A Transformer-Based Model for Preoperative Early Recurrence Prediction of Hepatocellular Carcinoma with Muti-modality MRI

Gan Zhan[1](✉), Fang Wang[2], Weibin Wang[1], Yinhao Li[1], Qingqing Chen[2], Hongjie Hu[2], and Yen-Wei Chen[1]

[1] College of Information Science and Engineering, Ritsumeikan University, Kyoto, Japan
{gr0502vs,gr0342he}@ed.ritsumei.ac.jp,
yin-li@fc.ritsumei.ac.jp, chen@is.ritsumei.ac.jp

[2] Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China
{wangfang11,qingqingchen,hongjiehu}@zju.edu.cn

Abstract. Hepatocellular carcinoma (HCC) is the most common primary liver cancer which accounts for a high mortality rate in clinical, and the most effective treatment for HCC is surgical resection. However, patients with HCC are still at a huge risk of recurrence after tumor resection. In this light, preoperative early recurrence prediction methods are necessary to guide physicians to develop an individualized preoperative treatment plan and postoperative follow-up, thus prolonging the survival time of patients. Nevertheless, existing methods based on clinical data neglect information on the image modality; existing methods based on radiomics are limited by the ability of its predefined features compared with deep learning methods; and existing methods based on CT scans are constrained by the inability to capture the details of images compared with MRI. With these observations, we propose a deep learning transformer-based model on multi-modality MRI to tackle the preoperative early recurrence prediction task of HCC. Enlightened by the vigorous capacity of context modeling of the transformer architecture, our proposed model exploits it to dig out the inter-modality correlations, and the performance significantly improves. Our experimental results reveal that our transformer-based model can achieve better performance than other state-of-the-art existing methods.

Keywords: HCC early recurrence prediction · Multi-modality MRI · Transformer

G. Zhan and F. Wang—First authors.

Y. Zheng et al. (Eds.): ACCV 2022, LNCS 13848, pp. 185–194, 2023.
https://doi.org/10.1007/978-3-031-27066-6_13

1 Introduction

Hepatocellular carcinoma, which is also known as HCC, is one of the primary liver cancer that accounts for a high mortality rate in clinical. It is the fifth most common malignancy and leads to the second most common death related to cancer around the world [1], especially in East Asia and sub-Saharan Africa, HCC occupies 82% of liver cancer cases [2]. There are many mature treatment choices for patients with HCC including liver transplantation, targeted therapy, immunotherapy, transarterial chemoembolization, surgical resection, and radiofrequency ablation. Among these treatment choices, surgical resection is the first-line treatment choice and is widely recommended by the clinical practice guidelines for patients with a well-preserved liver function [3–5]. However, patients with HCC are still facing a huge risk of recurrence after surgery, the recurrence rate can reach about >10% in 1 year, 70–80% in 5 years after tumor resection [6]. And the overall survival time varies in patients with early recurrence and late recurrence, patients with late recurrence tend to live longer than patients with early recurrence [7]. Therefore, it is vital to identify those HCC patients at high risk of early recurrence, which can guide physicians to develop a preoperative individualized treatment plan and postoperative follow-up, thus prolonging the survival time of patients.

So far, there are many existing methods [8,10,14,25] being proposed for preoperative HCC early recurrence prediction after tumor resection; existing methods based on clinical data mainly utilize machine learning algorithms to project quantitative and qualitative variables to the prediction space. For example, Chuanli Liu et al. [8] construct 5 machine learning algorithms based on 37 patients' characteristics, and they select K-Nearest Neighbor as their optimal algorithm for HCC early recurrence prediction task after comparison. The biggest flaw in clinical-based methods is it provides limited clinical features, and it lacks the medical image information which could well reflect the heterogeneity of the tumor. Besides those subjective assessments of the lesion provided by physicians are irreproducible due to human bias, thus the performance of prediction has bad generalization. Radiomics is brought up by Gillies et al. [9] as a quantitative tool for feature extraction from medical images, it can save time for physicians on repetitive tasks and facilitate a non-invasive tool for feature extraction. Existing methods based on radiomics provide the image feature that reveals the heterogeneity of the tumor across HCC patients, and it could achieve promising results. For example, Ying Zhao et al. [10] extract 1146 radiomics features from each image phase, and radiomics models of each phase and their combination are constructed by the multivariate logistic regression algorithm to well complete the HCC early recurrence prediction task. The drawback of radiomics-based methods is that radiomics features are predefined, it does not directly serve our prediction task, and thus the performance of image information to our prediction task could not be fully tapped.

