Paper ID:1407

EXPLAINABLE PREDICTION OF RENAL CELL CARCINOMA FROM CONTRAST-ENHANCED CT IMAGES USING DEEP CONVOLUTIONAL TRANSFER LEARNING AND THE SHAPLEY ADDITIVE EXPLANATIONS APPROACH

Fuchang Han, Shenghui Liao(✉), Siming Yuan, Renzhong Wu, Yuqian Zhao, Yu Xie

School of Computer Science and Engineering, Central South University, China

More about me

Contents

ICIP 2021

- Renal cell carcinoma (RCC) is the most common type of malignant kidney cancer.

- The visual identification method relies heavily on the experience and state of clinicians.

➢ It has become increasingly important to automatically and efficiently diagnose RCC.

ICIP 2021

Current state-of-the-art methods：

- The first class is conventional machine learning methods. Logistic regression (LR), kernel machines, tree-based methods, and ensemble methods (boosting, bagging and random forest).
- The second class is deep learning. These methods usually append one prediction branch to the final output of a convolutional neural network (CNN) framework.

➢ **However,** (1) these machine learning models remain mostly black boxes, and (2) these approaches lack multiscale feature extraction and fusion steps for better performance.

Challenges：

- The first aspect is limited clinical data.
- The second aspect is the inconsistent number of region of interest (ROI) segmentations.
- The third aspect is trust in the model and clinical utility evaluation.

ICIP 2021

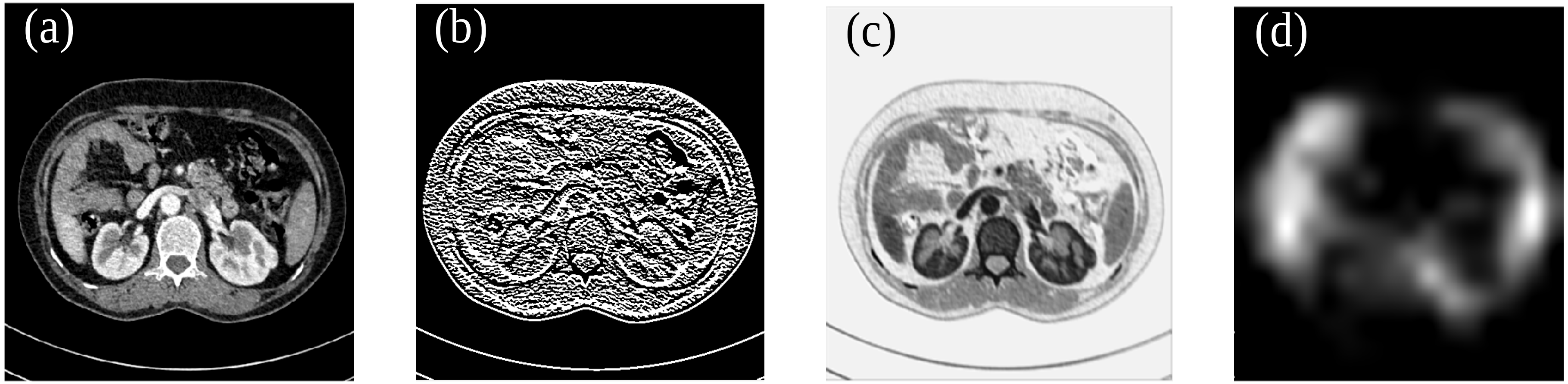


Fig. 1. Image comparison. (a) CT image. (b) Texture image. (c) Feature map from the shallow layer. (d) Feature map from the deep layer.

Motivation:

- We focus on deep convolutional transfer learning and the SHAP approach.
- Multiscale feature extraction is performed to obtain comprehensive features (texture features, deep features, and shallow features).
- A decision curve analysis (DCA) module is performed for the clinical utility evaluation.

