

RESEARCH

Open Access



# Identifying stroke diagnosis-related features from medical imaging reports to improve clinical decision-making support

Xiaowei Xu<sup>1†</sup>, Lu Qin<sup>1†</sup>, Lingling Ding<sup>2</sup>, Chunjuan Wang<sup>2</sup>, Meng Wang<sup>2</sup>, Zixiao Li<sup>2\*</sup> and Jiao Li<sup>1\*</sup>

## Abstract

**Background** Medical imaging reports play an important role in communication of diagnostic information between radiologists and clinicians. Head magnetic resonance imaging (MRI) reports can provide evidence that is widely used in the diagnosis and treatment of ischaemic stroke. The high-signal regions of diffusion-weighted imaging (DWI) images in MRI reports are key evidence. Correctly identifying high-signal regions of DWI images is helpful for the treatment of ischaemic stroke patients. Since most of the multiple signals recorded in head MRI reports appear in the same part, it is challenging to identify high-signal regions of DWI images from MRI reports.

**Methods** We developed a deep learning model to automatically identify high-signal regions of DWI images from head MRI reports. We proposed a fine-grained entity typing model based on machine reading comprehension that transformed the traditional two-step fine-grained entity typing task into a question-answering task.

**Results** To prove the validity of the model proposed, we compared it with the fine-grained entity typing model, of which the F1 measure was 5.9% and 3.2% higher than the F1 measures of the models based on LSTM and BERT, respectively.

**Conclusion** In this study, we explore the automatic identification of high-signal regions of DWI images from the description part of a head MRI report. We transformed the identification of high-signal regions of DWI images to an FET task and proposed an MRC-FET model. Compared with the traditional two-step FET method, the model we proposed not only simplifies the task but also has better performance. The comparable result shows that the work in this study can contribute to improving the clinical decision support system.

**Keywords** Ischaemic stroke, High-Signal intensity regions, DWI, MRI reports, Fine-grained entity typing

<sup>†</sup>Xiaowei Xu and Lu Qin contributed equally to this work.

\*Correspondence:

Zixiao Li  
lizixiao2008@hotmail.com

Jiao Li  
lijiao@imicams.ac.cn

<sup>1</sup>Institute of Medical Information, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

<sup>2</sup>Beijing Tiantan Hospital, Capital Medical University, Beijing, China



**Introduction**

The application of artificial intelligence technologies to medical images is the foundation of intelligent clinical decision-making support. Currently, a medical imaging report consists of not only timely patient data but also evidence-based approaches and multiparametric computational and analytical methods, including patient-specific and population-based clinical, laboratory, genomic, demographic and quantitative data.

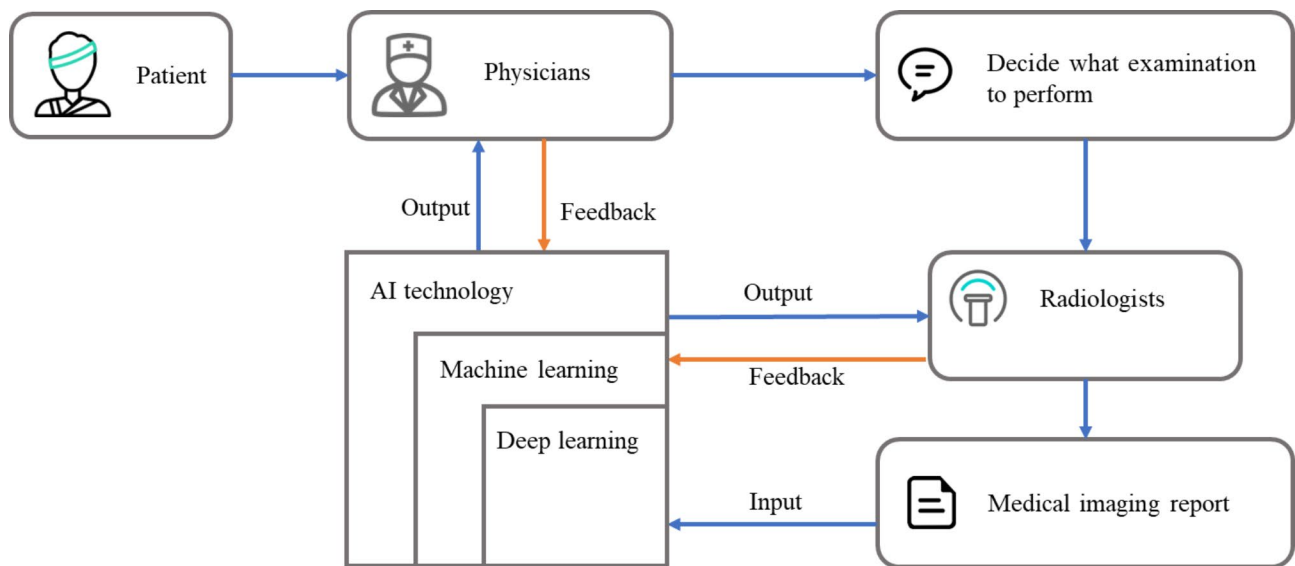
In clinical practice, when a patient arrives at the hospital, the physician will decide what examination to perform and request for imaging study based on the patient’s information. The order will be transmitted to the radiologist through the hospital information system. Radiologists usually manually identify the important information needed case by case and subsequently generate the image report. The manual film-reading method is labour intensive and inefficient. Studies report that, in some cases, an average radiologist must interpret one image every 3–4 s in an 8-hour workday to meet workload demands [1]. As radiology involves visual perception as well as decision-making under uncertain conditions [2], errors are inevitable—especially under such constrained conditions. At the same time, the lack of standards in the record, such as the phraseology difference and writing habits of radiologists, leads to the misunderstanding between the radiologist and the physician. With the development and application of deep learning technology in medical imaging, different artificial intelligence technologies have been applied to computer-aided medical film reading [3] [4], management options with probabilities grounded in evidence-based appropriate-use criteria can be presented to physicians so they can make better clinical decisions. Meanwhile, the feedback of the physicians and the

