Diagnose with Uncertainty Awareness: Diagnostic Uncertainty Encoding Framework for Radiology Report Generation

Sixing Yan[1,2*], Haiyan Yin[2], Ivor W. Tsang[2,3,4], and William K. Cheung[1]

[1] Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China {cssxyan,william}@comp.hkbu.edu.hk
[2] CFAR and IHPC, Agency for Science, Technology and Research, Singapore {yin_haiyan,ivor_tsang}@cfar.a-star.edu.sg
[3] CCDS, Nanyang Technological University, Singapore
[4] AAII, University of Technology Sydney, Australia

Abstract. Automated generation of radiology reports from X-ray images serves as a crucial task to streamline the diagnostic workflow for medical imaging and enhance the efficiency of radiologist decision-making. For clinical accuracy, most existing approaches focus on achieving accurate predictions of the existence of abnormalities, despite the inherent uncertainty impacting the reliability of the generated report, which is often clarified by radiologists simultaneously. In this paper, we present a unified report generation framework featuring a novel diagnostic uncertainty estimation model, named *Diagnostic Uncertainty Encoding* framework (DiagUE). Inspired by the clinician's uncertainty-aware radiology decision-making behavior, DiagUE first formulates belief-based diagnostic uncertainty metrics that effectively capture the variability of radiology abnormalities. Then, the estimated uncertainty-aware abnormality prediction is integrated with a report generation model under a novel visual-language encoding mechanism. Extensive experiments on two public benchmark datasets demonstrate that DiagUE could outperform SOTA baselines in ensuring the clinical accuracy of both abnormality description and diagnostic uncertainty of the report generation.

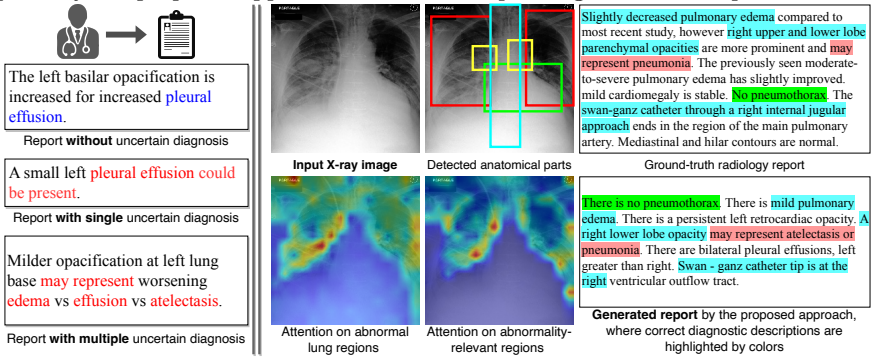
Keywords: Uncertainty Estimation · Radiology Report Generation · Medical Analysis.

1 Introduction

Nowadays, the radiology reporting plays a crucial role in expediting medical workflow in hospitals [3, 5, 13]. Radiology images are complex and difficult to interpret, with abnormalities manifesting with varying appearances. In reality, there are numerous instances where experienced radiologists generate diagnostic reports with inherent uncertainties [2, 17]. For instance, as shown in Fig. 1, given

* This work is done when Sixing Yan is a visiting student in CFAR, Agency for Science, Technology and Research, Singapore.

Fig. 1. Left: Examples of radiology report with uncertain diagnosis. **Right**: Radiology reports by the proposed approach with the correponding attention maps.



an X-ray image exhibiting developmentally abnormality observation, even a well-trained radiologist may only be able to identify that the observation falls within one of several possible abnormality categories, making it challenging to reach a definitive conclusion at the current stage. Acknowledging and managing such diagnostic uncertainty is essential for advancing the accuracy and reliability of radiology report generation systems.

Despite its importance, most existing approaches only consider predicting the appearance of abnormality as a certain decision, while neglecting the crucial diagnostic uncertainty information. Recently, there has been a growing interest in uncertainty-aware models in the medical imaging community. Some initial efforts involve inferring uncertainty scores from predicted probability for abnormality classification to align the features from classification and segmentation tasks [18, 23]. Another approach leverages variational inference techniques to enhance the posterior distribution [14]. However, developing uncertainty models would require the radiology knowledge, and notably, the existing approaches exhibit limited connection with the clinician's uncertainty estimation behavior. We also highlight that existing works could not offer human comprehensible uncertainty estimation results, where the incorporation of uncertainty awareness could add an additional dimension of diagnostic information for end-users.

