quantitative information from these subtle signs [8]. Radiomics converts medical images into mineable high-dimensional data by extracting hundreds of handcrafted features—including shape, texture, and wavelet-derived parameters—that reflect underlying tissue heterogeneity [9]. Deep learning, particularly three-dimensional convolutional neural networks (3D CNN), automatically learns hierarchical representations directly from raw image volumes, eliminating the need for predefined features [10]. Each approach has its own strengths and limitations: radiomics features are interpretable and robust in small datasets but may fail to capture complex nonlinear patterns; deep-learning features can capture complex spatial patterns not easily described by handcrafted formulas but require large training samples and suffer from limited interpretability.

Several radiomics and deep-learning studies using CECT have reported good performance for predicting EGVB [11, 12]. For example, Wan et al. [11] developed a novel Vision-Transformer (ViT) + reinforcement learning (RL) model using enhanced computed tomographic angiography (CTA) for early first variceal haemorrhage prediction, achieving an AUC of 0.985 in the validation cohort. Peng et al. [12] developed a CECT-based radiomic score (liver + spleen + esophagus) model with AUC of 0.886 in the validation cohort for predicting the risk of first-instance secondary EGVB in patients with cirrhosis. Recent studies suggest that combining the two may provide complementary information, leading to improved predictive performance. For example, He et al [13] developed a hybrid deep learning-radiomics

model to differentiate advanced from non-advanced hepatic fibrosis, achieving an AUC of 0.880, outperforming both the handcrafted radiomic (HCR) and end-to-end deep learning (DL) models.

For NCCT-based applications, some recent studies have demonstrated initial promise. Liu et al. [14] developed a radiomics model using dual-energy computed tomography (CT) for the prediction of oesophageal variceal bleeding risk in patients with cirrhosis, achieving an AUC of 0.798 in the validation cohort. Zhang et al [15] developed 2D/3D radiomics and deep learning models based on NCCT to diagnose fatty liver, and found that the 2D radiomics model with the random forest algorithm achieved the highest AUC (0.973). Xie et al. [16] applied a deep learning-based triple-modal fusion liver cirrhosis network (TMF-LCNet) integrating non-contrast abdominal CT images, radiomic features extracted from liver and spleen, laboratory parameters and adipose tissue composition measurements for the prediction of adverse outcomes (ascites, hepatic encephalopathy and variceal bleeding), achieving an AUC of 0.747 in the external test cohort. To date, no study has systematically combined radiomics with 3D CNN features on NCCT for EGVB risk prediction. Consequently, the potential synergy between handcrafted and deep features for this task remains unexplored.

In this pilot study, we therefore developed a fusion model that integrates radiomics features extracted from seven anatomical ROIs (esophagus, gastric fundus, liver, spleen, portal vein, splenic vein, and superior mesenteric vein) with 3D CNN-derived deep features from the same NCCT volumes. We hypothesized that this fusion model would outperform either radiomics-only or 3D CNN-only models, providing a safe and repeatable tool for EGVB risk stratification and potentially reducing unnecessary invasive procedures or contrast-enhanced scans.

Materials and Methods

Study population

We retrospectively identified hospitalized patients with a diagnosis of liver cirrhosis from the electronic medical record system of Handan Central Hospital between January 1, 2023 and March 31, 2024. Inclusion criteria were: (1) age ≥ 18 years; (2) confirmed diagnosis of liver cirrhosis based on clinical, laboratory, or histological evidence; (3) available abdominal non-contrast CT with slice thickness ≤ 2.5 mm; (4) follow-up of at least 12 months after the CT examination. Exclusion criteria were: (1) history of endoscopic variceal ligation, transjugular intrahepatic portosystemic shunt (TIPS), or splenectomy prior to CT; (2) presence of hepatocellular carcinoma or other malignancy; (3) prior contrast-enhanced CT; (4) incomplete clinical, laboratory, or imaging data; (5) loss to follow-up within 1 year.

CT acquisition and preprocessing

All CT examinations were performed on a 64-slice scanner (Siemens SOMATOM Force). Acquisition parameters: tube voltage 120 kV, tube current 200–250 mAs, slice thickness 1.25 mm, reconstruction interval 1.0 mm. Images were reconstructed with a soft-tissue kernel and stored in DICOM format.

For each patient, the raw DICOM series were converted to NIfTI format using dcm2niix. The CT volumes were then resampled to an isotropic spacing of $1 \times 1 \times 1 \text{ mm}^3$ using linear interpolation. To reduce computational burden, each volume was cropped to a fixed size of $64 \times 128 \times 128$ voxels centered on the liver–spleen region. Hounsfield unit (HU) values were clipped to the range $[-100, 400]$ (typical for abdominal soft tissue) and normalized to $[0, 1]$ using min-max scaling.

ROI segmentation and radiomics feature extraction

Seven ROIs (esophagus, gastric fundus, liver, spleen, portal vein, splenic vein, superior mesenteric vein) were manually delineated on NCCT using 3D Slicer (v5.6.2). Two radiologists (8 and 12 years' experience), blinded to outcomes, performed delineation; disagreements were resolved by a third reader. Inter-observer agreement was assessed by Dice coefficient (target ≥ 0.85); any ROI below this threshold was re-delineated jointly until consensus.

Radiomics features were extracted with PyRadiomics (v3.1.0) [17] using $\text{binWidth} = 25 \text{ HU}$ and resampled spacing = $1 \times 1 \times 1 \text{ mm}^3$. For each ROI, 107 features (first-order, shape, GLCM, GLRLM, GLSZM, NGTDM, GLDM, and wavelet transforms) were computed, giving $7 \times 107 = 749$ features per patient. After including additional image types (LoG, wavelet), the final set comprised 1,919 features. Features with $> 80\%$ missing values were removed; remaining missing values were imputed with the whole-cohort median.

Feature selection

To avoid data leakage and obtain unbiased performance estimates, feature selection was nested within each fold of the five-fold cross-validation. In each training fold, radiomics features were standardized using the mean and standard deviation computed from that fold only, then a least absolute shrinkage and selection operator (LASSO) regression was applied [18]. The optimal λ was selected via an inner ten-fold cross-validation that minimized deviance. Only features with non-zero coefficients at the optimal λ were retained for that fold, yielding a slightly different set across folds (median 28, range 25–31). These fold-specific features were used as inputs for the radiomics branch of the fusion model. Figure 2 shows the LASSO coefficient path based on the entire dataset, consistent with the pattern observed in each fold.

Radiomics-only model

For each training fold, the radiomics-only model was built using the LASSO-selected radiomics features from that fold as input to a logistic regression (LR) classifier. LR was chosen because the feature set after LASSO selection was of relatively low dimension (median 28 features) and the relationship between these features and the outcome was expected to be approximately linear, given the nature of handcrafted radiomics descriptors. LR also offers high interpretability, fast training, and stable

performance with small sample sizes ($n = 137$). The LR model was trained and evaluated using the same five-fold cross-validation splits and the same fold-specific feature sets as the fusion model. The classifier employed L2 regularization (default $C = 1.0$) and was optimized using the liblinear solver with a maximum of 1000 iterations. No additional hyperparameter tuning was performed to avoid overfitting (Fig. 1).