radiologists will help to improve the performance of the deep learning algorithms, as shown in Fig. 1.

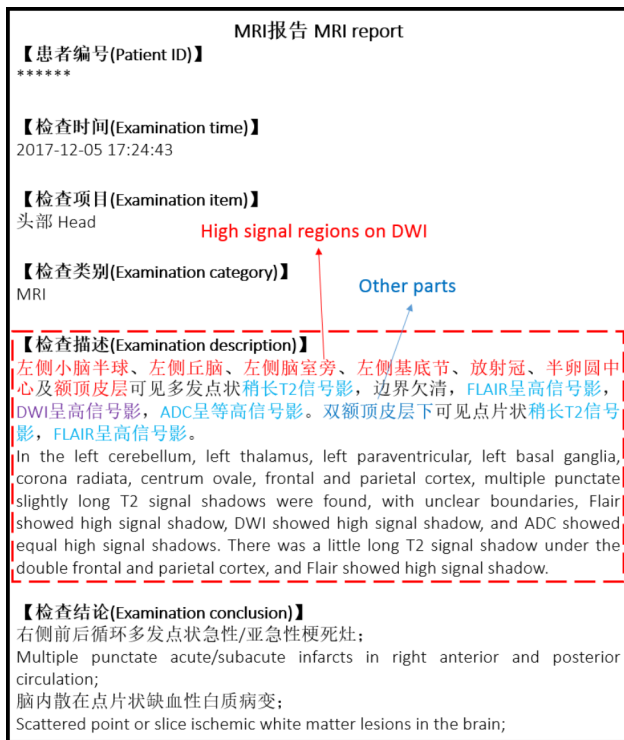
MRI of the head, as a key category of medical imaging reports, using a powerful magnetic field, radio waves and a computer to produce detailed pictures of the brain and other cranial structures, is widely used in clinics [5]. Common head MRI sequences T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), fluid attenuated inversion recovery (FLAIR), DWI and so on, among which DWI is used to detect the Brownian motion of water molecules in human tissues [6]. When the diffusion of water molecules in tissues is normal, the image has an equal signal, while when diffusion is limited, abnormally high-intensity signals are observed in DWI images. The signal intensity of DWI images depends on T2WI and the apparent diffusion coefficient (ADC), and the formula is as follows:

$$I_{DWI} \propto I_{T2WI} \times e^{-b \times ADC}$$

In the above formula, b represents the diffusion sensitivity factor. The larger b is, the greater the detection ability of the diffusion function. The greater the contrast between the lesion and normal tissue is, the higher the sensitivity. The causes of high signal intensity in DWI images are the prolongation of the T2WI signal (such as T2 shine-through) and the decrease in ADC. Many diseases, such as ischaemic stroke, brain tumours, brain abscesses, and lymphoma, show high signals in DWI. High signal intensity in DWI is the most sensitive sequence for detecting the infarct focus of acute ischaemic stroke, also known as the “stroke sequence” [7] [8] [9]. Therefore, the identification of high-signal regions



**Fig. 1** Information flow of a medical imaging report



**Fig. 2** Example head MRI report. The paragraph in the dotted box is the detailed examination description. The text in red means the high signal regions on DWI, and the text in dark blue means the other region parts. The text in purple means the statement of DWI, while the text in light blue means statement of the other modalities

in DWI is helpful for the treatment of ischaemic stroke patients.

Woo et al.[10] used convolutional neural networks (CNNs) to identify high-signal regions in DWI images to automatically segment acute ischaemic lesions and compared them with conventional algorithms. The results showed that the CNN algorithm had superior performance compared to the other conventional algorithms. Based on previous studies, Chang et al.[11] used the advanced 3D U-Net neural network to automatically identify high-signal regions from DWI images, and the intraclass correlation coefficient of the recognition results on the test set reached 0.974. Meanwhile, natural language processing (NLP) technologies have been applied to identify subjects from radiology reports. Kim et al.[12] used NLP and machine learning algorithms to identify acute ischaemic stroke from brain MRI reports. Carrodegua E et al.[13] used traditional machine learning and deep learning models to identify follow-up recommendations in radiology reports. Kang et al.[14] used NLP technology to identify incidental lung nodules in radiology reports to assess management recommendations. However, as far as we know, there is no research on the automatic extraction of the high-signal regions of DWI images from head MRI reports. The application

of natural language processing technology in the medical field, especially in the structured processing of MRI reports, shows great advantages[15] and provides an opportunity to obtain the high-signal regions of DWI images from MRI reports.