Recently, deep learning has yielded brilliant performance in the medical image area compared with conventional approaches [11]. The reason behind these successes is that deep learning can automatically explore robust and generalized image features directly related to the task without human intervention. And this process is mainly conducted by the convolutional neural networks [12], through the hierarchical structure, the convolutional neural network continuously combines low-level pixel features to obtain the ultimate high-level semantic feature that classification tasks demand. The disadvantage of deep learning methods is it needs a huge amount of data, which is normally not feasible for medical images. But with the technology of finetune [13], we now can also exploit the performance of neural networks on small sample medical datasets. Deep learning has been applied to the HCC early recurrence prediction task, Weibin Wang et al. [14] proposed the Phase Attention prediction model based on multi-phase CT (computed tomography) scans. But certain study [15] has shown that MRI(magnetic resonance imaging) is superior to CT scans in the detection of HCC given the preponderance of capturing the image details, especially the soft tissue of the tumor. Thus utilizing MRI has the advantage of image information over CT scans in our prediction task.

With these observations, we aim to develop a deep learning method based on MRI to complete our early recurrence prediction task on HCC. Compared with single-modality MRI, multi-modality MRI can observe the tumor and analyze the internal composition of the tumor in a more comprehensive way, which is more conducive to judging the degree of malignancy [16], thus we conduct our study on multi-modality MRI. Considering the unique characteristics of each modality in our prediction task, how to well combine them is the biggest challenge in our task. Enlightened by the vigorous capacity of context modeling of transformer model [17], we formulate each modality image feature as a token of the sequence, then we utilize transformer architecture to dig out the inter-modality correlations. Our contributions are as follows: (1) We propose a transformer-based model for the preoperative early recurrence prediction of HCC, which can effectively use MRI images for computer-aided diagnosis. (2) Our model utilizes the transformer architecture to efficiently combine image features from multiple MRI modality, and detailed experimental results show that it achieves superior performance to other existing methods. To the best of our knowledge, we are the first to propose a deep learning model based on multi-modality MRI to tackle the early recurrence prediction task for HCC.

2 Proposed Method

We propose an end-to-end deep learning method based on multi-modality MRI to tackle our early recurrence prediction task. Since portal venous (PV) phase and arterial (ART) phase are helpful for physicians to judge benign and malignant tumors, diffusion-weighted with $b = 1000\,\mathrm{s/mm^2}$ (DWI1) is useful to judge the water content and hemorrhage and necrosis of tumors, and outline of tumors is generally clearer on axial T2-weighted imaging with fat suppression(T2) [18,19]. We select them as our research subjects, and we propose our model based on these 3 modalities (PV phase and ART phase belong to the same modality).

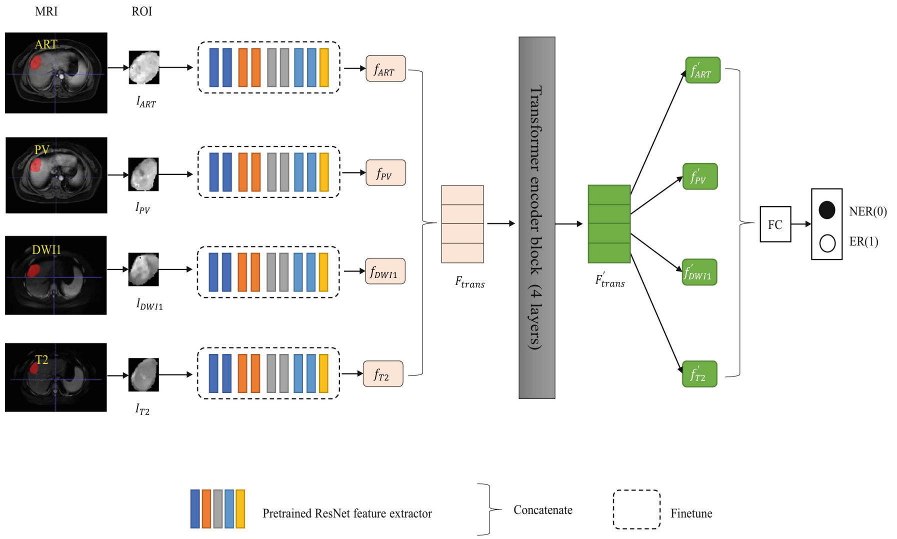

Fig. 1. The overall pipeline of our transformer-based model on multi-modality MRI. Image features are obtained through pre-trained ResNet feature extractor and transformer encoder block to complete our prediction task.