Developing uncertainty estimation models also has immense potential to generate clinically accurate reports. To achieve that, one prevailing direction is to formulate the generation process as a two-stage task [15, 22] of predicting medical keywords from images and generating report by the image feature and predicted keywords. However, radiologists' decision-making is more informative and context-aware since subtle details may not be explicitly represented as keywords. To enhance the two-stage model, it becomes imperative to incorporate more informative tasks in the initial stage and offer high-quality features imbued with richer clinical contexts at the second stage. In this paper, we present an uncertainty-aware report generation approach that can serve the aforementioned two objectives simultaneously. Specifically, we formulate a novel

clinically-inspired uncertainty prediction task, which assesses diagnostic uncertainty based on two main sources of variabilities: ***abnormality variability*** and ***observer variability***. Abnormality variability indicates both variation between observations of individuals and variation between subjects [16], while observer variability corresponds to the intra- and inter-variation of radiologists in medical imaging [1]. Then, we compute the diagnostic uncertainty score by propagating ***diagnostic belief*** representing subjective logistics (SL) [8] of radiologists in drawing abnormality conclusions. Consequently, we extract context-rich uncertainty-aware visual and language embeddings to generate radiology reports.

Overall, the contributions of this paper are three-fold: **i)** we introduce a clinically-inspired formulation of the abnormality uncertainty estimation model featuring *abnormality*-level and *observer*-level uncertainties in the abnormality diagnosis process; **ii)** a unified uncertainty-aware encoder-decoder architecture is presented, which leverages the uncertainty prediction outcome and uncertainty-aware abnormality feature for report generation; and **iii)** we demonstrate extensive experiments on two publicly available chest X-ray datasets to evaluate DiagUE, with SOTA performance of clinical accuracy on abnormality descriptions and diagnostic uncertainty in the generated reports.

2 Related Works

Radiology Report Generation X-ray report generation models offer convenient decision-making support for radiologists to discern abnormalities from medical images and provide clinical conclusions by end-to-end training [21, 20, 27]. Initial approaches utilized an encoder-decoder architecture to achieve automatically radiolilogy image report generation [5, 4]. Recently, to improve the clinical accuracy of the report, knowledge graph-based approaches have been explored [28, 30]. While they enhance the intricate semantic connections, clinical correctness is constrained by assuming indifferent abnormality prediction-based diagnostic behavior. Our work is the first one to propose modeling diagnostic uncertainty to enhance clinical accuracy for such methods.

Uncertainty Modeling in Medical Imaging Despite the importance of uncertainty modeling in medical imaging [2, 17], limited works have been proposed in this area. Most deep models tend to approximate deterministic functions, overlooking uncertainty prediction in clinical diagnosis. In [9, 18, 23], evidence-based methods are introduced to address uncertainty in medical image classification. Recently, the uncertainty of the input X-ray image is also considered in the report generation [24]. Our work is related to [14], where a variational inference framework is proposed to align visual-language modalities in the latent space with posterior from latent topic variables. Where their uncertainty comes from incorporating a variational framework, we formulate a unique clinically-inspired uncertainty prediction task aimed at classifying clinicians' uncertainty scores and utilizing the uncertainty-aware features to improve the downstream report generation task. To the best of our knowledge, our work is the first one to in-