The head MRI reports used in this study are from the electronic medical record (EMR) system of Beijing Tiantan Hospital, Capital Medical University. An example of an MRI report is shown in Fig. 2. In traditional medical natural language processing tasks, named-entity recognition technology is usually used to identify diseases, body parts, etc.[18, 16]). However, as the report in Fig. 2 shows, an MRI report records not only the high-signal regions of DWI images but also the high-signal regions from FLAIR, long-signal regions from T2WI, and some other regions but also high/long-signal regions from other MRI sequences, which causes considerable linguistic interference when using NER technologies to identify the high-signal regions of DWI images. Owing to the above factors, we transform the task of identifying the high-signal regions of DWI images to a fine-grained entity typing task.

### Related work

According to the different granularities of entity types, there are mainly two kinds of tasks in academia: named-entity recognition (NER) [17] [18] and fine-grained entity typing (FET) [19]. The former task extracts the mention from the text and predicts its type in the context, which is usually coarse-grained, such as diseases and body parts. Therefore, the subtask of predicting the type of mentions can be regarded as coarse-grained entity typing. The latter task predicts the entity type according to the mention type given. For example, the high-signal region in DWI is a subtype of the body part. Figure 3 shows the input and output of a traditional FET system. We therefore define the task as a combination of NER and FET tasks.

From the output of the example, we can divide the task of high-signal region recognition in DWI into two stages: first, recognizing all the body parts in an MRI report, namely, an NER task, and second, dividing the identified parts into high-signal regions of DWI images and other regions, namely, an FET task.

The BiLSTM-CRF model is one of the most widely used technologies in the named-entity recognition task [19–21]. In comparison with the CRF, BiLSTM and BiLSTM-CRF models in English EMRs, the BiLSTM-CRF model performed best[22]. With the development of pretraining language models such as BERT[23], the performance of NLP tasks has been further improved. Zhang et al.[24] used the BERT-based BiLSTM-CRF model to identify breast cancer from progress notes and discharge summaries of 100 breast cancer patients, for which the F1 score reached 93.53%, better than that of the BiLSTM-CRF

输入(input):	
上下文(context)	右侧丘脑见小片长T1长T2信号影，边界欠清，DWI呈高信号。 (In the right thalamus, there were small piece of long T1 and long T2 signal shadows with unclear boundary and high signal shadow on DWI.)
提及(mention)	右侧丘脑(right thalamus)
提及位置(span)	[1,4]
输出(output):	
类型(type)	/body part/High signal region on DWI