Figure 1 shows the overall pipeline of our proposed method. We feed the ROI of each modality into a pre-trained ResNet feature extractor and transformer encoder block to obtain the demand modality image feature, and then these image features will be combined and projected to the prediction space by an fully connected (FC) layer. In the following sections, we will introduce our prediction model in detail in terms of the pre-trained ResNet feature extractor and transformer encoder block.

2.1 Pre-trained ResNet Feature Extractor

ResNet [20] designed to solve the degradation problem of deep neural networks has achieved remarkable success in computer vision, so we select it to construct our modality image feature extractor. Since the last FC layer in ResNet serves as the classifier, we remove it to obtain our ResNet feature extractor for modality image feature extraction. Considering the small sample set in our study, we specifically select the 18-layer ResNet (ResNet18) pre-trained in ImageNet [21] as our feature extraction backbone, ResNet18 has few parameters, so it better fit our dataset size, and with the finetune technology, we can give our modality image feature extractor a very good initialization point.

Considering the pre-trained ResNet is originally designed and trained for natural images, which well accept input image with shape of 224 × 224 × 3 (both width and height are 224, number of channels is 3). For each modality, we selected the slice which has the largest tumor area and its two adjacent slices from each MRI modality, and we crop the ROI of the tumor region from these 3 slices, which can well represent the HCC charateristic of this input MRI modality volume. After we obtain this 3-channel ROI, we resize its width and height to 224 × 224, to obtain the input of 3 modalities $\{I_{PV}, I_{ART}, I_{DWI1}, I_{T2}\}$, then we feed them to our pre-trained ResNet feature extractor. After pre-trained ResNet feature extractor, the output features in the ART phase are denoted as f_{ART}, so are the f_{PV} in PV phase, f_{DWI1} in DWI1 modality and f_{T2} in T2 modality, and shape of each modality image feature, that we obtained before transformer encoder block, are all $1 \times 1 \times 512$.

2.2 Transformer Encoder Block

Transformer model [17] was first designed for natural language processing tasks, which accept tokens in the sequence and utilize the self-attention mechanism (Eq. 1) to dig out the inter-token correlations. Transformer has been widely used in computer vision, for example, images are divided into a collection of patches in vision transformer [22], and each patch could be served as the token in the image sequence, then the semantic of this image could be formulated as the composition of tokens and token grammar. Inspired by these intuitions, we formulate each modality image feature obtained in the pre-trained ResNet feature extractor as a token in the modality sequence, and inter-modality correlations serve as the modality grammar for our prediction task. We construct our transformer encoder block with 4 transformer encoder layers, and we first concatenate the 3 modality image features $\{f_{ART}, f_{PV}, f_{DWI1}, f_{T2}\}$ to obtain F_{trans}, then through linear projections, the concatenated modality image features are mapped to the key(K) vector, query(Q) vector and value(V) vector in self-attention described below:

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

Then the inter-modality correlations can be calculated by the dot product between Q vector and K^T vector, after the softmax operation scale the attention score value to 0–1, it will multiply the V, that is to encode the modality grammar into modality tokens, due to the large values of dk, the dot products grow large in magnitude, pushing the softmax function into regions where it has extremely small gradients, so we scale the dot products by $\frac{1}{\sqrt{d_k}}$, hereby we obtain the image feature F'_{trans}, and we split it by the token dimension to obtain $\{f'_{ART}, f'_{PV}, f'_{DWI1}, f'_{T2}\}$ as the ultimate image features. And finally, we concatenate these image features by the channel dimension and utilize one FC layer to complete our prediction task.

Table 1. Dataset arrangement of 5-fold cross-validation.

Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Total
57	58	58	58	58	289

3 Experiments

3.1 Patient Selection

From 2012 to 2019, 659 HCC patients from Run Run Shaw Hospital Affiliated to Medical College of Zhejiang University, who has undergone liver resection, pathologically confirmed as hepatocellular carcinoma and received enhanced MRI examination before surgery were recruited in this retrospective study. Under the following exclusion criteria: (1) patient received other anti-tumor treatments before surgery, such as TACE, RFA; (2) the interval between preoperative MR examination and surgery is more than 30 days; (3) image quality is not good, and (4) less than 2 years of follow-up after surgery; a total of 289 patients are included in our study. Patients who have a recurrence in 2 years are denoted as early recurrence (ER) [23], and patients who have a recurrence of more than 2 years or no recurrence are denoted as Non-early recurrence (NER) in our study.