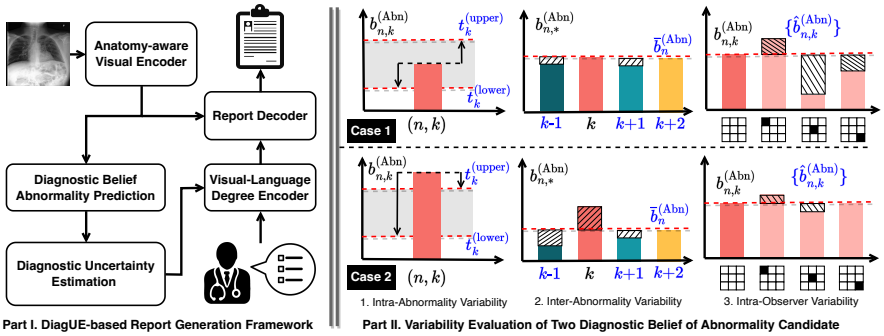


Fig. 2. Part I: An illustration of the DiagUE-based report generation framework. **Part II**: Demonstrating the numerical value of abnormality beliefs for three diagnostic variation, where the upper row corresponds to a relatively uncertain prediction (case 1) and the lower row corresponds to a relatively certain prediction (case 2).

troduce the clinically-inspired uncertainty-aware diagnostic prediction task for medical reporting.

3 Diagnostic Uncertainty Estimation in Radiology Report Generation

Radiologists commonly adhere to typical diagnostic procedures for generating clinical reports. Initially, they identify abnormal observations in the images to gather clinical evidence of abnormalities. Subsequently, radiologists form subjective logistics (i.e., ***diagnostic belief***) to draw a diagnosis conclusion of a particular abnormality, which is then detailed in written reports with comprehensive context information, e.g., reflecting the degree of diagnostic uncertainty. In this section, we introduce a clinically-inspired ***Diag****nostic* ***U****ncertainty* ***E****ncoding* framework (DiagUE) to enhance clinical accuracy of radiology reports through uncertainty-aware decision-making (shown in Fig. 2).

Overall, the uncertainty-aware report generation involves the following steps. First, we encode the anatomy-aware abnormality features F from the visual features I of a frontal chest X-ray image to estimate the diagnostic belief of *abnormality existence* b and *diagnostic uncertainty* u. Afterward, the estimated b and u together with a customized report format $\mathcal{X}$ are encoded to form a degree-based visual-language embeddings d. Accordingly, F and d are used to guide the neural decoder to output the radiology report R.

3.1 Estimating Diagnostic Belief of Anatomy-aware Abnormalities

Anomalies are usually detected in the anatomical regions that are considered abnormal in the prior examination. Given the questionable region detected, clinical evidence of possible abnormalities is then collected to form the diagnostic belief

to make the follow-up diagnosis. This motivates us to first obtain the vision embedding of anatomical parts and subsequently estimate the diagnostic belief of each abnormality candidate by the obtained anatomy embedding.

Given an input X-ray image, the visual extractor (e.g., pre-trained Vision Transformer [20]) encodes it as $I \in \mathbb{R}^{\mathcal{H}\mathcal{W}\times\mathcal{D}}$ of $\mathcal{H}\mathcal{W}$ feature map with $\mathcal{D}$ dimensional features. To detect potential abnormalities on each anatomical part, we also encode the regions of $\mathcal{N}$ anatomical parts detected by a pre-trained object detector [25] with each anatomical region represented as $V_n \in \mathbb{R}^{\mathcal{H}_n\mathcal{W}_n\times\mathcal{D}}$. We then learn the visual feature of $\mathcal{K}$ pre-defined abnormalities on each anatomical part, and estimate diagnostic belief of whether these abnormalities are diagnosed as clinical findings or not. To achieve that, we introduce a learnable memory matrix $E_k^{(\text{Abn})} \in \mathbb{R}^{\mathcal{E}\times\mathcal{D}}$ of each k-th abnormality with a number of $\mathcal{E}$ memory slots. $F_{n,k}$ the visual feature of anatomy-aware abnormality is then computed by the cross-attention network $\text{CrossAttn}(\cdot,\cdot)$, and $b_{n,k}$ the diagnostic belief of the existence of (n,k)-th abnormality is obtained:

$$F_{n,k} = \text{CrossAttn}(V_n, E_k^{(\text{Abn})}); \quad b_{n,k} = \underset{\mathcal{H}_n\mathcal{W}_n\to 1}{\text{AvgPooling}}(\underset{\mathcal{D}\to 1}{\text{FFN}}(F_{n,k})). \tag{1}$$

where $\text{FFN}(\cdot)$ is a two-layer fully-connected neural network [29]. Then, $b_{n,k}$ is used as the prediction probability of abnormality existence, optimized by multi-label classification of whether (n,k)-th abnormality is observed or not.

