

LKAN: LLM-BASED KNOWLEDGE-AWARE ATTENTION NETWORK FOR CLINICAL STAGING OF LIVER CANCER: SUPPLEMENTAL MATERIAL

Our experiments are all based on Chinese radiology reports. The English translation in the main text is only for illustration. The corresponding explanations (figures or tables) from English to Chinese are as follows:

Dataset Description The original dataset includes two parts: radiology reports and their corresponding staging (as shown in Table 1). The radiology report contains descriptions of lesions related to CSoLC, from which clinicians extract relevant information and compare it with the rules of CNLC one by one to stage each patient and develop subsequent treatment plans (hallmarks are detailed in the last column of Table 1). Our proposed method aims to learn from the experience of clinical doctors in first class hospitals and make staging judgments on radiology reports.

Table 1. Example sentences (Chinese version).

Sentences	Staging	Hallmark(s)
肝脏边缘欠光整，肝叶比例失调，肝裂增宽。肝S4及S5/6见两个类圆形低密度影，大小分别约31mm×28mm×28mm、33mm×33mm×32mm，增强动脉期不均匀强化，强化程度接近或稍低于邻近肝实质，门脉期强化程度减退，邻近肝静脉稍受压。	Ila	Size of tumors > 3 (cm) Number of tumors > 1
平扫肝见多发大小不等、边界不清略低密度灶，双期增强扫描后，动脉期病灶周边部见多个小结节状强化灶，密度高于肝接近同层主动脉，门脉期病灶内强化灶范围扩大并逐渐向中央充填，最大者位于S5，大小约50mm×48mm。余肝实质未见异常强化灶。肝内、外胆管无扩张，肝门结构清晰。	Ilb	Size of tumors > 3 (cm)
S6见类圆形稍低密度影，直径约18mm，增强扫描动脉期呈等密度，门脉期呈低密度。门脉左支见卵圆形充盈缺损影，增强扫描轻度强化。	IIIa	Vascular Invasion
门静脉主干及其分支未见充盈缺损。肝内外胆道无扩张。肝门区可见一个增大的淋巴结，大小约为26.6mm×21.4mm，增强扫描后较明显强化。	IIIb	Extrahepatic Metastasis

The data format provided by the clinicians is as follows: the reports and their corresponding staging are shown in the first two columns of the table. We list the hallmarks in the last column to explain how clinicians summarize information related to CSoLC from radiology reports based on CNLC.

From Fig.2, it can be seen that the radiology report is lengthy, which increases the burden on doctors. We preprocess the report content to remove information unrelated to liver cancer, and use the content of "imaging findings" and "imaging diagnosis" to construct domain-specific word vectors (as shown in Fig.2(a)). Cosine similarity is used to calculate the adjacency relationship between different word vectors, and we can see that words related to stage-IIIa are close to each other (as shown in Fig.2(b)).

Visualization Fig.4 shows the correlation between similar words. Fig.5 shows that some important medical terms are given higher weights, such as "S6/7", "肝细胞癌", and "癌栓".