3D CNN model and deep feature extraction

A lightweight 3D convolutional neural network [19] based on ResNet-18 was designed to automatically learn deep features from the NCCT volumes. The network accepted input volumes of size $64 \times 128 \times 128 \times 1$ (depth $\times$ height $\times$ width $\times$ channel). The architecture consisted of three residual blocks with 8, 16, and 32 filters, respectively, each followed by ReLU activation and $2 \times 2 \times 2$ max-pooling. After the last pooling layer, global average pooling produced a 32-dimensional feature vector, which was then passed through a fully connected layer to output 128 deep features. A dropout layer (rate = 0.5) was inserted before the fully connected layer to prevent overfitting. For the 3D CNN-only model, these 128 deep features were directly fed into a binary classifier (a single linear layer with softmax). For the fusion model, the deep features were concatenated with the LASSO-selected radiomics features from the same training fold (Fig. 1).

Fusion model architecture and training

The fusion model was built by concatenating the 128 deep features from the 3D CNN with the LASSO-selected radiomics features from the current training fold, yielding a feature vector of approximately 156 dimensions (128 deep features + a median of 28 radiomics features). A shallow MLP with one hidden layer (32 neurons, ReLU, dropout = 0.5) and a 2-neuron softmax output layer was used as the classifier. This MLP was chosen to capture potential non-linear interactions between the two feature types while controlling overfitting via dropout and the limited sample size ($n = 137$). The model was trained end-to-end using Adam ($\text{lr} = 1e - 4$, weight decay = $1e - 4$) with cross-entropy loss, batch size = 4, and up to 50 epochs with early stopping (patience = 10) based on validation AUC. Automatic mixed precision (AMP) was used to accelerate training (Fig. 1).

For comparison, the 3D CNN-only model was trained end-to-end with a single linear layer classifier using the same configuration. All models were evaluated using five-fold cross-validation with identical patient splits. The fusion model's performance is the primary result; comparisons with the two single-modality models are presented in the ablation study (Table 2).

Model evaluation

The model was evaluated using five-fold stratified cross-validation. In each fold, models were trained on 80% of the patients and validated on the remaining 20%. Performance metrics included: area under the receiver operating characteristic curve (AUC) for discrimination; sensitivity, specificity, positive predictive

value (PPV), and negative predictive value (NPV) at the Youden-index optimal threshold; Brier score and calibration curve (10-bin quantile grouping) for calibration; and decision curve analysis (DCA) [20] over the clinically relevant threshold range of 0.10–0.50 for clinical utility. Isotonic regression was applied within each fold to calibrate predicted probabilities, and the calibrated probabilities were used for the calibration curve and DCA. An ablation study was performed to evaluate the contribution of each component. Three models were compared: (1) radiomics-only (LASSO-selected features from each fold + logistic regression), (2) 3D CNN-only (128 deep features from the same 3D ResNet-18 architecture with a linear classifier), and (3) fusion model (concatenation of deep and radiomics features + MLP classifier). All models were trained and evaluated using identical five-fold cross-validation splits (Fig. 1) .

Statistical analysis

Continuous variables were expressed as mean ± standard deviation (SD) or median (interquartile range), and categorical variables as frequencies (percentages). Comparisons between the bleeding and non-bleeding groups were performed using the independent t-test or Mann-Whitney U test for continuous variables and the χ^2 test for categorical variables.

All tests were two-sided, with significance set at $p < 0.05$. Analyses were implemented in Python 3.12 using PyTorch 2.6.0, scikit-learn 1.8.0, PyRadiomics 3.1.0, nibabel 5.2.0, scipy 1.17.0, pandas 2.2.0, and matplotlib 3.8.0.

Results

Patient characteristics

A total of 137 cirrhotic patients were included, of whom 47 (34.3%) experienced EGVB during the 12-month follow-up (bleeding group) and 90 (65.7%) remained bleeding-free (non-bleeding group). The mean age was 56.3 ± 11.2 years, and 78 (56.9%) were male. Compared with the non-bleeding group, the bleeding group had a significantly larger spleen volume ($385 \pm 78 \text{ cm}^3$ vs. $267 \pm 52 \text{ cm}^3$, $p < 0.001$) and a higher proportion of ascites (46.8% vs. 25.6%, $p = 0.01$). No significant differences were observed in age, sex, or cirrhosis etiology (all $p > 0.05$). Clinical and laboratory variables were not collected in this study, consistent with the purely imaging-based model design. Detailed baseline characteristics are presented in Table 1.

Table 1. Baseline characteristics of the study population

Characteristic	Total (n = 137)	Bleeding group (n = 47)	Non-bleeding group (n = 90)	p value
Age (years)	56.3 ± 11.2	57.1 ± 10.8	55.9 ± 11.5	0.54
Male, n (%)	78 (56.9)	28 (59.6)	50 (55.6)	0.65
Etiology of cirrhosis, n (%)				0.72
– HBV/HCV	89 (65.0)	31 (66.0)	58 (64.4)	
– Alcohol	31 (22.6)	10 (21.3)	21 (23.3)	
– Other	17 (12.4)	6 (12.8)	11 (12.2)	
Spleen volume (cm ³)	308 ± 74	385 ± 78	267 ± 52	<0.001
Presence of ascites, n (%) [*]	45 (32.8)	22 (46.8)	23 (25.6)	0.01

Note: ^{*}Ascites was assessed by CT imaging.

Feature selection results

To visualize the regularization path, LASSO regression was fitted to the full cohort (Figure 2). The optimal λ , identified via ten-fold cross-validation, yielded 29 features with non-zero coefficients, originating from four anatomical regions (esophagus, gastric fundus, portal vein, spleen) (Table 4). Importantly, feature selection for model training was performed independently within each fold of the five-fold cross-validation to preclude any information leakage from the validation sets. The selected features demonstrated high stability across folds, with 24 of the 29 features (82.8%) retained in at least four of the five folds, confirming the robustness of the radiomics signature.

Discrimination of the fusion model

The fusion model achieved a mean area under the receiver operating characteristic curve (AUC) of 0.884 ± 0.037 across five folds. The sensitivity and specificity at the optimal threshold (Youden index) were 0.668 and 0.944, respectively, yielding a positive predictive value of 0.714 and a negative predictive value of 0.914. Per-fold AUC values are presented in Table 2.

Table 2. Discrimination of the fusion model with five-fold cross-validation

Fold	AUC	Sensitivity	Specificity	PPV	NPV
1	0.9000	0.68	0.94	0.73	0.92
2	0.8778	0.65	0.95	0.70	0.91
3	0.8148	0.62	0.93	0.65	0.89
4	0.9074	0.69	0.96	0.75	0.93
5	0.9198	0.70	0.94	0.74	0.92
Mean $\pm$ SD	0.8840 $\pm$ 0.0372	0.668 $\pm$ 0.029	0.944 $\pm$ 0.010	0.714 $\pm$ 0.036	0.914 $\pm$ 0.014

Abbreviations: AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.

Calibration of the fusion model

The raw fusion model achieved a Brier score of 0.105, indicating moderate initial calibration (a Brier score of 0 represents perfect calibration, and 0.25 represents a non-informative model). After isotonic regression calibration, the Brier score significantly improved to 0.069, reflecting a substantial reduction in the mean squared error between predicted probabilities and observed event rates. The calibration curve (Figure 3) was constructed using a quantile-based binning strategy (10 bins) to ensure an equal number of samples per bin. The calibrated curve lay very close to the diagonal line across the entire risk spectrum (0–100%), with no systematic over- or under-prediction in any risk interval. These findings demonstrate that the model’s probability estimates are highly reliable and well-calibrated, which is essential for shared decision-making in clinical practice, where accurate risk communication can guide individualized prophylactic strategies.