3.2 Estimating Belief-based Diagnostic Uncertainty

In radiology tests, radiologists could only report the appearance of abnormality given sufficient diagnostic belief; otherwise, they would report possible abnormalities as suspected findings and suggest follow-up examination or treatment. To mimic this, we present a novel belief-based diagnostic uncertainty estimation mechanism, where we formulate the clinically-inspired uncertainty based on three types of variability in the abnormality detection on the radiology imaging: **i)** *Intra-abnormality variability* for single abnormality with uncertain appearance; **ii)** *Inter-abnormality variability* for multiple confusing disjoint abnormalities; and **iii)** *Intra-observer variability* for one observer with instance noises.

i) Intra-Abnormality Variability (Intra-Abn) $v^{(\text{Intra-Abn})}$ in radiology imaging is usually observed when an abnormality has varied appearances and the prediction toward the detected observation is hard to confirm. This variability could be captured by the amount of diagnostic belief in the existence of an abnormality, when it is not enough to make a positive diagnosis ("*Definitely present*") while too much to make a negative diagnosis ("*Definitely absent*").

To collect the related variability $v^{(\text{Intra-Abn})}$, we first learn a numerical range of diagnostic belief that indicates the expected uncertain diagnosis, denoted $[t_k^{(\text{lower})}, t_k^{(\text{upper})}]$. $t_k^{(\text{lower})}, t_k^{(\text{upper})} \in (0,1)$ are the lower and upper bounds of the numerical range learned by two scalar parameters. Then, $v_{n,k}^{(\text{Intra-Abn})}$ is measured

by the relative value of $b_{n,k}$ with reference to $[t_k^{(\text{lower})}, t_k^{(\text{upper})}]$ as:

$$v_{n,k}^{(\text{Intra-Abn})} = (b_{n,k} - t_k^{(\text{upper})})(b_{n,k} - t_k^{(\text{lower})}). \tag{2}$$

ii) Inter-Abnormality Variability (Inter-Abn) $v^{(\text{Inter-Abn})}$ is considered when one abnormal observation could be related to conflict abnormalities with similar appearances. If several abnormalities on one anatomical part are assigned with high but close diagnostic beliefs, then the predictions of these abnormalities will become conflict where none of the abnormalities can be distinguished from others as a confirmed diagnosis. To measure the variability among multiple abnormalities $v^{(\text{Inter-Abn})}$, we consider the deviation of diagnostic beliefs as,

$$v_{n,k}^{(\text{Inter-Abn})} = 1 - \frac{|\bar{b}_n - b_{n,k}|}{\sum_{k=1}^{\mathcal{K}} |\bar{b}_n - b_{n,k}|}\bar{b}_n; \quad \bar{b}_n = \frac{1}{\mathcal{K}}\sum_{k=1}^{\mathcal{K}} b_{n,k} \tag{3}$$

iii) Intra-Observer Variability (Intra-Obs) $v^{(\text{Intra-Obs})}$ is found in the radiology imaging where varying predictions can be made by one observer (radiologist) for the same observation. This variability is indicated where the predictions vary due to the fine-grained changes of one abnormality observation. Motivated by the Monte Carlo dropout used in the uncertainty validation [26, 12], we capture this variability by estimating additional diagnostic beliefs using the input image feature with random dropout operations. Then $v^{(\text{Intra-Obs})}$ is measured by the deviation of the diagnostic belief from the image with/without dropouts as,

$$v_{n,k}^{(\text{Intra-Obs})} = \frac{1}{|B_{n,k}|}\sum_{\hat{b}_{n,k} \in B_{n,k}} \frac{|\hat{b}_{n,k} - b_{n,k}|}{b_{n,k}}, \tag{4}$$

where $|\hat{b}_{n,k} - b_{n,k}|$ calculates the deviation of diagnostic beliefs between the input image feature with and without dropouts, and $B_{n,k} = \{\hat{b}_{n,k}\}$ is the set of diagnostic beliefs estimated by $F_{n,k}$ processed with different dropout operations.