Fig. 3 Example of fine-grained entity typing task input and output

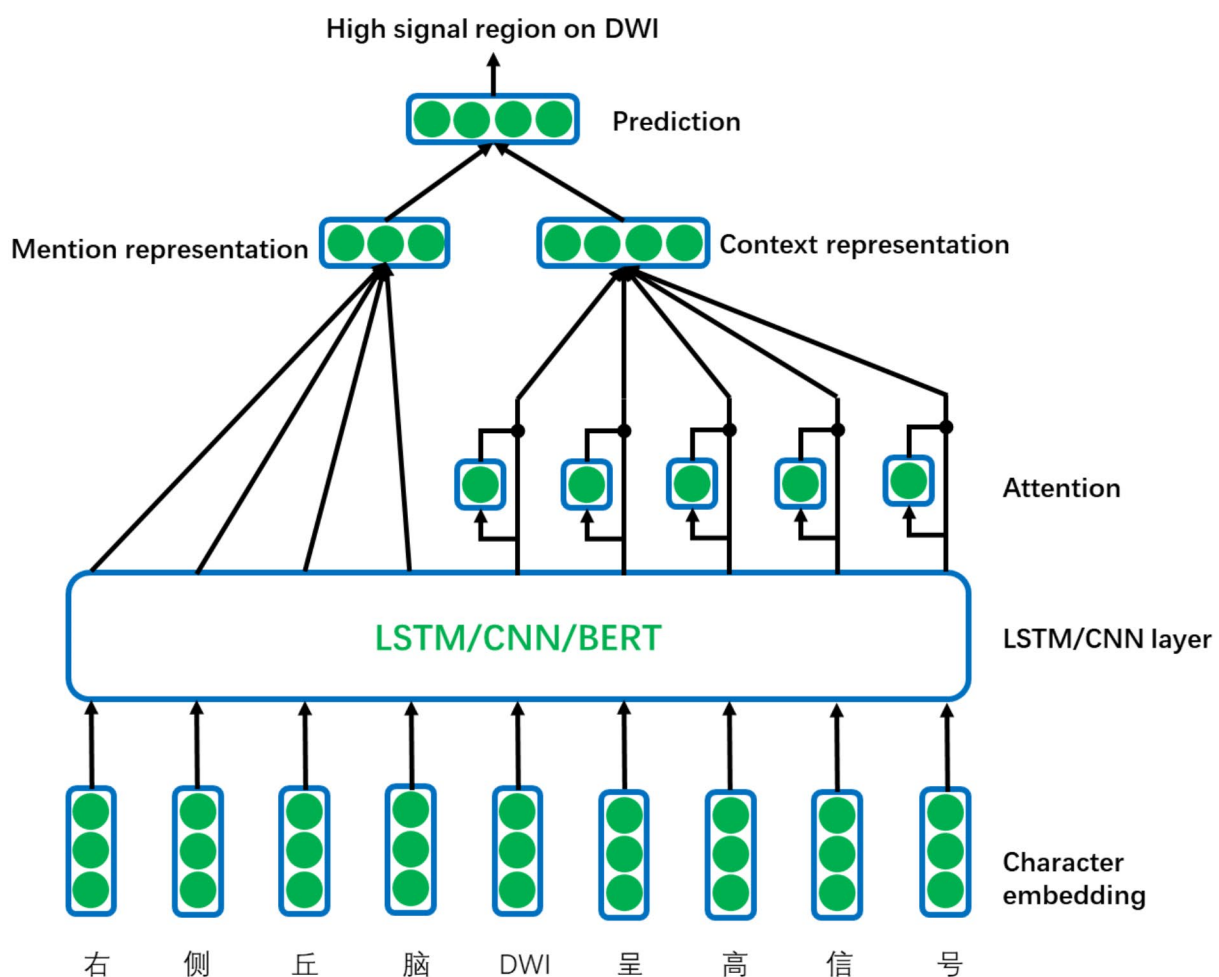


Fig. 4 General structure of the fine-grained entity typing model

model. On the other hand, FET is mainly used in the general field and rarely in the medical field. The essence of fine-grained name-entity typing is a multilabel classification problem. When the entity and its context are known, the entity can be classified into a specific category. The commonly used structure of the fine-grained named entity typing model is shown in Fig. 4. Generally, it is a

three-layer structure. The first layer is the text input layer, which transforms the text into the form of a character vector or word vector as the input of the second layer. The second layer is the representation layer of entity and context, and the CNN, LSTM, and BERT models and attention mechanism are commonly used for feature extraction to obtain vector representations of an entity

and entity context, respectively. The third layer is the prediction layer, which predicts the entity category[25][26]. Lee et al. [27] compared the performances of fine-grained entity typing models based on LSTM and BERT on a Chinese corpus, and the results showed that the performance of the latter is much better than that of the former. If we simply combine the NER model and the FET model, errors will accumulate, which will lead to poor performance. Li et al.[28] recently proposed a machine reading comprehension method to transform the traditional NER task (sequence labelling) into a question-answering task. Since the problem itself could provide additional information, this method could be introduced into an FET task. Therefore, we used the machine reading comprehension method to identify the high-signal regions of DWI images in this study.

## Method

### Dataset

The head MRI reports used in this study are from Beijing Tiantan Hospital, Capital Medical University. There were 2527 head MRI reports of 2527 patients, as shown in Fig. 2. MRI reports included patient ID, examination time, examination item, examination category, examination description and examination conclusion. The description of the DWI signal is mainly recorded in the examination description. Therefore, we chose the description part of the report to identify the high-signal regions of DWI images. After removing spaces and line breaks and converting upper-case English letters to lower-case English letters, we used the open-source software BRAT[29] to annotate the data. The annotation process is divided into two stages. The first stage is to annotate all body parts. The semiautomatic annotation method was adopted. First, 200 reports were manually annotated. All body parts appearing in the description part of MRI reports were annotated in the 200 reports (without identifying the high-signal regions of DWI images from other parts). The double-blind annotation method has been adopted for the annotation of body parts in the 200 MRI reports. Two junior radiologists are involved in the first round of annotation. For the inconsistent annotation results, an audit expert senior radiologist, with ten-years working experience, was invited to compare and determine the final answer, which acts as golden standard of the annotation results. From the feedback of the audit expert, the main inconsistency between the two annotators is the omission of some body parts.

Then, we trained a BiLSTM-CRF model to identify body parts of MRI reports with the 200 annotated reports, which was performed to automatically annotate the remaining 2327 reports. To ensure the accuracy of the corpus, the same audit expert was invited to check the annotation results identified by the BiLSTM-CRF

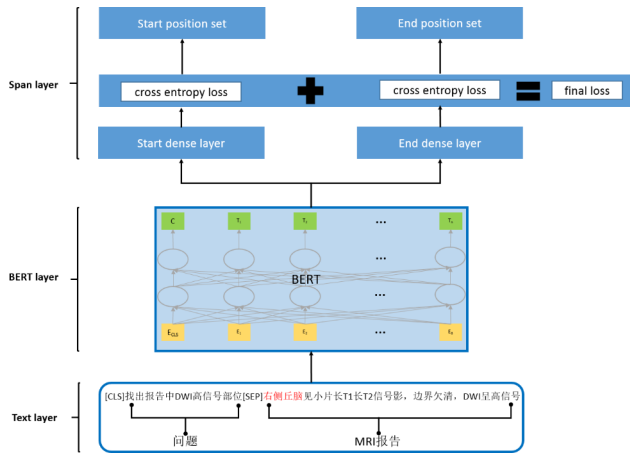
model. The audit expert modified 331 annotated body parts and added 43 body parts annotations. This semiautomatic annotation method greatly reduces the annotation workload in the first stage. Based on the first stage, we manually annotated parts in the description part of the MRI report as the high-signal regions of DWI images and other parts in the second stage. After annotation, clinical experts were also invited to review the corpus annotation and they approved the annotation results. Therefore, we obtained the final labelled corpus.