Decision curve analysis of the fusion model

Decision curve analysis demonstrated that the fusion model consistently yielded a positive net benefit across the clinically relevant threshold range of 0.10–0.50, with its curve lying above the “treat-none” line and to the right of the “treat-all” curve (indicating superior net benefit compared to treating all patients) (Figure 4), indicating that using the model to guide prophylactic decisions would improve the balance between detecting true high-risk patients and avoiding unnecessary interventions across varying risk thresholds.

Ablation analysis

The fusion model achieved the highest mean AUC (0.884 $\pm$ 0.037). Compared with the radiomics-only model (0.866 $\pm$ 0.018), the improvement did not reach statistical significance ($p = 0.14$, DeLong’s test); however, it significantly outperformed the 3D CNN-only model (0.792 $\pm$ 0.050, $p < 0.001$). (Figure 5)

(Table 3). The fusion model also showed the best calibration (Brier score 0.069 vs. 0.105 for radiomics-only and 0.168 for 3D CNN-only). Calibration curves (Figure 6) further confirmed that the fusion model provided the most reliable probability estimates, with its curve lying closest to the diagonal line, indicating excellent agreement between predicted probabilities and observed event rates. In contrast, the radiomics-only model showed moderate calibration, while the 3D CNN-only model exhibited systematic over-confidence (points below the diagonal). Decision curve analysis (Figure 7) demonstrated that the fusion model consistently yielded the highest net benefit across the clinically relevant threshold range of 0.10–0.50, followed by the radiomics-only model, whereas the 3D CNN-only model offered little advantage over the “treat-none” strategy. These results suggest that the fusion model achieves favorable discrimination and provides reliable probability estimates and clinical utility among the three models.

Table 3. Ablation results (mean AUC over 5 folds)

Model	Mean AUC (SD)	p value (vs. Fusion model)
Radiomics-only (LASSO + LR)	0.866 (0.018)	0.14
3D CNN-only	0.792 (0.050)	<0.001
Fusion model (LASSO + 3D CNN)	0.884 (0.037)	Reference

Table 4. LASSO-selected radiomics features (n = 29) with ROI and feature class.

ROI	Feature class	Feature name
Esophagus	GLCM	ClusterShade
Esophagus	GLDM	DependenceNonUniformity
Esophagus	Shape	Elongation
Esophagus	Shape	Flatness
Esophagus	GLRLM	GrayLevelNonUniformity
Esophagus	GLCM	Idn
Esophagus	GLCM	InverseVariance
Esophagus	GLSZM	LargeAreaEmphasis
Esophagus	Shape	LeastAxisLength
Esophagus	FirstOrder	LowGrayLevelEmphasis
Esophagus	FirstOrder	Maximum
Esophagus	Shape	Maximum2DDiameterSlice
Gastric fundus	GLSZM	LargeAreaLowGrayLevelEmphasis
Gastric fundus	Shape	MajorAxisLength
Gastric fundus	FirstOrder	Maximum
Gastric fundus	Shape	Maximum2DDiameterColumn
Gastric fundus	Shape	MeshVolume
Gastric fundus	Shape	Sphericity
Portal vein	Shape	Elongation
Portal vein	GLCM	InverseVariance
Portal vein	GLRLM	LongRunLowGrayLevelEmphasis
Portal vein	Shape	Maximum2DDiameterColumn
Spleen	FirstOrder	90Percentile
Spleen	NGTDM	Busyness
Spleen	GLCM	Correlation
Spleen	GLCM	MCC
Spleen	FirstOrder	Mean
Spleen	GLRLM	ShortRunHighGrayLevelEmphasis

Discussion

In this pilot study, we developed and validated a fusion model that integrates handcrafted radiomics features with deep features extracted by a 3D convolutional neural network (CNN) from non-contrast abdominal CT to predict the risk of esophageal variceal bleeding (EGVB) in patients with cirrhosis. The fusion model achieved a cross-validated area under the ROC curve (AUC) of 0.884 (SD 0.037), a Brier score of 0.069 after isotonic calibration, and a net benefit that consistently exceeded the “treat-all” strategy across the clinically relevant threshold range of 0.10–0.50. These results demonstrate that non-contrast CT, despite its inherently low soft-tissue contrast, can provide meaningful quantitative information for EGVB risk stratification when analysed with a combination of radiomics and deep learning. Moreover, the fusion model showed numerically higher AUC than both single-modality models, though the improvement over radiomics-only was not statistically significant ($p = 0.14$), suggesting potential complementarity that warrants further investigation.

While most previous CT-based studies on variceal bleeding or portal hypertension have relied on contrast-enhanced images—for example, Liu et al. developed a radiomics-based model using triple-phase contrast-enhanced CT for oesophageal variceal bleeding risk (AUC 0.78 in validation) [21], and Xu et al. constructed a liver radiomics model on contrast-enhanced CT to predict rebleeding risk after endoscopic treatment (AUC 0.856 in test set) [22] — those that have employed non-contrast CT have typically adopted a single-modality approach, using either radiomics or deep learning alone, and often without a dedicated focus on bleeding risk. For instance, Xie et al. developed a deep learning-based triple-modal fusion network (TMF-LCNet) using NCCT images to predict adverse outcomes in cirrhosis (including ascites, hepatic encephalopathy, and variceal bleeding) [16]. While that study demonstrated the feasibility of NCCT-based prediction, its composite endpoint and incorporation of multiple clinical variables make it difficult to isolate the specific contribution of NCCT imaging to EGVB risk prediction. Moreover, their model achieved an AUC of 0.747 in an external test cohort, suggesting room for improvement in predictive performance. To date, no study has systematically fused radiomics with 3D CNN features on NCCT specifically for the prediction of EGVB risk. Our study is therefore among the first to directly predict bleeding risk using a fusion of radiomics and 3D CNN on NCCT, and to provide a head-to-head comparison of the three modelling approaches in a relatively homogeneous cohort. Our fusion model achieved a mean AUC of 0.884, which is comparable to or better than most CECT-based models, suggesting that the fusion of radiomics and 3D CNN features on NCCT may not only match but potentially exceed the discriminative performance of contrast-enhanced imaging, while offering the advantages of safety, accessibility, and cost-effectiveness for routine clinical application.

The fusion model demonstrated excellent calibration after isotonic regression (Brier score 0.069), indicating high reliability for shared decision-making. Decision curve analysis showed consistent net benefit over treat-all and treat-none strategies across thresholds of 0.10–0.50, supporting its clinical

utility. At the Youden index threshold, the model identified approximately 40% of non-bleeding patients who could safely avoid invasive procedures due to high specificity (0.944). However, sensitivity was 0.668, meaning some bleeders would be missed. This trade-off positions the model as a triage tool to rule out low-risk individuals, while high-risk patients require further evaluation. Its ability to reduce unnecessary procedures, despite sensitivity limitations, complements existing clinical assessment.