ICIP 2021

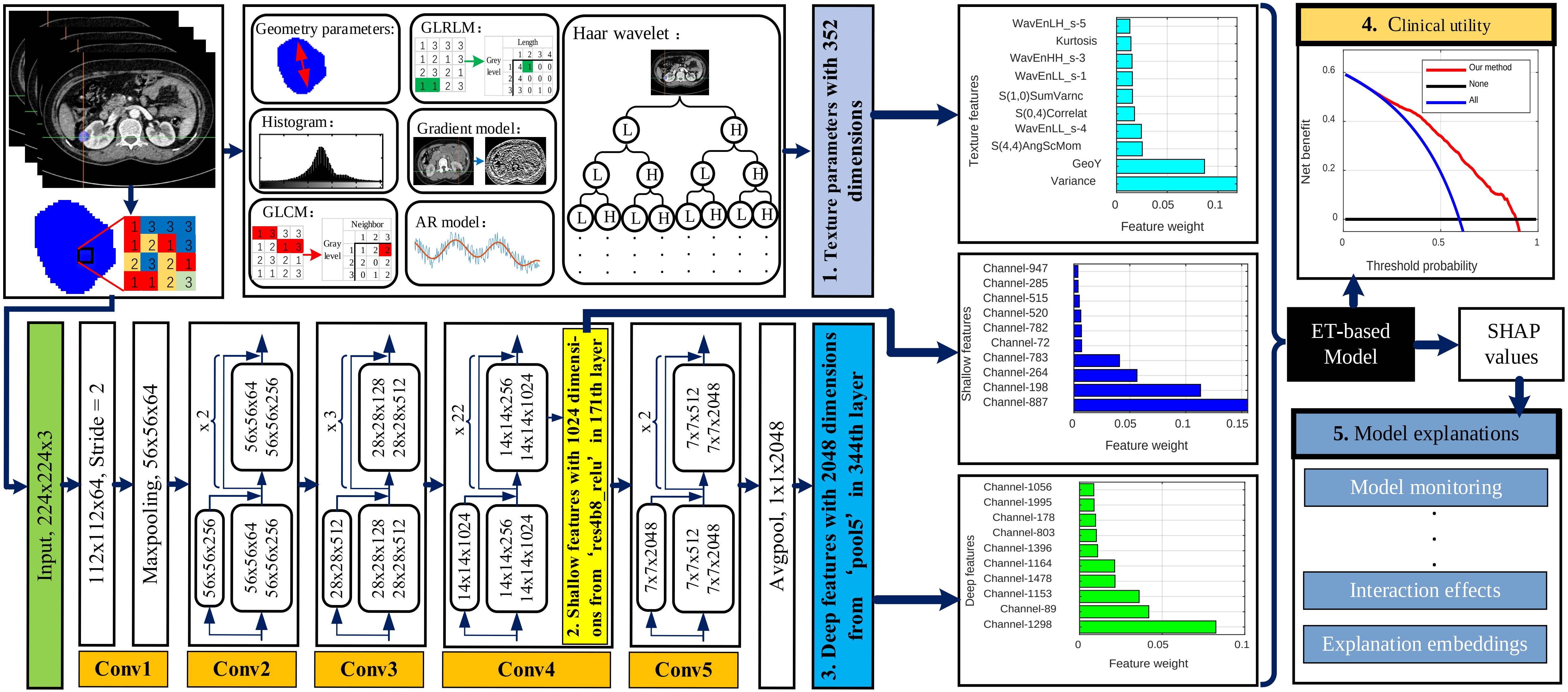


Fig. 1. Flowchart of the proposed method.

ICIP 2021

A. Image Preprocessing

- Preprocessing pipeline: (1) tumor stripping, (2) standardization, and (3) sampling.

B. Feature Engineering

- (1) Texture parameters with 352 dimensions; (2) Shallow features with 1024 dimensions from the res4b8_relu block in the 171st layer of ResNet-101; (3) Deep features with 2048 dimensions from the pool5 block in the 344th layer of ResNet-101. All features are evaluated by the mRMR method, and the high-importance features are selected.