3.2 Dataset Preparation and Metrics

For the MRI data, considering the affection of different types of artifacts to the MRI, we mainly use the preprocessing method described by Jose Vicente Manjon [24] on our multi-modality MRI, which includes denoise, bias field correction, resampling and normalization.

To fairly measure the performance of our proposed method, we use the 5-fold cross-validation on our 289 patients, we split it into 5 groups along the timeline, the arrangement of dataset is in Table 1.

We select AUC (area under the ROC curve), ACC (accuracy), SEN (sensitivity), SPE (specificity), PPV (positive predictive value), and NPV (negative predictive value) to comprehensively evaluate the model performance, among which, AUC, ACC, and SEN are the most important indicator in our task, especially SEN, we normally use SEN to validate the model's ability on identifying patients, and it is of great clinical value in our research if our model could identify more positive patients. Since we conduct 5-fold cross-validation, we calculate the average value of each metric as the indicator value; Cross-entropy loss is our model's loss function, and for every fold, we train 50 epochs, and we select the checkpoint that has the minimal training loss as our trained model.

3.3 Ablation Study

Selecting early fusion or late fusion as our baseline method for multiple modality image inputs is controversial, thus it is necessary to do this ablation on our task.

Table 2. Ablation on baseline method.

Model	AUC	ACC	SEN	SPE	PPV	NPV
Early fusion	0.6447	0.6469	0.4031	0.8101	0.5777	0.6926
Late fusion	**0.6829**	**0.6365**	**0.4307**	**0.7639**	**0.5237**	**0.6900**

Table 3. Ablation on transformer encoder block.

Model	Transformer	AUC	ACC	SEN	SPE	PPV	NPV
Baseline		0.6829	0.6365	0.4307	0.7639	0.5237	0.6900
Trans-based	✓	**0.6907**	**0.6782**	**0.4360**	**0.8296**	**0.5944**	**0.7117**

The comparison result between early fusion and late fusion is shown in Table 2. For the early fusion, we simply concatenate the 3 modality ROI images by the channel dimension, and feed it into a single pre-trained ResNet18 model, since the new input has 12 channels, we create a new convolution layer that would accept this input to replace the first convolution in Pre-trained ResNet18. And to fit our binary task, we replace the last FC layer which outputs 1000 neurons with a new FC layer outputting 1 neuron, and we will use the value 0.5 as the threshold to do the binary prediction. For the late fusion, we utilized 4 pre-trained ResNet feature extractors to extract image features from each modality, then we simply concatenate these 3 modality image features and utilize 1 FC layer to complete our binary prediction task.

We can see in Table 2 that the late fusion model is more appropriate for our task. In the indicators AUC and SEN, which we care about most, the late fusion model obtains a large gain compared with the early fusion model, so we select it as our baseline method. And based on the late fusion model, we add a transformer encoder block to construct our transformer-base model, to measure the effectiveness of our transformer encoder block, we also conduct a detailed ablation study, the results of which are shown in Table 3.

As we can see, when we dig out the inter-modality correlations using the transformer encoder block, our transformer-based model (denoted as Trans-based in Table 3) can achieve better performance on all metrics, which strongly proves the effectiveness of our transformer encoder block, and it is indeed necessary to dig out the inter-modality correlations in our study.

3.4 Comparison with Existing Methods

To further prove the effectiveness of our proposed method, we compare our proposed method with existing methods on the HCC early recurrence prediction task, including the radiomics-based method and deep learning-based methods. The comparison result is shown in Table 4.

Table 4. Comparison with existing methods

Model	AUC	ACC	SEN	SPE	PPV	NPV
Radiomics [10]	0.6861	0.6364	0.3337	0.8252	0.5502	0.6721
PhaseNet [25]	0.6450	0.6643	0.3148	0.8770	0.6068	0.6820
DPANet [14]	0.6818	0.6711	0.4308	0.8105	0.5817	0.7075
Trans-based	**0.6907**	**0.6782**	**0.4360**	**0.8296**	**0.5944**	**0.7117**

We can see from the Table 4, established on the multi-modality MRI only, our transformer-based model can achieve better performance than radiomics-based method (denoted as Radiomics in Table 4) and deep learning method (PhaseNet [25] and DPANet [14]).

4 Conclusion

In this paper, we construct our transformer-based model on the multi-modality MRI to tackle the preoperative early recurrence prediction of HCC patients. Our proposed model formulates each modality image feature as a token in the sequence and obtains the modality correlations related to our prediction task by the transformer encoder block. Detailed experiments reveal the effectiveness of our proposed method and it could achieve better performance compared with existing methods, including state-of-the-art radiomics-based method and deep learning-based method.