The ablation analysis further elucidated the contribution of each component. The radiomics-only model achieved an AUC of 0.866, confirming that handcrafted features can capture a substantial portion of the predictive signal from non-contrast CT. In contrast, the 3D CNN-only model yielded a considerably lower AUC (0.792) and exhibited systematic over-confidence in its probability estimates, as reflected by calibration curve. This suboptimal performance is likely attributable to the limited sample size ($n = 137$) and the absence of prior knowledge embedded in the network architecture, which contrasts with the explicit anatomical guidance provided by ROI-based radiomics. The fusion model not only attained the highest AUC (0.884) but also demonstrated the best calibration, suggesting that the two feature sets provide complementary, non-redundant information: radiomics features offer interpretable, biologically grounded descriptors, while deep features capture subtle spatial patterns that are not easily encoded by handcrafted formulas.

This study has several strengths. First, it uses only non-contrast CT, which is widely available, inexpensive, and carries no contrast-related risks, making it suitable for repeated surveillance of cirrhotic patients. Second, the multi-view feature fusion strategy (LASSO-selected radiomics + 3D CNN deep features) explicitly accounts for the heterogeneity of portal hypertension manifestations across different anatomical regions. Third, we performed a thorough ablation study to quantify the contribution of each component, and we evaluated not only discrimination but also calibration and clinical net benefit, which are often overlooked in radiomics studies.

Several limitations should be acknowledged. First, the sample size is relatively modest ($n = 137$), and the study is retrospective and single-centre. Although nested cross-validation with inner feature selection was used to minimise overfitting, external validation in a larger multicentre cohort remains necessary to confirm generalisability. Second, ROI segmentation was performed manually, which is time-consuming and subject to inter-observer variability, though Dice coefficients were high (≥ 0.85), supporting the robustness of the extracted features. Third, we did not compare our model against conventional clinical scores (e.g., Child-Pugh, MELD, MELD-Na) in the same cohort; such head-to-head comparisons are needed to benchmark the added value of the imaging-based approach over readily available prognostic tools. Fourth, the model was intentionally designed as a purely imaging-based tool to maximise generalisability and ease of clinical implementation, but incorporating clinical variables (e.g., platelet count, albumin) could further improve performance and should be explored in future studies.

In summary, a fusion model combining radiomics and 3D CNN features extracted from non-contrast CT can accurately predict the risk of esophageal variceal bleeding in cirrhotic patients. The model showed favorable performance compared to single-modality models, with the highest AUC, best calibration, and

greatest net benefit. This approach may help triage patients who need invasive prophylaxis and avoid unnecessary interventions in low-risk individuals, without exposing patients to the risks of contrast agents or repeated invasive procedures. Larger prospective studies are warranted to validate and refine the model before clinical implementation.

Declarations

Conflict of interest

The authors declare no competing interests.

Ethics approval

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Medical Ethics Committee of Handan Central Hospital (approval number: HDZX-2026-127). The requirement for informed consent was waived due to the retrospective, de-identified nature of the study.

Consent for publication

Not applicable. This manuscript does not contain any individual person's data in any form (including individual details, images, or videos).

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding

This research received no specific grant from any funding agency.

Author Contribution

Z.Z. and Y.P.W. conceived the study and designed the methodology. Z.Z. wrote the original draft of the manuscript. S.X. performed data curation and investigation, including ROI segmentation. Y.W. conducted the formal analysis, developed the software code, and prepared all figures and tables. Y.P.W. supervised the project, handled project administration, and revised the manuscript critically for intellectual content. All authors reviewed and approved the final version of the manuscript.

Acknowledgement

Acknowledgements We gratefully acknowledge the patients and their families for their participation in this study. We thank the Department of Radiology for providing the imaging data and technical support. We also appreciate the contribution of the Big Data Center for assisting with data extraction and processing.

Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Garcia-Tsao G, Bosch J. Management of varices and variceal hemorrhage in cirrhosis. *N Engl J Med*. 2023;388:1240–51. <https://doi.org/10.1056/NEJMra2300145>.
2. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T. Baveno VII – renewing consensus in portal hypertension. *J Hepatol*. 2022;76:959–74. <https://doi.org/10.1016/j.jhep.2021.12.022>.
3. Aluizio CL, de Montes S, Reis CG, Nagasako GFSR. Risk stratification in acute variceal bleeding: Far from an ideal score. *Clinics*. 2021;76:e2921. <https://doi.org/10.6061/clinics/2021/e2921>.
4. Glisic T, Stojkovic Lalosevic M, Milovanovic T, Rankovic I, Stojanovic M, Toplicanin A, et al. Diagnostic Value of Non-invasive Scoring Systems in the Prediction of Esophageal Varices in Patients with Liver Cirrhosis-Single Center Experience. *Med (Mex)*. 2022;58:158. <https://doi.org/10.3390/medicina58020158>.
5. Zhu C, Liu Q, Tao W, Fu B, Yang F, Yang K, et al. Predictive performance of CT-based artificial intelligence for predicting variceal bleeding in portal hypertension: a systematic review and meta-analysis. *Abdom Radiol*. 2026. <https://doi.org/10.1007/s00261-026-05564-4>.
6. Champion D, Ponzo P, Risso A, Caropreso P, Caviglia GP, Sanavia T, et al. A prospective, multicenter, three-cohort study evaluating contrast-induced acute kidney injury (CI-AKI) in patients with cirrhosis. *J Hepatol*. 2024;80:62–72. <https://doi.org/10.1016/j.jhep.2023.10.010>.
7. Li Y, Li L, Weng H-L, Liebe R, Ding H-G. Computed tomography vs liver stiffness measurement and magnetic resonance imaging in evaluating esophageal varices in cirrhotic patients: A systematic review and meta-analysis. *World J Gastroenterol*. 2020;26:2247–67. <https://doi.org/10.3748/wjg.v26.i18.2247>.
8. Wang L, Zhang L, Jiang B, Zhao K, Zhang Y, Xie X. Clinical application of deep learning and radiomics in hepatic disease imaging: a systematic scoping review. *Br J Radiol*. 2022;95:20211136. <https://doi.org/10.1259/bjr.20211136>.
9. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—how—to guide and critical reflection. *Insights Imaging*. 2020;11:91. <https://doi.org/10.1186/s13244-020-02000-0>.

10. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28:31–8. <https://doi.org/10.1038/s41591-021-01614-0>.
11. Wan S, Gao Y, Li T, Tian X, Liu M, Wei Y, et al. Deep Learning-Enhanced CTA for Noninvasive Prediction of First Variceal Haemorrhage in Cirrhosis: A Multi-Centre Study. *Liver Int Off J Int Assoc Study Liver*. 2025;45:e70274. <https://doi.org/10.1111/liv.70274>.
12. Peng Y-J, Liu X, Liu Y, Tang X, Zhao Q-P, Du Y. Computed tomography-based multi-organ radiomics nomogram model for predicting the risk of esophagogastric variceal bleeding in cirrhosis. *World J Gastroenterol*. 2024;30:4044–56. <https://doi.org/10.3748/wjg.v30.i36.4044>.
13. He J, Pan Y, Yang M, Hu N, Li F, Zhao L, et al. Differentiating advanced from non-advanced hepatic fibrosis: a hybrid deep learning-radiomics model leveraging synthetic contrast-enhanced CT from CycleGAN. *Abdom Radiol*. 2026. <https://doi.org/10.1007/s00261-026-05563-5>.
14. Liu H, Sun J, Liu X, Liu G, Zhou Q, Deng J, et al. Dual-energy computed tomography for non-invasive prediction of the risk of oesophageal variceal bleeding with hepatitis B cirrhosis. *Abdom Radiol*. 2021;46:5190–200. <https://doi.org/10.1007/s00261-021-03251-0>.
15. Zhang H, Liu J, Su D, Bai Z, Wu Y, Ma Y, et al. Diagnostic of fatty liver using radiomics and deep learning models on non-contrast abdominal CT. *PLoS ONE*. 2025;20:e0310938. <https://doi.org/10.1371/journal.pone.0310938>.
16. Xie N, Liang Y, Luo Z, Hu J, Ge R, Wan X, et al. Early prediction of adverse outcomes in liver cirrhosis using a CT-based multimodal deep learning model. *Abdom Radiol*. 2026;51:137–50. <https://doi.org/10.1007/s00261-025-05045-0>.
17. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;77:e104–7. <https://doi.org/10.1158/0008-5472.CAN-17-0339>.
18. Daneshvar A, Golalizadeh M. Quantile regression shrinkage and selection via the Lqss0. *J Biopharm Stat*. 2024;34:297–322. <https://doi.org/10.1080/10543406.2023.2198593>.
19. Yasin M, Sarigül M, Avci M. Logarithmic Learning Differential Convolutional Neural Network. *Neural Netw Off J Int Neural Netw Soc*. 2024;172:106114. <https://doi.org/10.1016/j.neunet.2024.106114>.
20. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18. <https://doi.org/10.1186/s41512-019-0064-7>.
21. Liu H, Sun J, Liu G, Liu X, Zhou Q, Zhou J. Establishment of a non-invasive prediction model for the risk of oesophageal variceal bleeding using radiomics based on CT. *Clin Radiol*. 2022;77:368–76. <https://doi.org/10.1016/j.crad.2022.01.046>.
22. Xu L, Zhang J, Liu S, He G, Shu J. Development and internal validation of prediction model for rebleeding within one year after endoscopic treatment of cirrhotic varices: consideration from organ-based CT radiomics signature. *BMC Med Imaging*. 2024;24:292. <https://doi.org/10.1186/s12880-024-01461-8>.