A total of 4369 high-signal regions of DWI images and 19260 other parts were annotated in the annotated corpus set. The average numbers of high-signal regions of DWI images and other parts in each examination description part of the report were 2 and 8, respectively. 90% of the examination descriptions in the reports did not contain any parts. The average length of each examination description was 170 Chinese characters, and the longest one consisted of 525 Chinese characters, while the shortest one consisted of only 2 Chinese characters. Then, we divided the corpus, all the annotated examination descriptions of the reports, into a training set and a testing set according to a ratio of 8:2. The training set contained 2022 examination descriptions of reports, including 3506 high-signal regions of DWI images and 15352 other parts. The testing set contained 505 examination descriptions of reports, including 863 high-signal regions of DWI images and 3908 other parts. The tagging scheme we used in this study is BIOES[30], in which B stands for ‘beginning’ (signifying the beginning of an entity), I stands for ‘inside’ (signifying that the character is inside an entity), O stands for ‘outside’ (signifying that the character is just a regular character outside of an entity), E stands for ‘end’ (signifying that the character is the end of an entity), and S stands for ‘singleton’ (signifying that the single character is an entity).

### Model

We proposed an FET model based on the machine reading comprehension (MRC-FET) method, which simplifies the original two-step fine-grained entity typing task into one step and compared the result with the work of Lee et al.[27] The traditional fine-grained entity typing task consists of two steps: NER and entity typing.

The structure of MRC-FET is shown in Fig. 5. The model consists of three layers. The first layer is the text layer. In the MRC-FET experiment, the question is “找出报告中DWI 高信号部位” (to find high-signal regions of DWI images in the MRI report), and the answer is high-signal regions of DWI images. First, we transform the MRI report into the format that the machine reading comprehension method in BERT needs, such as  $(CLS, q_1, q_2, \dots, q_m, SEP, x_1, x_2, \dots, x_n)$ , in which  $q_i$  is the question character and  $x_i$  is the MRI report character.



**Fig. 5** Structure of MRC-FET

Since the question shows what the experiment is to do, it can provide more information than the traditional fine-grained entity typing methods. The second layer is the BERT layer, and the output is matrix  $E \in R^{n \times d}$ , where n is the length of the MRI report description and d is the vector dimension of the last layer of the BERT model. The third layer is the span layer, which is mainly used to predict the start and end positions of the high-signal regions of DWI images in MRI reports according to the output of the BERT model. Start position prediction predicts the possibility that each character in the MRI report is the start position of the high-signal region of DWI images based on the matrix  $E$  output by BERT. The calculation formula is as follows:

$$P_{start} = (\text{softmax}_{eachrow}(E \bullet T_{start})) \in R^{n \times 2}$$

where  $T_{start}$  is the parameter that the model needs to learn. Each line of  $P_{start}$  represents the possibility that each character in the MRI report is the start position of the high-signal region of the DWI image. For example, if the first line of  $P_{start}$  is [0.3,0.7], the first character is not the start position of the high-signal region of the DWI image. If the second line of  $P_{start}$  is [0.8,0.2], the second character may be the start position of the high-signal region of the DWI image. Similarly, end position prediction predicts the possibility that each character in the MRI report is the end position of the high-signal region of the DWI image based on the matrix  $E$  output by BERT. The calculation formula is as follows:

$$P_{end} = (\text{softmax}_{eachrow}(E \bullet T_{end})) \in R^{n \times 2}$$

where  $T_{end}$  is the parameter that the model needs to learn. Each line of  $P_{end}$  represents the possibility that each character in the MRI report is the end position

of the high-signal region of the DWI image. Then, the argmax function acts on  $P_{start}$  and  $P_{end}$ , and two 0–1 sequences of length n  $I_{start}$  and  $I_{end}$  are obtained. If the i-th position is 1, then the i-th character may be the start or end position of the high-signal region of the DWI image.

$$I_{start} = (\text{argmax}_{eachrow}(P_{start}))$$

$$I_{end} = (\text{argmax}_{eachrow}(P_{end}))$$

After obtaining  $I_{start}$  and  $I_{end}$ , we match the start position and the end position according to the order of characters, and then, we obtain the position of the high-signal region of the DWI image in the MRI report. For example, if  $I_{start}$  is [0,1,0,0,0,0,1,0,0] and  $I_{end}$  is [0,0,0,1,0,0,0,0,0,0,1], then the locations of the high-signal regions of the DWI images are [2, 4] and [7, 10], respectively, because the start positions of the high-signal regions of the DWI images are 2 and 7 in  $I_{start}$  and the end positions are 4 and 10 in  $I_{end}$ , respectively. Thus, the positions of the high-signal regions of the DWI images in the description part of the MRI report are obtained by matching the start and end positions in sequence.

In the model training stage, we used two losses, the start position loss and the end position loss, which are defined as follows:

$$L_{start} = \text{Cross\_Entropy}(P_{start}, Y_{start})$$

$$L_{end} = \text{Cross\_Entropy}(P_{end}, Y_{end})$$

The cross entropy between the predicted and real results is determined, and then, the start position loss is added to the end position loss to obtain the final loss.

All methods were performed in accordance with the relevant guidelines and regulations.

