Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ30564

Deep Learning-Based Integrated Multi-Omics and Radiomics Model for Diagnosis of Kidney Renal Clear Cell Carcinoma

Fucheng Jia

School of Mathematics and Statistics, Shandong University, Weihai, China
202300820013@mail.sdu.edu.cn

Abstract. Kidney renal clear cell carcinoma (KIRC) is a highly aggressive malignancy that accounts for 70–80% of renal cell carcinomas. Current diagnostic approaches for KIRC face several limitations that compromise early detection capabilities. Conventional imaging modalities frequently fail to identify early-stage lesions due to their subtle radiographic characteristics. Molecular diagnostic methods, though promising, mostly rely on unimodal gene data, making it harder for further improvements in accuracy, particularly in the case of early-stage prediction. To address these issues, this paper proposed an attention-based deep learning framework with one-dimensional convolutional neural network (genomics) and three-dimensional residual network (imaging) branches. After highly relevant genes were identified through feature engineering methods, they were fused with representations extracted from CT scans using a 3D Residual Network. The framework then employed a dynamic cross-modal attention mechanism to generate a fused vector, which then goes to a fully connected layer to make final decisions. A series of experiments demonstrate that the integrated model achieves the highest prediction performance in all four clinical stages.

Keywords: Multimodal Deep Learning, Kidney Clear Cell Carcinoma, Early Diagnosis, Convolutional Neural Network

1. Introduction

The kidneys are important organs responsible for critical physiological functions, including filtering metabolic waste, regulating fluid and electrolyte balance, and maintaining blood pressure. Renal dysfunction can lead to a series of systemic health problems, such as chronic kidney disease, cardiovascular complications, and end-stage renal failure, which significantly affect the quality of life and survival rate of patients. As the main subtype of cancer, renal cell carcinoma (RCC) has seen a steady increase in its incidence rate in recent years, which plays an important role in the global cancer burden. Therefore, early detection of renal malignancies is not only crucial for improving treatment outcomes, but also for reducing the socio-economic costs associated with late-stage disease management.

In the research, we took kidney renal clear cell carcinoma (KIRC) into consideration for the study. Kidney renal clear cell carcinoma (KIRC) is the most common renal malignancy, accounting for approximately 70-80% of all cases. [1] Its 5-year survival rate exceeds 90% for early-stage patients but drops to less than 10% in advanced stages, underscoring the urgency of enhancing early

DOI: 10.54254/2755-2721/2026.TJ30564

diagnostic accuracy to improve patient prognosis [2]. Its histological feature is clear cytoplasm, and its genetic feature is frequent mutations in the VHL gene, leading to dysregulation of hypoxia inducible factors, promoting angiogenesis and tumor progression. Known risk factors include smoking, obesity, hypertension, and genetic syndromes such as von Hippel Lindau disease. According to the tumor, node, matastasized (TNM) staging, there are four clinical stages of KIRC: stage I, stage II, stage III, and stage IV. The disease usually shows few symptoms in the early stages (I and II), leading to a large proportion of cases being diagnosed only in the late or metastatic stages. The sharp decline in five-year survival rate from over 90% for local diseases to below 10% for metastatic diseases highlights the urgent need for more sensitive and reliable early diagnostic tools.

2. Literature review

Since the development in high-throughput sequencing technology, large quantities of genomic data have been obtained and analyzed. Due to its high dimension and information density, researchers have proposed several algorithms to select features to support diagnosis while diminishing the disturbance of the unnecessary and noise information it included.

In the field of genomics, researchers have developed several feature selection algorithms to improve diagnostic accuracy. Das et al. [3] proposed an innovative feature selection method using Support Vector Machine (SVM) and T-statistic. Wu et al. [4] designed a gene selection algorithm (CRIA) based on gene correlation, redundancy, and interaction analysis before applying machine learning methods such as LightGBM and CatBoost, achieving more accurate cancer-type prediction.