Figures

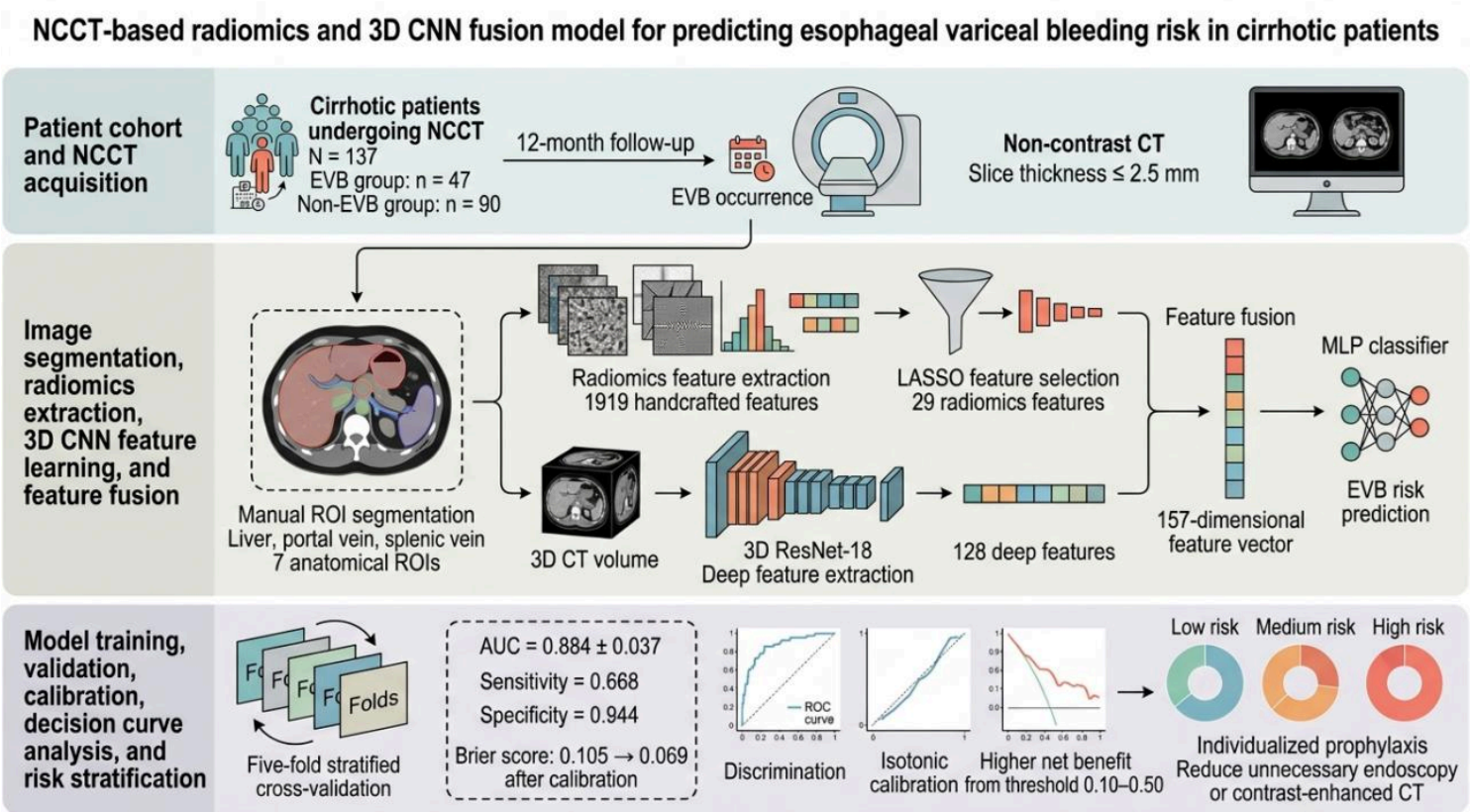


Figure 1

Workflow of the NCCT-based fusion model for predicting EGVB risk

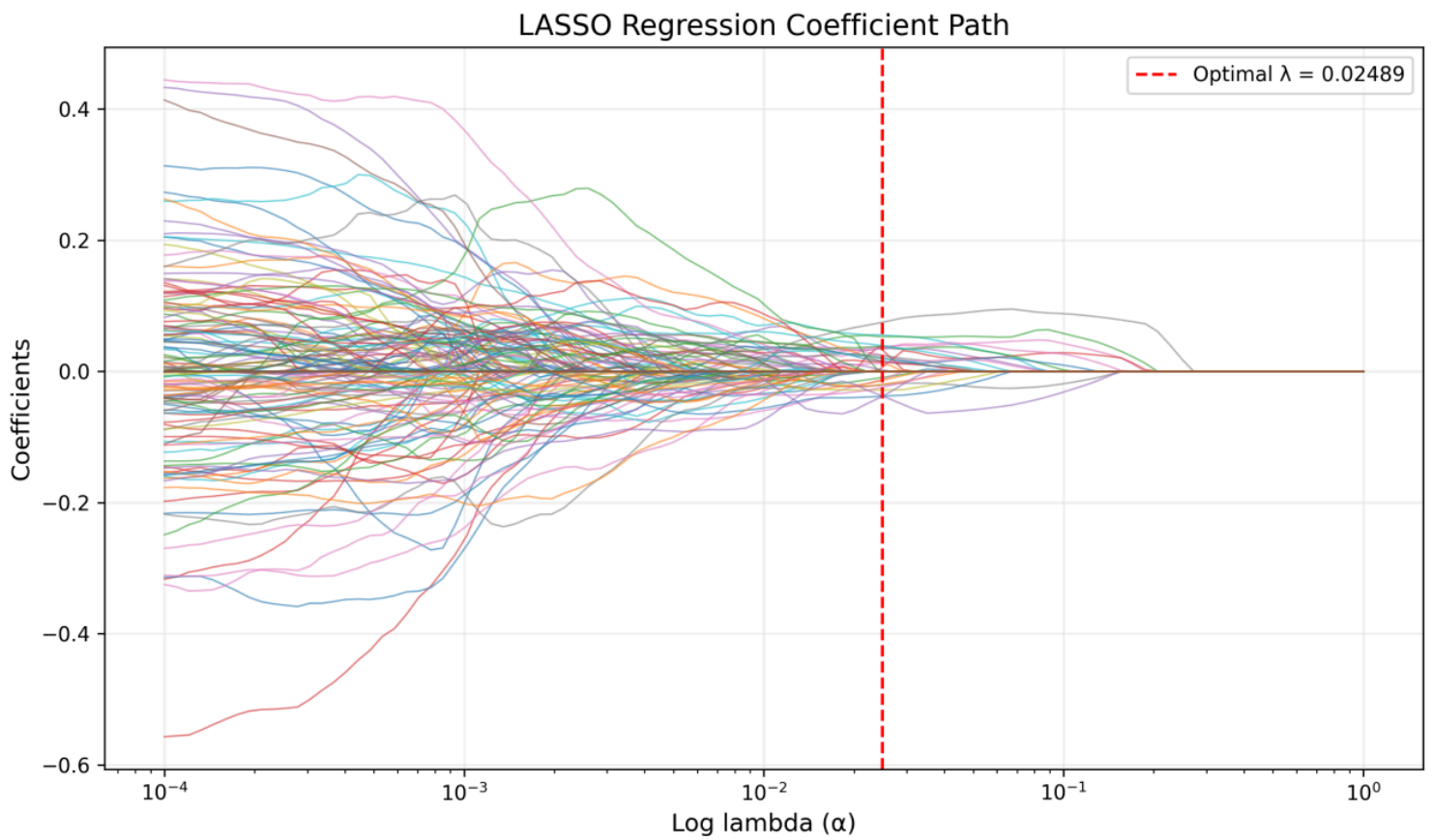


Figure 2

LASSO regression coefficient path. Each colored line represents the coefficient trajectory of one radiomics feature as the regularization parameter λ increases (log scale). The vertical red dashed line marks the optimal λ selected by ten-fold cross-validation, where 29 features retain non-zero coefficients.