## Experiment

### Evaluation criteria

In this study, we used the precision, recall and F1 measure to evaluate the models. We took the result predicted by the model on the test set as  $Pre=[p_1, p_2, \dots, p_i]$  and the result of manual annotation as  $GoI=[g_1, g_2, \dots, g_j]$ . The set element is a mention of the high-signal regions of DWI images and is represented by triple  $\langle d, pos_{start}, pos_{end} \rangle$ , where d is the serial number of the corpus, and  $pos_{start}$  and  $pos_{end}$  represent the start and end positions of high-signal regions in DWI images, respectively. When

$$p_i \cdot d = g_j \cdot d$$

$$p_i \cdot pos_{start} = g_j \cdot pos_{start}$$

**Table 1** Model parameter settings

Model	Character length	learning rate	batch size	epoch	optimizer
MRC-FET	400	0.00003	32	5	Adam
BERT-BiLSTM-CRF	400	0.00003	32	5	Adam
BiLSTM-FET	400	0.001	256	15	Adam
BERT-FET	400	0.00003	32	5	Adam

**Table 2** Results of identifying high-signal regions in DWI

Model	Precision (%)	Recall (%)	F1 measure (%)
BiLSTM-FET	86.39	92.7	89.44
BERT-FET	93.44	90.84	92.12
MRC-FET	97.27	93.62	95.41

$$p_i \cdot pos_{end} = g_j \cdot pos_{end}$$

the equation  $p_i = g_j$  holds, and the precision, recall and F1 measure are calculated as follows:

$$P = \frac{|intersection(Pre, Gol)|}{|Pre|}$$

$$R = \frac{|intersection(Pre, Gol)|}{|Gol|}$$

$$F1 = \frac{2PR}{P + R}$$

**Experimental parameters**

The experimental code of this study is based on the PyTorch deep learning framework. The parameter settings of MRC-FET, BERT-BiLSTM-CRF, FET based on

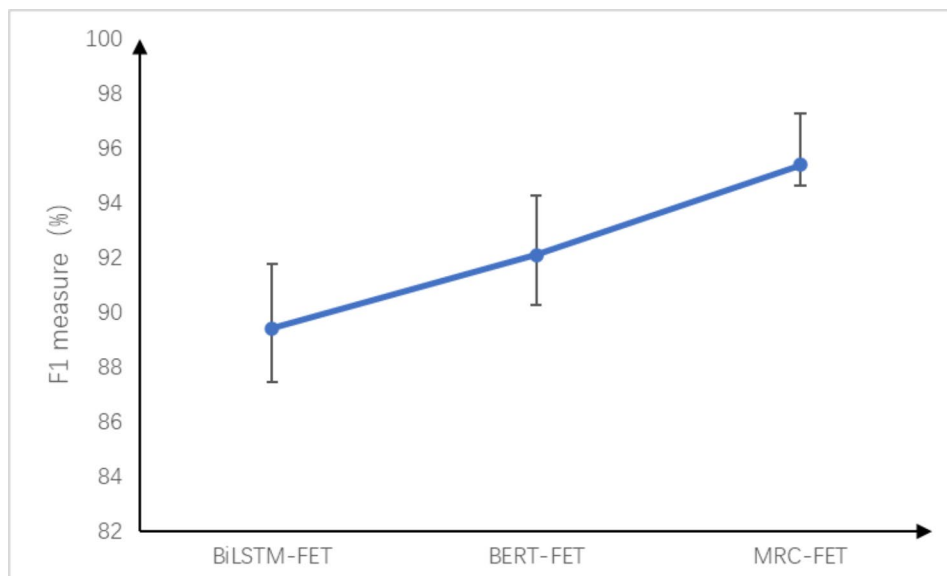
LSTM (LSTM-FET), and the FET model based on BERT (BERT-FET) are shown in Table 1, and the parameters except the character length in the last two models are consistent with those in Lee et al.[27]. Through our statistics, the character length of over 95% of MRI reports we obtained was less than 400. Then, we set the character length to 400.

**Results**

In the research of Jagannatha et al.[22] and Zhang et al.[24], the BERT and BiLSTM models were widely used in NER tasks and had good performance. Therefore, in the comparative experiments, the BERT-BiLSTM-CRF model identifying all parts in the MRI report description part was used in the first step. The BiLSTM-based FET model and BERT-based FET model were used in the entity typing task in the second step.

Table 2 shows that the MRC-FET model proposed in this study has the best performance on the testing set, with precision, recall and F1 measure values reaching 97.27%, 93.62% and 95.41%, respectively. The F1 measure is 5.9% and 3.2% higher than that of BiLSTM-FET and that of BERT-FET, respectively. We also report the error bars of the three models in Fig. 6. It can be seen that the error bar of the MRC-FET model we proposed is the shortest, which indicates that the model has better identification performance and stability.

The result shows that the MRC-FET model can not only simplify the task of identifying high-signal regions in DWI but also improve the performance.



**Fig. 6** Error bars of the three models

**Table 3** Error examples of the MRC-FET model's prediction results on the testing set

Error type	Example		
	Origin text	Gold standard	Prediction
broken writing	脑桥双底节放射冠双额	脑桥,双底节,放射冠,双额	脑桥
semantic similarity	xxxDWI等高信号xxx	None	Several regions
rare	右侧脑室体部旁脑白质	右侧脑室体部旁脑白质	None

**Discussion**

We analyse the reasons why the MRC-FET model performs well. First, the MRC-FET model is based on the BERT pretrained model, which can obtain bidirectional representations of deep-seated text and greatly improve the performance of downstream natural language processing tasks. Second, it draws lessons from the question-answering method in machine reading comprehension tasks and can clearly describe the task in question, transforming the traditional NER task (sequence labelling) into a question-answering task, “找出报告中DWI 高信号部位” (to find high-signal regions of DWI images in the MRI report), with the high-signal regions of DWI images as answers, which contributes to deal with the considerable linguistic interferences when using traditional NER technologies to identify the high-signal regions of DWI images, and also provides a new idea to solve fine-grained entity typing tasks. Finally, the MRC-FET model simplifies the traditional FET task and avoids the error accumulation of the original models.