C. Classification Models

- We build extra trees classifier (ET). The 30 best selected features are finally supplied to the ET model for final classification.

D. Model Explanations

- Shapley additive explanations (SHAP) is applied in this study to explain the output of machine learning model.

E. Clinical Utility Evaluation

- Decision curve analysis (DCA) provides a framework to evaluate the predictive models that incorporates the balancing of risks and benefits of treatment.

ICIP 2021

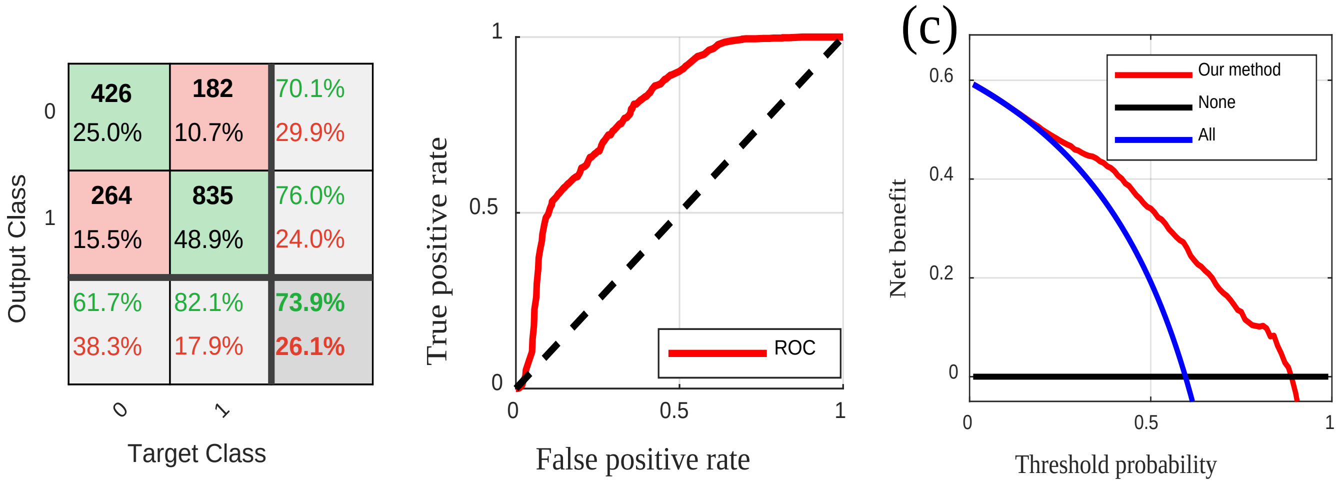
A. Data Processing and Experimental Results:

This study is a retrospective study approved by the institutional ethics review board at Hunan cancer hospital (Changsha, China).

Table 1. Characteristics and groups of patients.

Parameters	Benign	RCC
Patient & image number	29 & 2255	40 & 3105
Age (mean ± SD)	51.8 ± 9.06	49.2±10.6
Gender (male/femal)	7/22	18/22
Training set (p.no & i.no)	20 & 1,565	28 & 2,088
Testing set (p.no & i.no)	9 & 690	12 & 1,017

SD: standard deviation, p.no: the number of patient, i.no: the number of image.



The experimental results obtained by our model.
(a) Confusion matrix. (b) ROC curve. (c) Decision curve.

- Each CT image were resized to 224x224x3, and then three-stage feature extractions were performed.
- Our model achieves an accuracy of 73.87% and an area under the curve (AUC) of 0.8030.

ICIP 2021

B. Comparison with State-of-the-Art Methods：

Table 2. Performance comparison of some pretrained deep models with fine-tuned parameters.