In radiomics, deep learning techniques have been widely applied to the automatic analysis of images such as X-rays and CT scans. For instance, the work of Rahman et al. [5] presents a detailed examination of chest X-rays, demonstrating the effectiveness of deep learning in detecting diseases like tuberculosis. In contrast to deep learning approaches, Okmen et al. [6] demonstrated the efficacy of hand-crafted radiomic features for TNM staging of Clear Cell Renal Cell Carcinoma (CCRCC), achieving an 85.4% prediction accuracy by extracting and selecting critical features from manually defined tumor regions in CT images. Although these unimodal approaches have made significant progress, their accuracy faces bottlenecks for further improvement.

3. Methodology

This part explains the steps taken in the experiment. To conduct the experiment on a multimodal basis, we collected the needed genomic data from [7] and responding imaging data from [8]. Figure 1 is a concise summarization of the process.

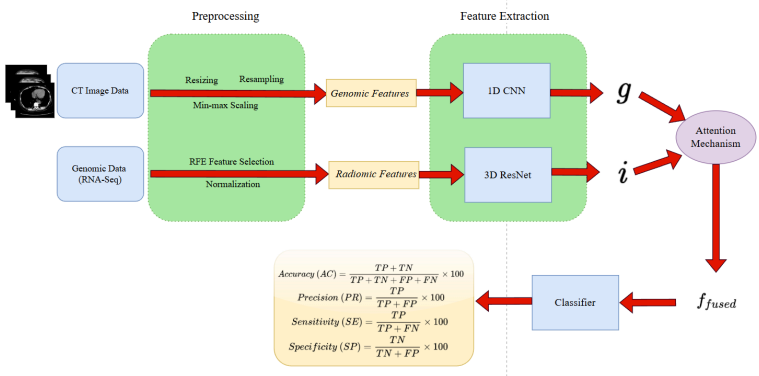


Figure 1. Proposed examination process

DOI: 10.54254/2755-2721/2026.TJ30564

3.1. Data collection and preprocessing

The experiments were conducted on radiomic data collected from The Cancer Imaging Archive, a large archive of medical images of cancer accessible for public download, and genomic data from The Cancer Genome Atlas. The TCGA dataset contains 537 samples of RNA-sequencing data with over 5000 files, among which 267 cases with corresponding radiomic data were finally selected to conduct cross-modal experiments.

For genomics data, the pipeline included quality control, differential expression analysis using DESeq2, and the selection of the most crucial 512 gene features with the Recursive Feature Elimination (RFE) algorithm. For CT image data, the steps involved resizing, resampling and min-max scaling.

3.2. Deep feature extraction

The feature extraction stage utilized a dual-branch deep neural network architecture. The genomic branch processed the preprocessed gene expression vector through a 1D Convolutional Neural Network (1D CNN) to capture local patterns among genes through convolution and pooling operations, producing a 128-dimensional feature vector as output. The imaging branch fed the preprocessed 3D CT image patches into a modified 3D ResNet-18 network, which learned three-dimensional spatial hierarchical features of the tumor, outputting a 512-dimensional deep image feature vector through global average pooling. These two branches automatically learned the most representative deep features from their respective modalities.

3.3. Cross-modal attention fusion

The fusion stage integrates the produced features from the two branches with a dynamic cross-modal attention mechanism. First, the 128-dimensional feature vector from the genomic branch $\mathbf{g} \in \mathbb{R}^{128}$ is first projected to $\mathbf{g}'$ via a fully connected layer. This gives it the same dimension as the image branch feature $\mathbf{i} \in \mathbb{R}^{512}$ to ensure they can be operated on in the same space. The aligned feature vectors $[\mathbf{i}, \mathbf{g}']$ are then concatenated to form a 1024-dimensional joint representation. This joint representation is then fed into a small neural network (the attention network), where raw energy score e is calculated for each modality through ReLU and two linear transformation layers. After obtaining the energy scores $[e_i, e_g]$, they are normalized via the Softmax function to generate the final attention weights. These weights reflect the relative importance the model assigns to each modality's information when making the final decision:

$$\alpha = \frac{\exp(e_i)}{\exp(e_i) + \exp(e_g)}, \beta = \frac{\exp(e_g)}{\exp(e_i) + \exp(e_g)}$$

where α and β represent the attention weights for the image and genomic modalities, satisfying $\alpha + \beta = 1$.