Variability-based Diagnostic Uncertainty $u_{n,k}$ is then obtained by weighted summing the diagnostic variability as $u_{n,k} = w_{n,k}^{(1)}(v_{n,k}^{(\text{Intra-Abn})}) + w_{n,k}^{(2)}(v_{n,k}^{(\text{Inter-Abn})}) + w_{n,k}^{(3)}(v_{n,k}^{(\text{Intra-Obs})})$ where the operation $w_{n,k}(\cdot) \in (0, 1)$ is composed of a SoftPlus activation function to ensure the variability value to be larger than 0 [23], and an importance weighting to be learned by a scalar parameter. $u_{n,k}$ is used as the probability of diagnostic uncertainty towards (n, k)-th abnormality, optimized by the multi-label classification of whether the diagnosis of the (n, k)-th abnormality is reported as uncertain findings or not.

3.3 Report Generation with Belief-Driven Uncertainty Estimation

Given the input X-ray image encoded as the visual embedding of anatomy-aware abnormalities $F_{n,k}$ and the diagnostic results $(b_{n,k}, u_{n,k})$ obtained, a radiology

report is generated that is also required to follow certain report formats (denoted $\mathcal{X} = \{x_{k,m}\}_{m=1,k=1}^{\mathcal{M},\mathcal{K}}$ with $\mathcal{M}$ report formats considered). With a pre-defined $\mathcal{K}$ abnormalities considered, in total $\mathcal{K}$ sub-reports are generated by the decoder, where each sub-report R_k is specified to the k-th abnormality. The complete report $\hat{R}$ will be concatenated from $\{R_1, ..., R_{\mathcal{K}}\}$ in the post-processing step.

Visual degree embedding To guide the neural decoder by the prediction results of abnormality existence and the associated diagnostic uncertainty, the numerical probabilities are usually transformed to a high-dimension vector by the linear projection, which could be easily over-fitting. To avoid that, we propose a ***degree mechanism*** to encode the prediction results by a set of disentangled feature vectors where each vector is related to certain input values. In particular, we first construct a learnable degree matrix $D \in \mathbb{R}^{\mathcal{C}\times\mathcal{D}}$ that represents $\mathcal{C}$ different degrees in $\mathcal{D}$-dimension. Given the predicted probability p in decimal value, we first scale it to a pre-defined degree range $\{1, 2, ...\mathcal{C}\}$. Then, the degree embedding of p is indexed from the degree matrix D as D_c (the c-th row of D).

To obtain the degree embedding of the existence probability toward k-th abnormality $d_k^{(\text{Abn})} \in \mathbb{R}^{1\times\mathcal{D}}$, the degree vectors are indexed by $b_{n,k}$ and aggregated as $d_k^{(\text{Abn})} = 1/\mathcal{N}\sum_{n=1}^{\mathcal{N}} D^{(\text{Abn})}[b_{n,k}]$

where $D^{(\text{Abn})}$ is the degree matrix of abnormality prediction. Similarly, we encode $u_{n,k}$ to the degree embedding of the diagnostic uncertainty $d_k^{(\text{Unc})} \in \mathbb{R}^{1\times\mathcal{D}}$ by the degree matrix of uncertainty estimation $D^{(\text{Unc})}$, given as $d_k^{(\text{Unc})} = 1/\mathcal{N}\sum_{n=1}^{\mathcal{N}} D^{(\text{Unc})}[u_{n,k}]$.

Language degree embedding The radiology report is a mission-oriented document following varied report formats, which are related to the language quality of the report, such as *level of details for different abnormalities*. To allow user customization, we also encode user-input report formats as language degree embedding to guide the report generation. We start by representing each report format using an integer value $x_{k,m}^{(\text{Lang})} \in \{1, 2, ...\mathcal{C}\}$, and construct the degree matrices of report formats $\{D_m^{(\text{Lang})}\}_{m=1}^{\mathcal{M}}$. Accordingly, the language degree embedding $d_k^{(\text{Lang})}$ is obtained by $d_k^{(\text{Lang})} = \text{FFN}_{\mathcal{M}\mathcal{D}\to\mathcal{D}}\big(\bigoplus_{m=1}^{\mathcal{M}} D_m^{(\text{Lang})}[x_{m,k}^{(\text{Lang})}]\big)$.