Network	Depth	Layers	Image input size	Training loss	Training accuracy	Testing loss	Testing accuracy
DenseNet-201	201	709	224 by 224	0.0456	97.5890%	2.1841	54.4206%
GoogleNet	22	144	224 by 224	0.0627	97.0411%	1.5518	60.5647%
Inception-v3	48	316	299 by 299	0.0356	97.3973%	2.1221	60.0211%
ResNet-101	101	347	224 by 224	0.0368	97.3973%	2.0440	**64.9889%**
ShuffleNet	50	172	224 by 224	0.0368	97.4795%	3.0737	53.3216%
Xception	71	170	299 by 299	0.0375	97.4795%	1.5424	56.8155%

Table 3. Performance comparison of ResNet-101 with state-of-the-art ensemble tree methods.

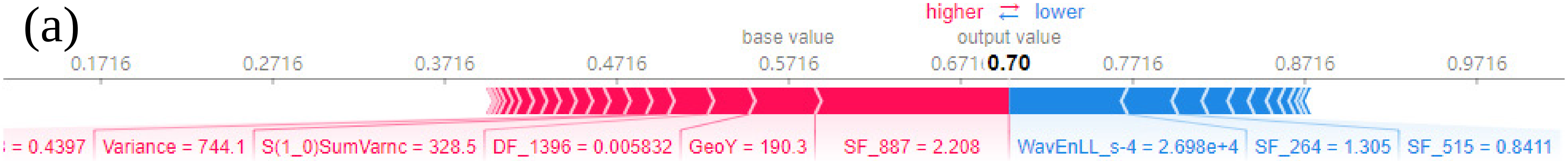
ResNet-101 + Trees	Training process		Testing process	
	Accuracy	AUC	Accuracy	AUC
DT	**0.9989**	**0.999998**	0.6608	0.6368
RF	0.9970	0.999974	0.6210	0.6810
GBC	0.9808	0.999097	0.6520	0.7346
XGBoost	0.9795	0.998949	0.6661	0.7764
CatBoost	**0.9989**	**0.999998**	0.6514	0.7806
Our method	**0.9989**	**0.999998**	**0.7387**	**0.8030**

- ResNet-101 yields better results than state-of-the-art pretrained CNN classification methods.
- It is obvious that the model built by ET classifier exhibits decent discriminating abilities.