Finally, the original high-dimensional features are fused using a weighted sum with the generated weights, producing the fused feature vector $\mathbf{f}_{fused}$:

DOI: 10.54254/2755-2721/2026.TJ30564

$$\mathbf{f}_{fused} = \alpha \cdot \mathbf{i} + \beta \cdot \mathbf{g}^{'}$$

This 512-dimensional fused feature vector integrates complementary information from both modalities and quantifies the model's confidence for each information source. This vector is then fed into the final classification layer for stage prediction.

3.4. Fully connected layer classifier

The classification stage uses a fully connected neural network layer to map the vector of merged features to the final four-class diagnosis. This network consists of two hidden layers with ReLU activation for non-linear transformation, followed by an output layer with four neurons corresponding to the clinical stages (I-IV). A Softmax activation function is applied to the output layer to generate a probabilistic distribution among the four classes. The class with the highest probability is selected as the final prediction, enabling accurate staging of clear cell renal carcinoma with clinical interpretability. This streamlined classification approach effectively leverages the discriminative information captured by the preceding multimodal fusion process.

4. Results and discussion

This part evaluates the model's performance with selected key metrics such as AC, PR, SE and SP. Confusion matrix and the Precision-Recall curve are also included to demonstrate the model's effectiveness. Table 1 and 2 show the metrics, figure 2 shows the confusion matrix heatmap and the Precision-Recall curve.

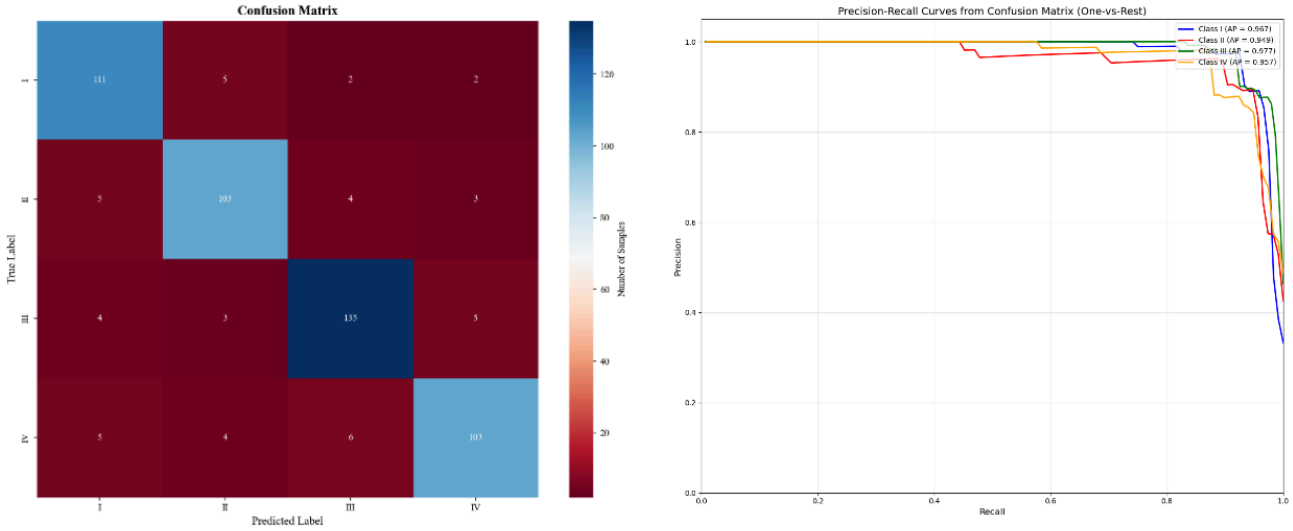
Table 1. Classification results using bimodal data