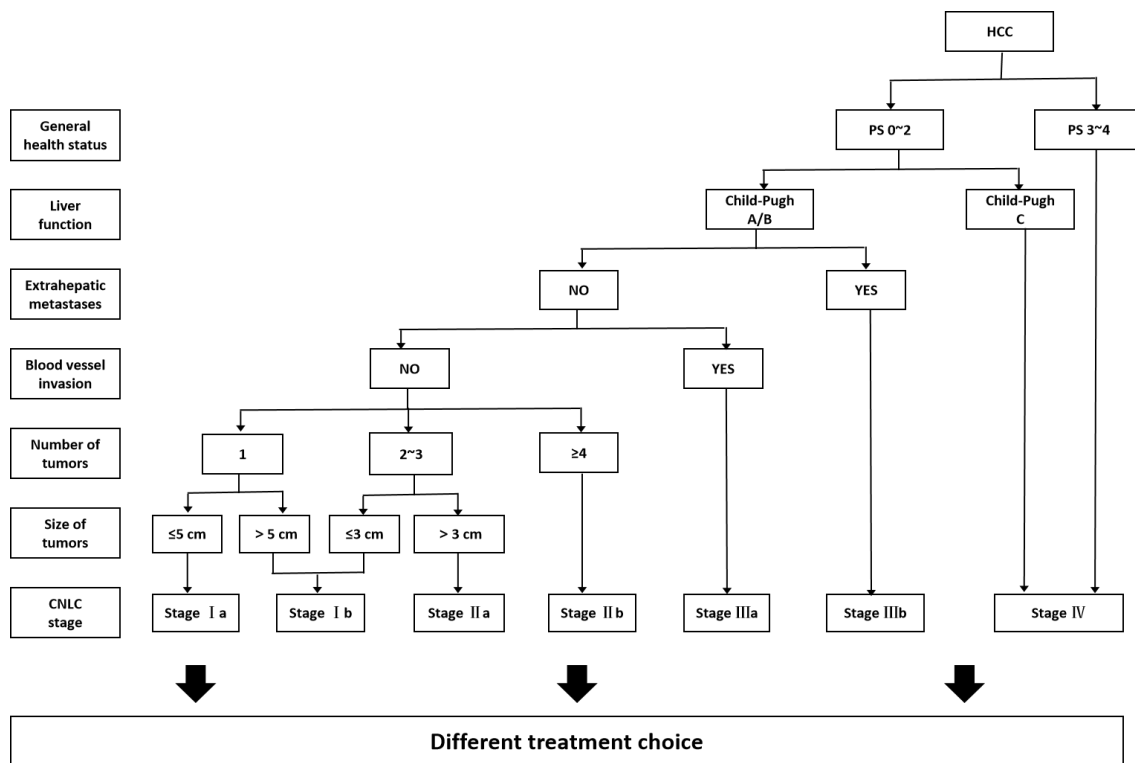


Fig. 1. China clinical staging and treatment pathway for PLC.