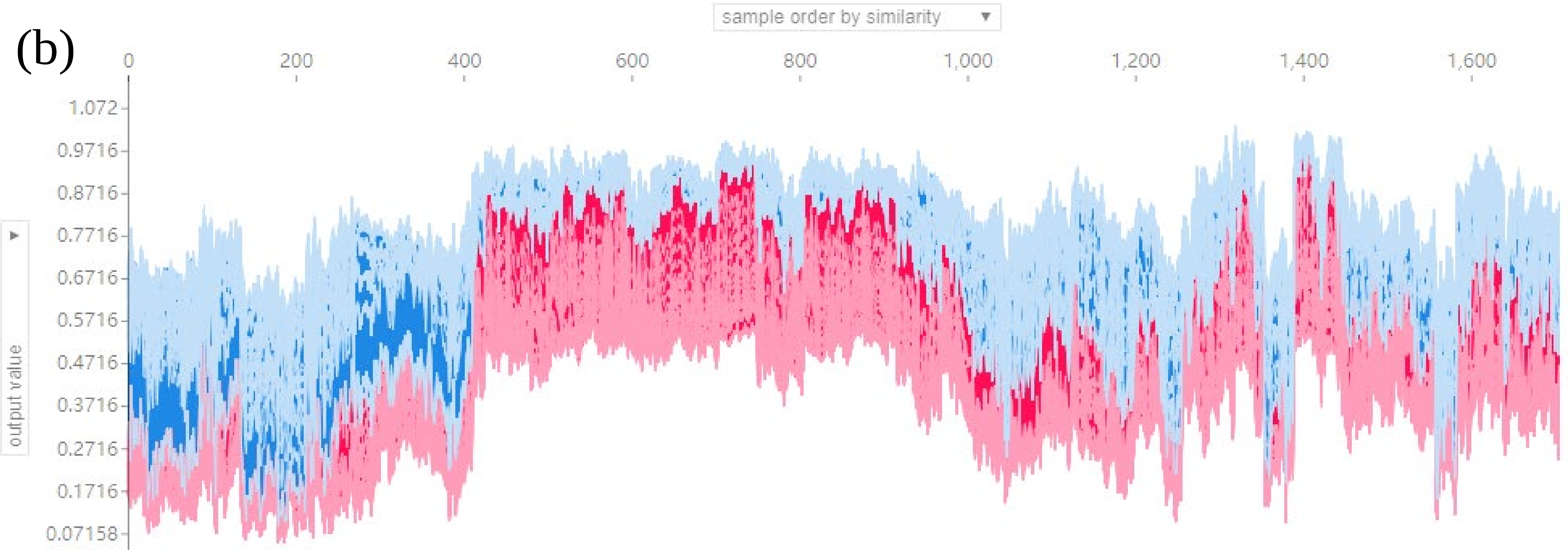
ICIP 2021

C. Model Explanation:

(a)

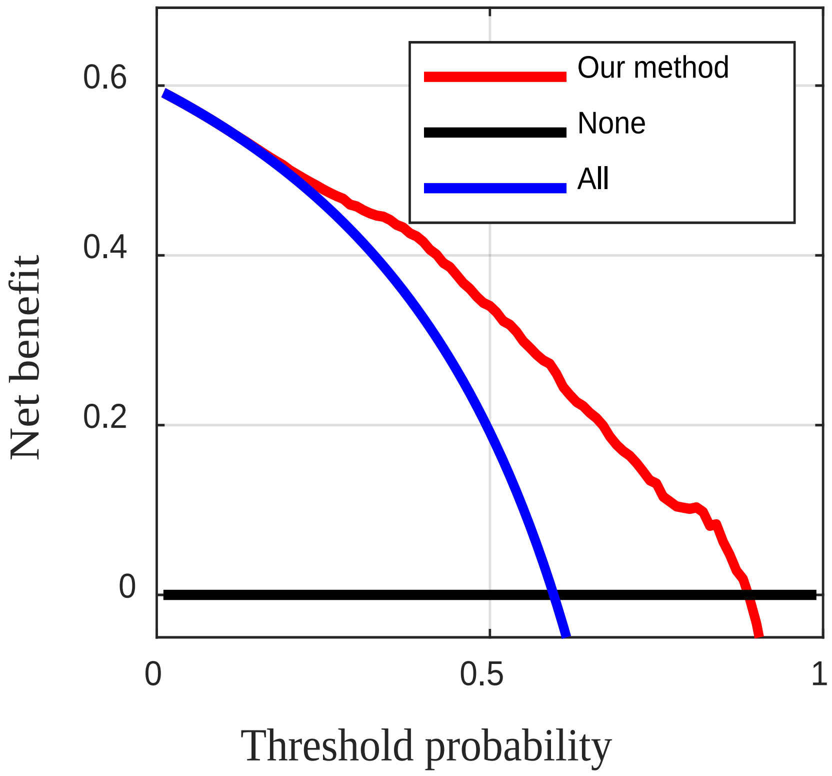


(b)



ICIP 2021

D. Clinical Utility Evaluation:



- If the threshold probability of a patient or doctor is approximately 20%, using our model to predict RCC adds more benefit than either the treat-all-patients scheme or the treat-none scheme.

ICIP 2021

Contribution：

- A multiscale feature extraction module;
- An attribute optimization module based on mRMR method;
- Appending a SHAP module to the framework to automatically and efficiently interpret the prediction of the model;
- A decision curve analysis (DCA) module is performed for the clinical utility evaluation.

➢ Our ET model achieves higher accuracies than the state-of-the-art deep CNN models and other ensemble tree methods.

Limitations:

➢ In future work, we will test the proposed explainable model with more labeled clinical datasets to expand the application range and improve the robustness and accuracy of the model.

➢ The authors are also working to explore the utilization of 3D CT information for providing further insights into the prediction mechanism of RCC.

THANKS FOR ALL