In the comparative experiment, the two stages of BERT-FET are both based on the BERT pretrained model. Before the experiment, we speculated that the F1 scores of the BERT-FET and MRC-FET models on the testing set should have a small difference, which was 3.2%. We then conducted an experiment to analyse the reasons for the large difference. We used the MRC-FET model to identify all parts in the description part of the MRI report in the first step of fine-grained entity typing. In MRC-FET, we changed the question to “to find all parts in the MRI report” (origin is “to find high-signal regions of DWI images in the MRI report”). The precision, recall and F1 measure of the MRC-FET model were 95.93%, 96.70% and 96.31%, respectively, while those of the BERT-BiLSTM-CRF model were 94.91%, 93.37%, 94.13%, respectively. The F1 measure of the BERT-BiLSTM-CRF model is 2.18% worse than that of the MRC-FET model in recognizing all parts in the first step, which is the main reason why the performance of body part recognition is poor in the first step. On the other hand, the MRC-FET model proposed in this study can not only achieve good performance in the FET task but also perform well in the traditional NER task, which unifies NER and FET tasks. In addition, introducing machine reading comprehension

methods into sequence annotation can inspire the development of other natural language processing tasks.

Finally, we analysed the reasons for the error in MRC-FET prediction, which can be divided into three categories, as shown in Table 3. The first is due to the broken writing of MRI reports; in “脑桥双底节放射冠双额”, the standard writing is “脑桥、双底节、放射冠、双额”, the model only identifies the “脑桥”, the “双底节”, “放射冠” and “双额” were not identified, “双侧丘脑底节放射冠” is a similar situation, the standard writing is “双侧丘脑、底节、放射冠”, and the model also only identified the “双侧丘脑”. To address this issue, we can expand the dataset or strictly regulate the writing of MRI reports from the source to make improvements. The second reason is semantic similarity. For example, one of the testing sets describes “DWI等高信号” (“DWI等信号”) as its similarity with the high-signal region of the DWI image, and the model identifies some parts from the report that do not have to be identified. For this error, we expect a much stronger language model than BERT to appear in the future to help understand the context. The third type is caused by the few appearances of words in the corpus, such as “右侧脑室体部旁脑白质”, which only appears once in the training set and results in incomplete learning of the training model and poor scores in the testing set. For these errors, we can expand the scale of the training set and add more regular expressions in future research.

**Conclusion**

In this study, we were the first to explore the automatic identification of high-signal regions of DWI images from the description part of a head MRI report. Considering the construction of the report, we transformed the identification of high-signal regions of DWI images to an FET task and proposed an MRC-FET model. Compared with the traditional two-step FET method, the model we proposed not only simplifies the task but also has better performance, with an F1 measure that is 9% and 3.2% higher than the F1 measures of the LSTM-FET and BERT-FET models, respectively. In the future, we will expand the dataset and use regular expressions and a more advanced pretrained language model to improve the performance of identifying the accuracy of high-signal regions of DWI images from Chinese head MRI reports.

**List of abbreviations**

T1WI	T1-weighted imaging.
T2WI	T2-weighted imaging.
FLAIR	fluid attenuated inversion recovery.
DWI	diffusion-weighted imaging.
ADC	apparent diffusion coefficient.
EMR	electronic medical record.
NER	named-entity recognition.
FET	fine-grained entity typing.
MRC-FET	fine-grained entity typing model based on machine reading comprehension.
LSTM-FET	fine-grained entity typing model based on LSTM.



BERT-FET fine-grained entity typing model based on BERT.

#### Acknowledgements

Not applicable.

#### Authors' contributions

XWX and LQ processed the data, built the machine learning model, completed the experiments, and drafted the manuscript. LLD, CJW and MW provided medical knowledge, contributed to the experimental design, and reviewed the experimental results. ZXL and JL conducted the study, designed the data analysis protocol, and wrote and revised the manuscript. All authors read and approved the final manuscript.

#### Funding

This work was supported by the Beijing Natural Science Foundation (Grant No. Z200016), Chinese Academy of Medical Sciences (Grant No.2021-I2M-1-056, No. 2018-I2M-AI-016), and Ministry of Science and Technology of the People's Republic of China (National Key R&D Program of China, 2016YFC0901901). The funding agreement ensured the authors' independence in designing this study, interpreting the data, writing, and publishing the manuscript, and the funding provided us with sufficient experimental equipment.

#### Data Availability

Although we obtained permission from the institutional ethics committee to use the data, we did not obtain informed consent from patients to disclose medical history data. Therefore, the data are not available. However, if you are interested in our research, you can contact xiaowei xu (xu.xiaowei@imicams.ac.cn) for further information.