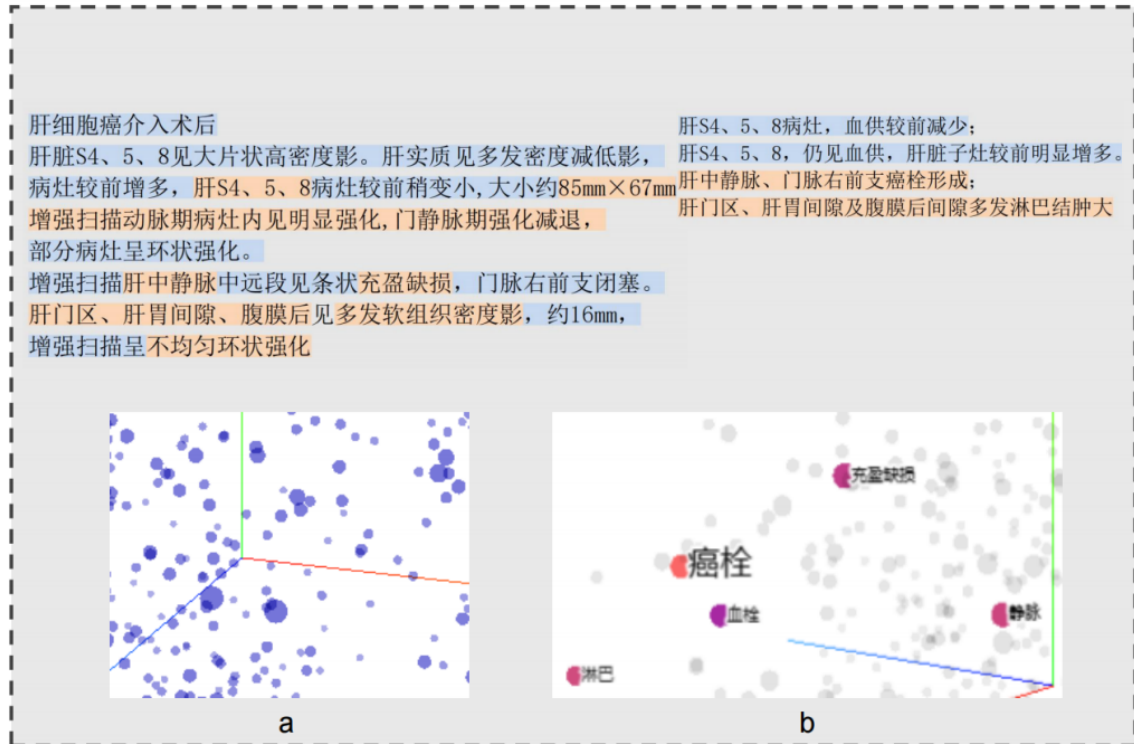
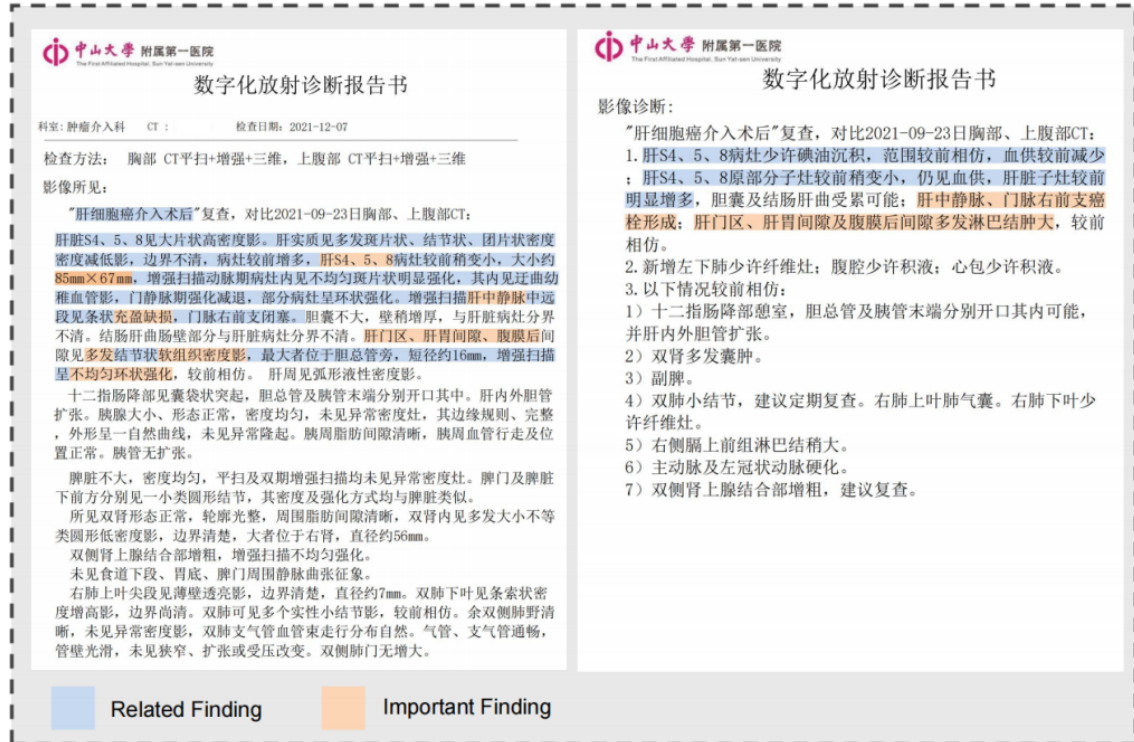


Fig. 2. An example of Chinese radiology reports (Chinese version). Blue blocks represent normal findings; Orange blocks represent relevant findings; Others indicate that they are not related to CSOLC. The distribution of pre-training word vectors is shown in (a). (b) showing spatial relationships of similar words.