The visual-language degree embedding d_k is then obtained by fusing all degree embeddings as $d_k = \text{FFN}_{3\mathcal{D}\to\mathcal{D}}(d_k^{(\text{Abn})} \oplus d_k^{(\text{Unc})} \oplus d_k^{(\text{Lang})})$.

Uncertainty-aware Report Decoder In addition to the prediction results, we also extract the visual features of uncertain regions to improve the report generation. In particular, we learn the embedding of visual uncertainty of the k-th abnormality, denoted as U_k, by the abnormality embedding $\{F_{n,k}\}$ as:

$$\hat{F}_{n,k} = \text{CrossAttn}(F_{n,k}, E_k^{(\text{Unc})}); \quad U_k = \underset{\mathcal{N}\times\mathcal{H}\mathcal{W}\times\mathcal{D}\to 1\times\mathcal{D}}{\text{AvgPooling}} \bigoplus_{n=1}^{\mathcal{N}} \hat{F}_{n,k} \tag{5}$$

where $E_k^{(\text{Unc})} \in \mathbb{R}^{\mathcal{E}\times\mathcal{D}}$ is a learnable memory matrix. To guide the report generation, we concatenate d_k and U_k with the last hidden state of the token sequence

Table 1. Performance comparison on report generation by MIMIC CXR data (222,705/1,807/3,269 for train/valid/test).

Model	CE		RadRQI-F1		Uncertainty Acc.	
	(14)	(19)	TopK	Hits	Hard	Soft
Transformer [19]	0.360	0.550	0.199	24.0	0.287	0.319
$\mathcal{M}^2$Transformer [6]	0.385	0.584	0.231	**35.7**	0.303	0.342
R2Gen [5]	0.333	0.589	0.199	26.0	0.277	0.317
R2Gen-CMN [4]	0.415	0.607	0.282	30.0	0.321	0.366
WCL [27]	0.572	0.555	0.191	22.0	0.262	0.313
XProNet [21]	0.534	0.564	0.141	20.0	0.268	0.311
GIT [20]	<u>0.660</u>	<u>0.642</u>	<u>0.307</u>	17.0	<u>0.367</u>	<u>0.422</u>
DiagUE (proposed)	**0.664**	**0.688**	**0.319**	<u>31.5</u>	**0.441**	**0.473**

(sentence) of (k-1)-th abnormality (denoted as $h_{k-1,T} \in \mathbb{R}^{\mathcal{D}}$ at the last time step T) to feed to the language decoder in the k-th sentence generation, given as $y_{k,t}, h_{k,t} = \text{Decoder}(U_k \oplus d_k \oplus h_{k-1,T}, y_{k,1}, ..., y_{k,t-1})$. The generated report $R_k = \{y_{k,1}, y_{k,2}, ..., y_{k,T}\}$ is optimized by the language modeling loss [6].

4 Experiment

4.1 Experimental Settings

We use the publicly available report dataset MIMIC CXR [11] and IU Xray [7] for evaluation. We collect $\mathcal{K} = 19$ abnormalities in $\mathcal{N} = 10$ anatomical regions with two types of labels for training and testing by RadGraph [10] and Chest ImaGemone [25]: i) whether the abnormalities exist; and ii) whether the diagnosis is uncertain. *Clinical quality* of the generated reports is evaluated by clinical efficacy (`CE`) [5], radiology report quality index (`RadRQI-F1`) [28] and diagnostic uncertainty accuracy. Detailed task settings including hyperparameter configurations for DiagUE can be found in appendix.