Stage (I-IV)	AC	PR	SE	SP
I	92.5	90.2	92.5	92.3
II	89.6	90.4	89.6	93.0
III	91.8	94.5	91.8	91.1
IV	87.3	91.2	87.3	93.7

Table 2. Classification results using unimodal (genetic) data

Stage (I-IV)	AC	PR	SE	SP
I	86.0	86.9	86.1	90.8
II	87.1	87.9	87.0	89.8
III	88.2	87.1	88.2	89.2
IV	89.0	87.4	88.3	90.1

DOI: 10.54254/2755-2721/2026.TJ30564



(a) Confusion Matrix (b) Precision-Recall Curve

Figure 2. Confusion matrix and Precision-Recall curve

The experimental results demonstrate the clear superiority of our multimodal model over the unimodal reference model, with consistent performance gains across all four stages of KIRC, as shown in Tables 1 and 2. This is particularly important for early detection (stages I and II), where high sensitivity (92.5% and 89.6%, respectively) is crucial for reducing diagnostic errors. This confirms that the integration of radiomics effectively compensates for subtle genomic signals in early lesions. However, the discernible misclassification between stage II and III samples, likely due to their similar imaging manifestations, highlights an area for future improvement in the fusion strategy.

5. Conclusion

This study proposed a novel deep learning framework that integrates genomic data and CT radiomics for the diagnosis of Kidney Renal Clear Cell Carcinoma (KIRC). By employing a dual-branch network coupled with a dynamic cross-modal attention mechanism, the model effectively captures and fuses complementary information from multi-omics data. This idea can be implemented in KIRC detection with further developments. Despite preliminary good-looking results, it is important to acknowledge the major limitations of this work, including (a) the partial data loss makes establishment of strict temporal alignment of the bimodal data difficult and (b) the size of gathered bimodal dataset limits the model's generalization capability. Further research will focus on increasing the dataset size along with attempts to enhance explainability of the decision, such as displaying the precise areas in CT images that contributes most to the prediction.

References

[1] National Cancer Institute. (2018). Diagnostic and prognostic biomarkers in renal clear cell carcinoma [Fact sheet]. NIH. https: //www.cancer.gov/types/kidney/research/biomarkers

[2] JAMA Network. (2025). What is renal cell carcinoma? JAMA Patient Page, 323(4), 1. https: //doi.org/10.1001/jama.2024.30123

[3] P. Das, "sigFeature: Novel significant feature selection method for classification of gene expression data using support vector machine and t statistic, " Front. Genet., vol. 11, 2020, Art. no. 247.

DOI: 10.54254/2755-2721/2026.TJ30564

[4] Q. Wu and D. Li, “CRIA: An interactive gene selection algorithm for cancers prediction based on copy number variations, ” Front. Plant Sci., vol. 13, 2022, Art. no. 839044.
[5] Rahman, T., Khandakar, A., Kadir, M. A., Islam, K. R., Islam, K. F., Mazhar, R., ... & Chowdhury, M. E. (2020). Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. Ieee Access, 8, 191586-191601.
[6] H. B. Okmen, H. Uysal, and A. Guvenis, "Automated prediction of TNM stage for clear cell renal cell carcinoma disease by analyzing CT images of primary tumors, " in 11th International Conference on Electrical and Electronics Engineering (ELECO 2019), Bursa, TURKEY, 2019, pp. 456 - 459, doi: 10.23919/eleco47770.2019.8990496.
[7] https: //portal.gdc.cancer.gov/projects/TCGA-KIRC
[8] https: //www.cancerimagingarchive.net/