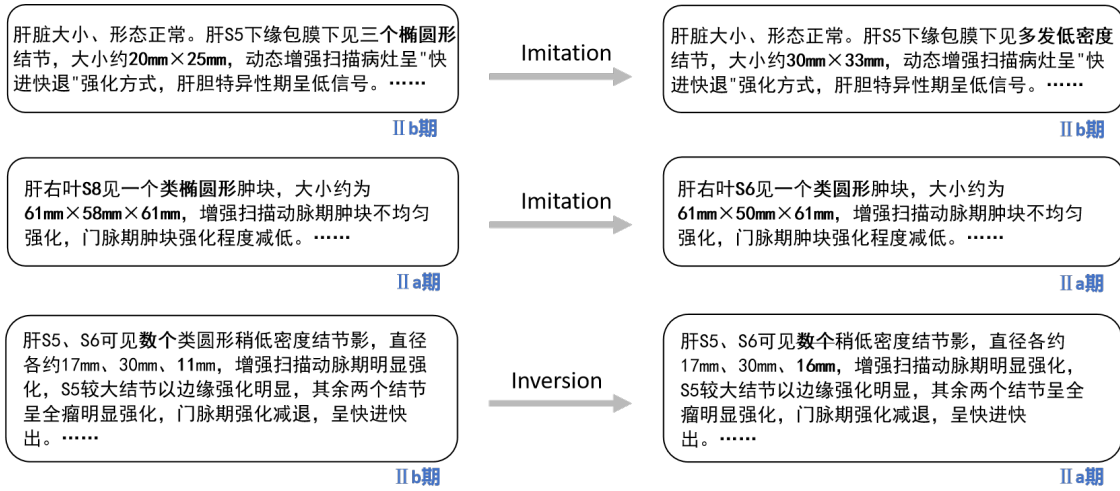


Fig. 3. Examples of “inversion” and “imitation” strategies (Chinese version).

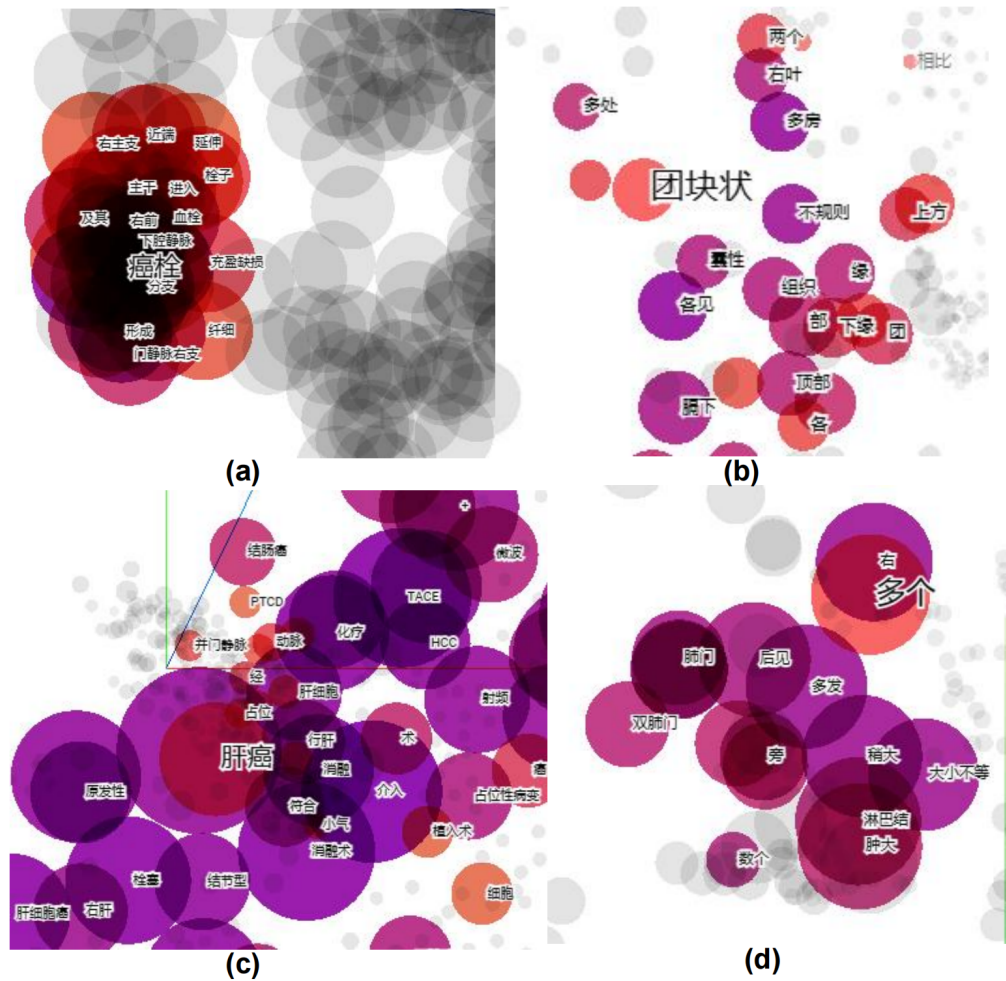


Fig. 4. Visualization of t-SNE distribution of word embedding (Chinese version).

肝 S6/7 肿块，考虑肝细胞癌，伴门静脉右支癌栓形成；肝门区、胃小弯多发稍大淋巴结，不排除转移。肝硬化，脾大。

Fig. 5. Visualization of attention weights (Chinese version).