Acknowledgments. This work was supported in part by the Grant in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant Nos. 20KK0234, 21H03470, and 20K21821, and in part by the Natural Science Foundation of Zhejiang Province (LZ22F020012), in part by Major Scientific Research Project of Zhejiang Lab (2020ND8AD01), and in part by the National Natural Science Foundation of China (82071988), the Key Research and Development Program of Zhejiang Province (2019C03064), the Program Co-sponsored by Province and Ministry (No. WKJ-ZJ-1926) and the Special Fund for Basic Scientific Research Business Expenses of Zhejiang University (No. 2021FZZX003-02-17).

References

1. Elsayes, K.M., Kielar, A.Z., Agrons, M.M.: Liver imaging reporting and data system: an expert consensus statement. J. Hepatocel. Carcinoma **4**, 29–39 (2017)
2. Zhu, R.X., Seto, W.K., Lai, C.L.: Epidemiology of hepatocellular carcinoma in the Asia-Pacific region. Gut Liver **10**, 332–339 (2016)
3. Thomas, M.B., Zhu, A.X.: Hepatocellular carcinoma: the need for progress. J. Clin. Oncol. **23**, 2892–2899 (2005)
4. association, E.: EASL clinical practice guidelines: management of hepatocellular carcinoma. J. Hepatol. **69**, 182–236 (2018)

5. Marrero, J.A., et al.: Diagnosis, staging and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the study of liver diseases. Hepatology **68**, 723–750 (2018)
6. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **68**, 394–424 (2018)
7. Cheng, Z., Yang, P., Qu, S.: Risk factors and management for early and late intrahepatic recurrence of solitary hepatocellular carcinoma after curative resection. HPB **17**, 422–427 (2015)
8. Liu1, C., Yang, H., Feng, Y.: A k-nearest neighbor model to predict early recurrence of hepatocellular carcinoma after resection. J. Clin. Translat. Hepatol. **10**, 600–607 (2022)
9. Gillies, R., Kinahan, P.E., Hricak, H.: Radiomics: images are more than pictures, they are data. radiology. Radiology **278**, 563–577 (2015)
10. Zhao, Y., Wu, J., Zhang, Q.: Radiomics analysis based on multiparametric MRI for predicting early recurrence in hepatocellular carcinoma after partial hepatectomy. J. Magn. Reson. Imaging **53**, 1066–1079 (2021)
11. Litjens, G., Kooi, T., Bejnordi, B.E.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
12. Krizhevsky, A., Sutskever, I., HintonImagenet, G.E.: ImageNet classification with deep convolutional neural networks. Communications ACM **60**, 84–90 (2017)
13. Kolesnikov, A., Beyer, L., Zhai, X., Puigcerver, J., Yung, J., Gelly, S., Houlsby, N.: Big Transfer (BiT): general visual representation learning. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12350, pp. 491–507. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58558-7_29
14. Wang, W., et al.: Phase attention model for prediction of early recurrence of hepatocellular carcinoma with multi-phase CT images and clinical data. Front. Radiol. **8** (2022)
15. Burre, M., et al.: MRI angiography is superior to helical CT for detection of HCC prior to liver transplantation: an explant correlation. Hepatology **38**, 1034–1042 (2003)
16. Armbruster, M., et al.: Measuring HCC tumor size in MRI-the sequence matters! Diagnostics **11** (2002)
17. Vaswan, A., et al.: Attention is all you need. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA (2017)
18. Lee, Y., et al.: Benign versus malignant soft-tissue tumors: differentiation with 3t magnetic resonance image textural analysis including diffusion-weighted imaging. Investig. Magn. Resonance Imaging **25**, 118–128 (2021)
19. Chartampilas, E., Rafailidis, V., Georgopoulou, V., Kalarakis, G., Hatzidakis, A., Prassopoulos, P.: Current imaging diagnosis of hepatocellular carcinoma. Cancers **14**, 3997 (2022)
20. He, K., Zhang, X., Ren, S.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
21. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Li, F.-F.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)
22. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

23. Xing, H., Zhang, W.G., Cescon, M.: Defining and predicting early recurrence after liver resection of hepatocellular carcinoma: a multi-institutional study. HPB **22**, 677–689 (2020)
24. Manjon, J.V.: MRI preprocessing. In: Imaging Biomarkers, pp. 53–63 (2017)
25. Wang, W., et al.: Deep learning-based radiomics models for early recurrence prediction of hepatocellular carcinoma with multi-phase CT images and clinical data. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4881–4884 (2019)