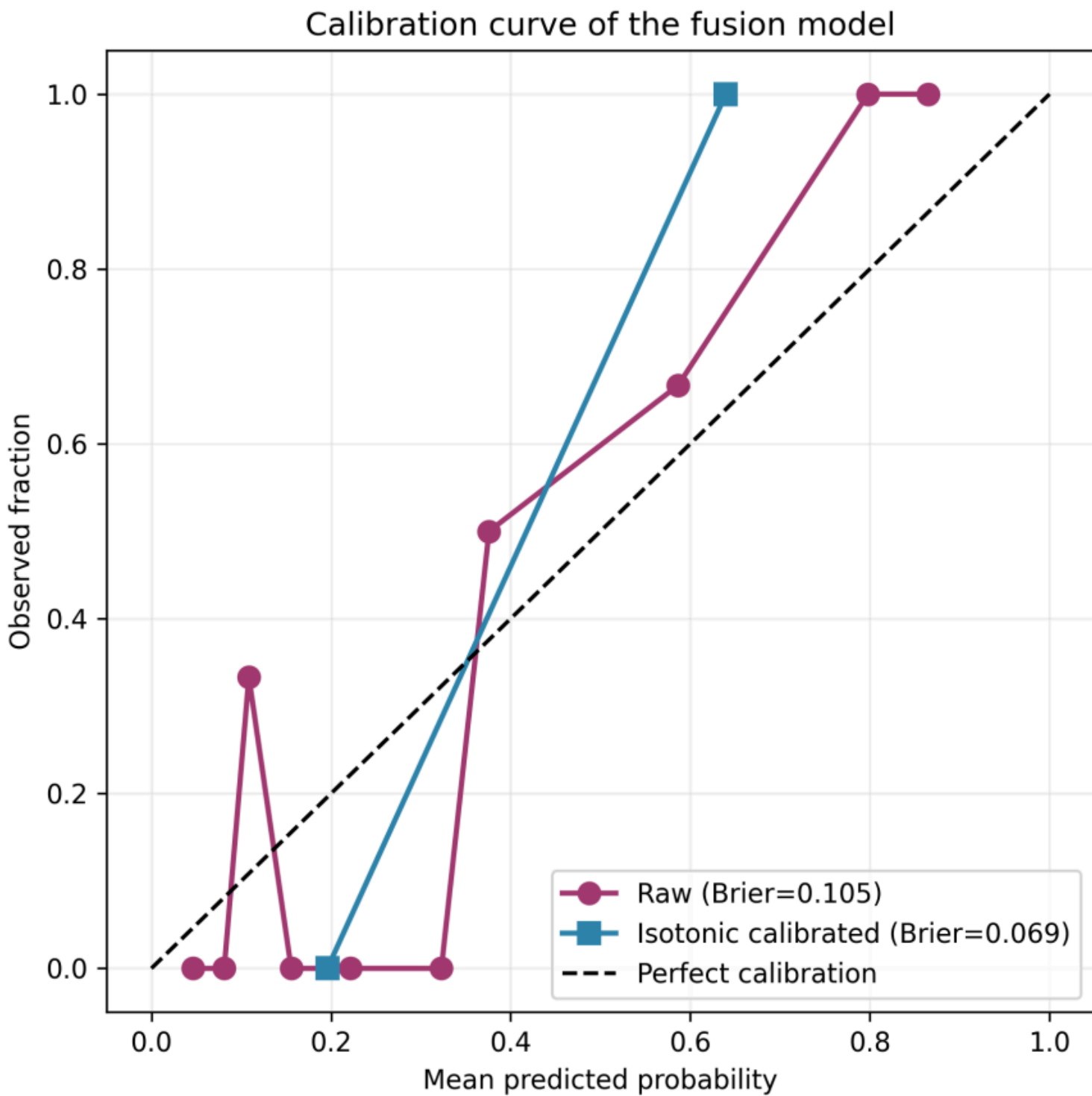


Figure 3

Calibration curve of the fusion model

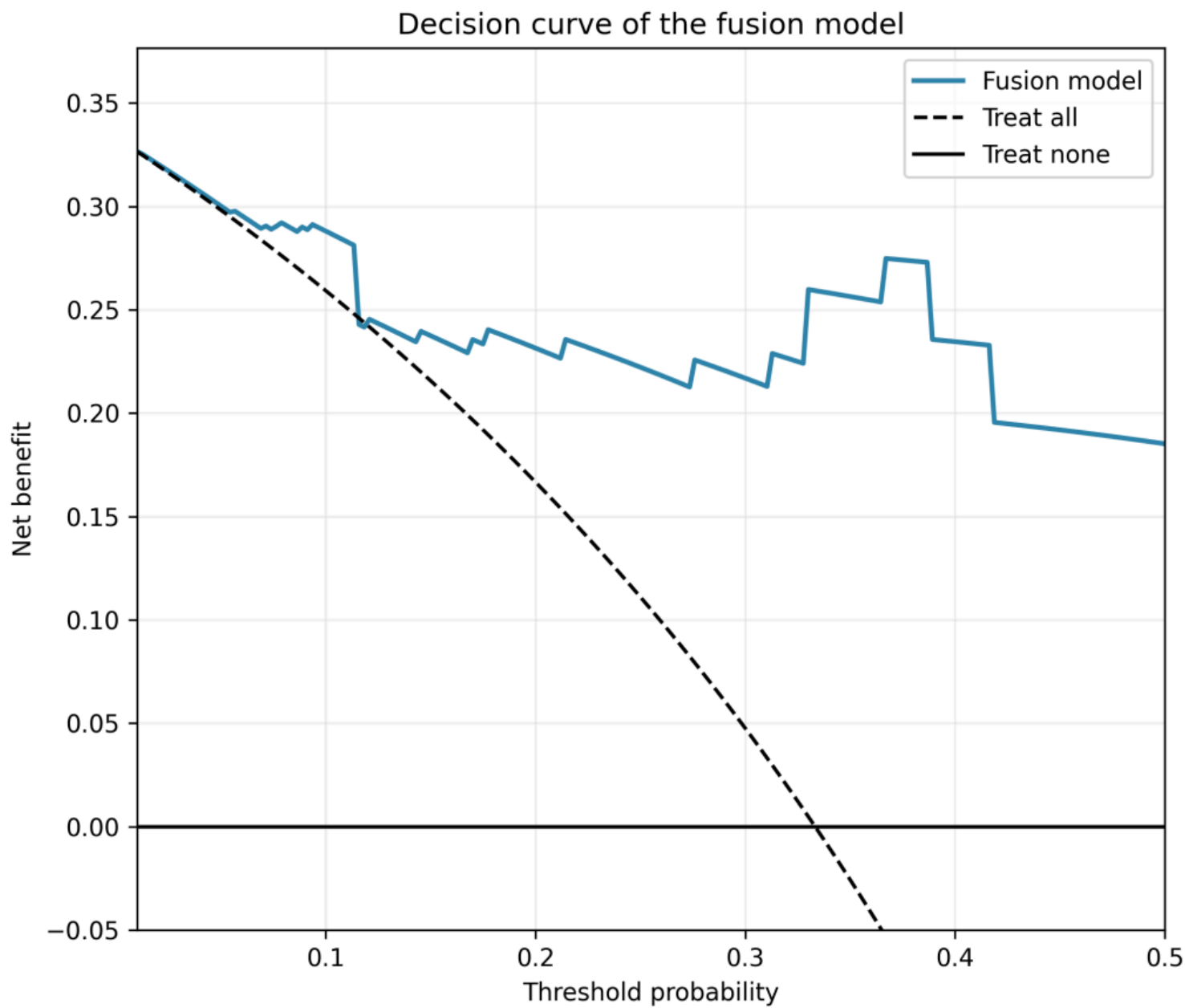


Figure 4

Decision curve for the fusion model

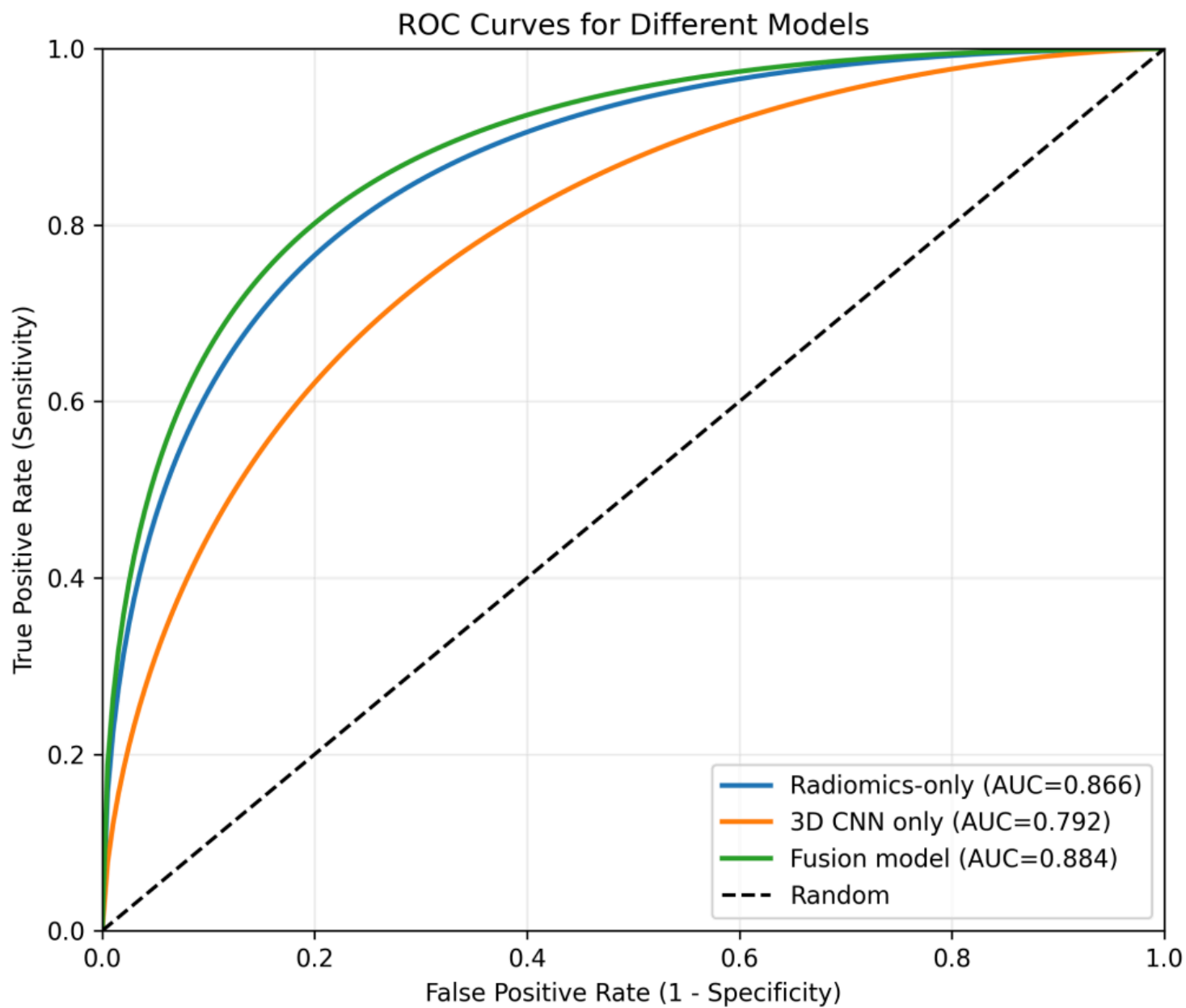


Figure 5

Receiver operating characteristic (ROC) curves of the three models.