#### Declarations

##### Ethics approval and consent to participate

The Human Research Ethics Committees of Beijing Tiantan Hospital and Capital Medical University approved this study (NO. KY2020-155-02). All participants gave informed consent.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

Received: 21 April 2022 / Accepted: 30 September 2022

Published online: 20 October 2022

#### References

1. McDonald RJ, Schwartz KM, Eckel LJ, et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad Radiol.* 2015;22(9):1191–8.
2. Fitzgerald R. Error in radiology. *Clin Radiol.* 2001;56(12):938–46.
3. Altaf F, Islam SM, Akhtar N, et al. Going deep in medical image analysis: Concepts, methods, challenges, and future directions. *IEEE Access.* 2019;7:99540–72.
4. Taghanaki SA, Abhishek K, Cohen JP, et al. Deep semantic segmentation of natural and medical images: A review. *Artificial Intelligence Review*, 2020; 1–42.
5. Morris SA, Slesnick TC. Magnetic resonance imaging. *Visual Guide to Neonatal Cardiology.* 2018; 104–108.
6. Chung JW, Park SH, Kim N, et al. Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification and vascular territory of ischemic stroke lesions diagnosed by diffusion-weighted imaging. *J Am Heart Assoc.* 2014;3(4):e001119.
7. Brunser AM, Cavada G, Venturelli PM, et al. Diffusion-weighted imaging determinants for acute ischemic stroke diagnosis in the emergency room. *Neuroradiology.* 2018;60(7):687–92.
8. CSO N, Society CS. Chinese guidelines for diagnosis and treatment of acute ischemic stroke 2018. *Chin J Neurol.* 2018;51(9):666–82.
9. Bang OY, Li W. Applications of diffusion-weighted imaging in diagnosis, evaluation, and treatment of acute ischemic stroke. *Precision and Future Medicine.* 2019;3(2):69–76.
10. Woo I, Lee A, Jung SC, et al. Fully automatic segmentation of acute ischemic lesions on diffusion-weighted imaging using convolutional neural networks: comparison with conventional algorithms. *Korean J Radiol.* 2019;20(8):1275–84.
11. Chang K, Brown J, Beers A, et al. Abstract WMP17: Fully-Automated Ischemic Brain Infarct Volumetric Segmentation in Diffusion Weighted MR using Deep Learning. *Stroke.* 2019;50(Suppl\_1):AWMP17–7.
12. Kim C, Zhu V, Obeid J, Lenert L. Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke. *PLoS ONE.* 2019;14(2):e0212778.
13. Carrodeguas E, Lacson R, Swanson W, Khorasani R. Use of Machine Learning to Identify Follow-Up Recommendations in Radiology Reports. *J Am Coll Radiol.* 2019;16(3):336–43.
14. Kang SK, Garry K, Chung R, et al. Natural Language Processing for Identification of Incidental Pulmonary Nodules in Radiology Reports. *J Am Coll Radiol.* 2019;16(11):1587–94.
15. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys.* 2019;29(2):102–27.
16. Wu S, Roberts K, Datta S, et al. Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc.* 2020;27(3):457–70.
17. Wu Y, Jiang M, Xu J, Zhi D, Xu H. Clinical Named Entity Recognition Using Deep Learning Models. *AMIA Annu Symp Proc.* 2018;2017:1812–1819.
18. Giorgi JM, Bader GD. Towards reliable named entity recognition in the biomedical domain. *Bioinformatics.* 2020;36(1):280–6.
19. Huang L, May J, Pan X, et al. Liberal Entity Extraction: Rapid Construction of Fine-Grained Entity Typing Systems. *Big Data.* 2017;5(1):19–31.
20. Ji B, Liu R, Li S, et al. A hybrid approach for named entity recognition in Chinese electronic medical record. *BMC Med Inform Decis Mak.* 2019;19(Suppl 2):64.
21. Spasic I, Nenadic G. Clinical Text Data in Machine Learning: Systematic Review. *JMIR Med Inform.* 2020;8(3):e17984.
22. Jagannatha AN, Yu H. Structured prediction models for RNN based sequence labeling in clinical text. *Proc Conf Empir Methods Nat Lang Process.* 2016;2016:856–65.
23. Devlin J, Chang MW, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
24. Zhang X, Zhang Y, Zhang Q, et al. Extracting comprehensive clinical information for breast cancer using deep learning methods. *Int J Med Inform.* 2019;132:103985.
25. Choi E, Levy O, Choi Y, Zettlemoyer L. Ultra-Fine Entity Typing. *arXiv preprint arXiv:1807.04905*, 2018.
26. Shimaoka S, Stenetorp P, Inui K, Riedel S. Neural architectures for fine-grained entity type classification. In *Long Papers - Continued. Association for Computational Linguistics (ACL).* 2017. p. 1271–1280.
27. Lee C, Dai H, Song Y, Li X. A Chinese Corpus for Fine-grained Entity Typing. *arXiv preprint arXiv:2004.08825*, 2020.
28. Li X, Feng J, Meng Y, et al. A unified mrc framework for named entity recognition. *arXiv preprint arXiv:1910.11476*, 2019.
29. Stenetorp P, Pyysalo S, Topic G, et al. brat: a Web-based Tool for NLP-Assisted Text Annotation. The 3th Conference of the European Chapter of the Association for Computational Linguistics; Avignon, France. 102–107, 2012.
30. Reimers N, Gurevych I. Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks. *arXiv preprint arXiv: 1707.06799*, 2017.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.