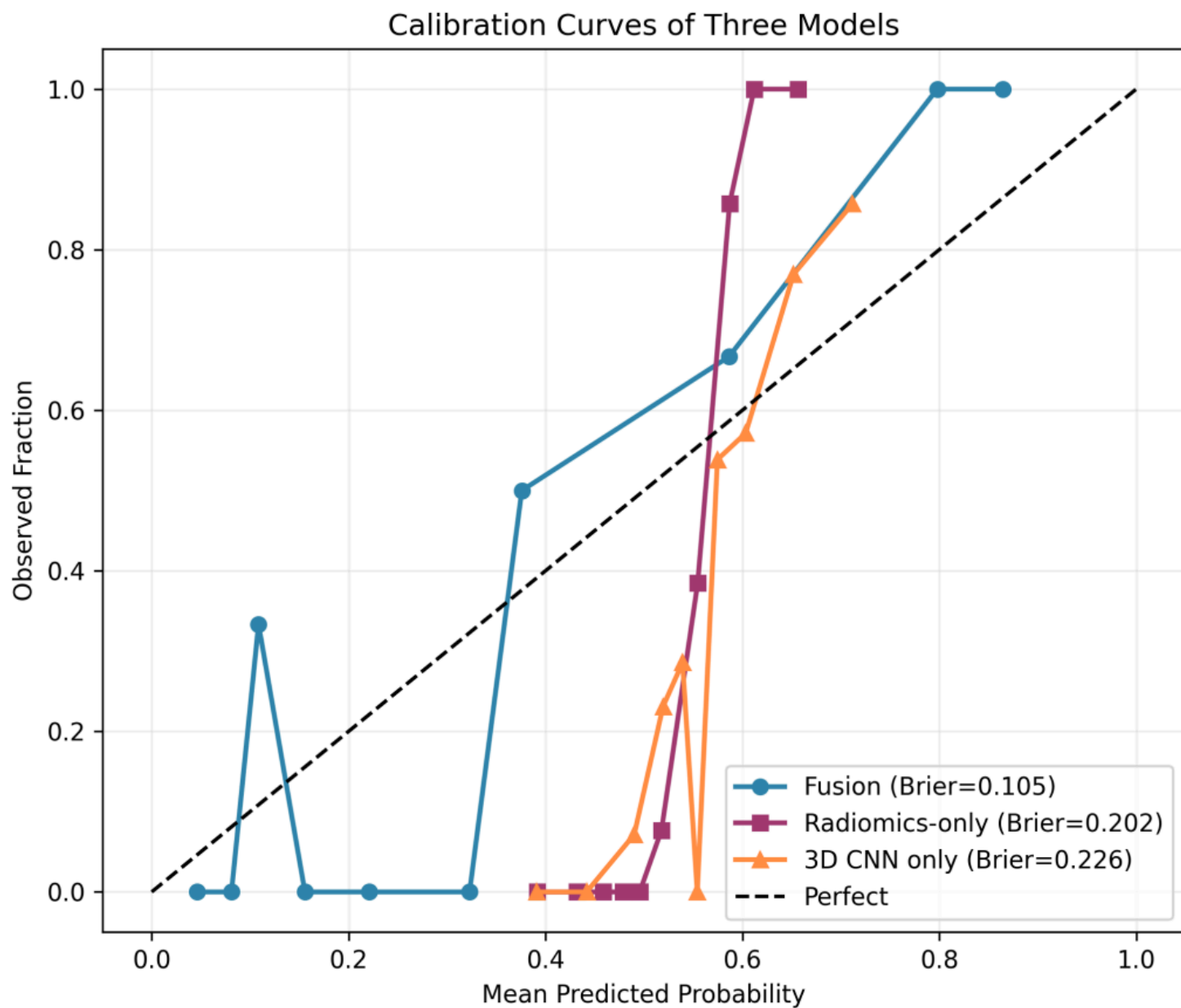


Figure 6

Calibration curves of the three models.

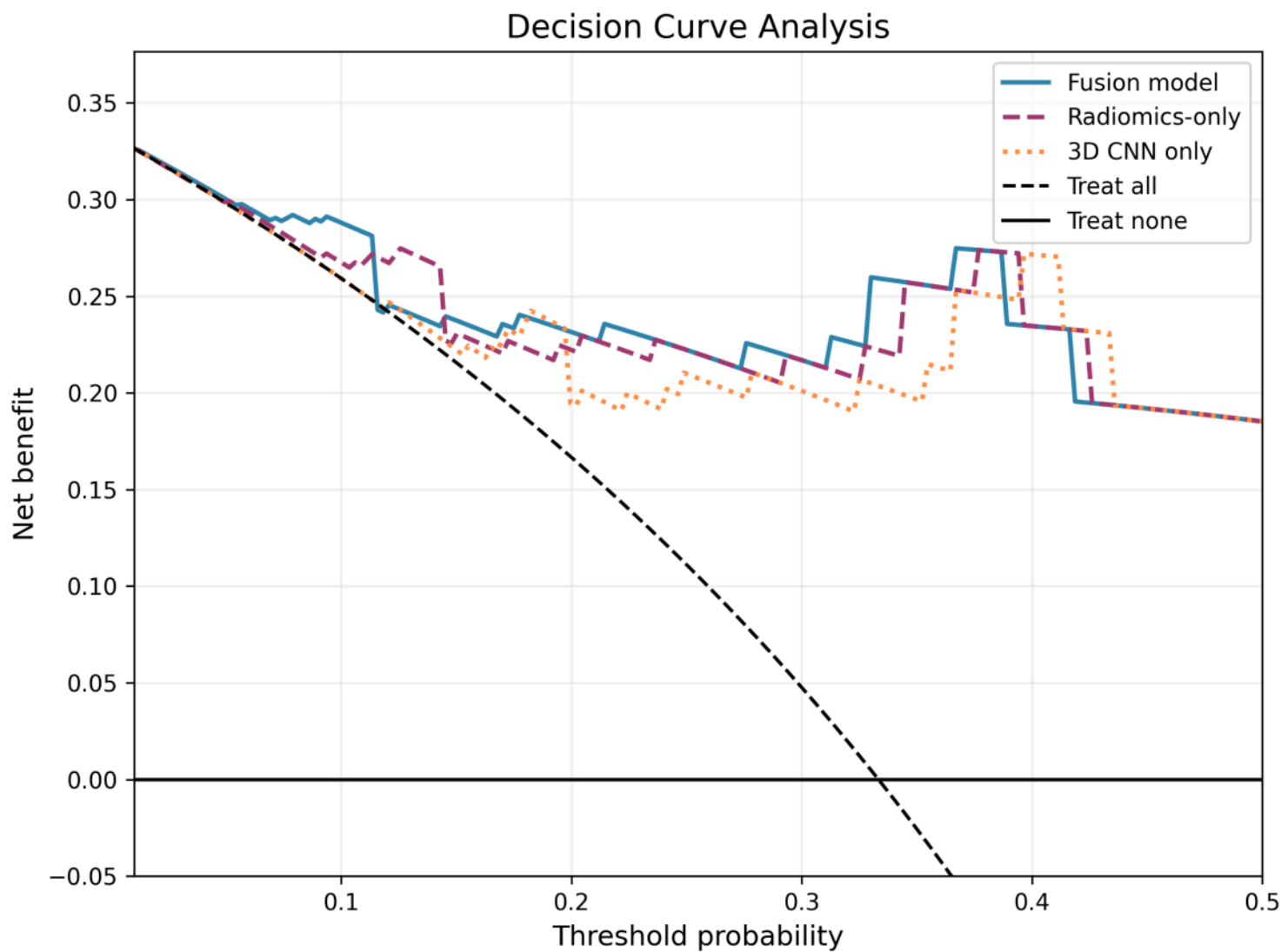


Figure 7

Decision curve analysis of the three models.