4.2 Evaluation Results

We conduct extensive benchmark experiments to evaluate the reports generated by DiagUE (shown in Table 1)[5]. Among the baselines, GIT demonstrates strong overall performance, surpassing other baselines on most of `CE` metrics for abnormality prediction and `RadRQI-F1` metrics for attribute keywords prediction. Notably, GIT also showcases high standards in diagnostic uncertainty estimation. Remarkably, DiagUE achieves a performance improvement of 7.5% over GIT in `CE(19)`. Also, an average performance improvement of 15.8% is achieved in the uncertainty estimation metrics. This suggests that incorporation of DiagUE can effectively enhance clinical accuracy in terms of both abnormality prediction and diagnostic uncertainty estimation in radiology report generation.

[5] Additionally, we present evaluation results on NLG quality, abnormality detection scores on another IU Xray dataset, and a thorough ablation study in the appendix.

5 Conclusion

We present DiagUE, a belief-based diagnostic uncertainty encoding framework featuring a clinically-inspired diagnostic uncertainty estimator that effectively models the variability of radiology abnormalities. To the best of our knowledge, this is the first attempt to simultaneously consider diagnostic predictions on both the abnormality existence and diagnostic uncertainties in the radiology imaging domain. Through comprehensive empirical evaluations, we demonstrate that DiagUE surpasses state-of-the-art methods in terms of both clinical accuracy for abnormalities and uncertainty prediction on two benchmark datasets.

Acknowledgement

This research is partially supported by General Research Fund RGC/HKBU12202621 from the Research Grant Council and the Research Matching Grant Scheme RMGS2021_8_06 from the Hong Kong Government.

References

1. Al-Khawari, H., Athyal, R.P., Al-Saeed, O., Sada, P.N., Al-Muthairi, S., Al-Awadhi, A.: Inter-and intraobserver variation between radiologists in the detection of abnormal parenchymal lung changes on high-resolution computed tomography. Annals of Saudi medicine **30**(2), 129–133 (2010)
2. Bruno, M.A., Petscavage-Thomas, J., Abujudeh, H.H.: Communicating uncertainty in the radiology report. American Journal of Roentgenology **209**(5), 1006–1008 (2017)
3. Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G., Murphy, K.: Deep learning for chest x-ray analysis: A survey. Medical Image Analysis **72**, 102125 (2021)
4. Chen, Z., Shen, Y., Song, Y., Wan, X.: Cross-modal memory networks for radiology report generation. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. pp. 5904–5914 (2021)
5. Chen, Z., Song, Y., Chang, T.H., Wan, X.: Generating radiology reports via memory-driven transformer. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 1439–1449 (2020)
6. Cornia, M., Stefanini, M., Baraldi, L., Cucchiara, R.: Meshed-memory transformer for image captioning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10578–10587 (2020)
7. Demner-Fushman, D., Kohli, M.D., Rosenman, M.B., Shooshan, S.E., Rodriguez, L., Antani, S., Thoma, G.R., McDonald, C.J.: Preparing a collection of radiology examinations for distribution and retrieval. Journal of the American Medical Informatics Association **23**(2), 304–310 (2016)
8. Han, Z., Zhang, C., Fu, H., Zhou, J.T.: Trusted multi-view classification. In: International Conference on Learning Representations (2020)
9. Han, Z., Zhang, C., Fu, H., Zhou, J.T.: Trusted multi-view classification with dynamic evidential fusion. IEEE transactions on pattern analysis and machine intelligence **45**(2), 2551–2566 (2022)

10. Jain, S., Agrawal, A., Saporta, A., Truong, S.Q., Duong, D.N., Bui, T., Chambon, P., Zhang, Y., Lungren, M.P., Ng, A.Y., et al.: Radgraph: Extracting clinical entities and relations from radiology reports. arXiv preprint arXiv:2106.14463 (2021)
11. Johnson, A.E., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Mark, R.G., Horng, S.: Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. Sci. Data **6**(1), 317 (2019)
12. Luo, J., Sedghi, A., Popuri, K., Cobzas, D., Zhang, M., Preiswerk, F., Toews, M., Golby, A., Sugiyama, M., Wells, W.M., et al.: On the applicability of registration uncertainty. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd. pp. 410–419. Springer (2019)
13. Miura, Y., Zhang, Y., Tsai, E., Langlotz, C., Jurafsky, D.: Improving factual completeness and consistency of image-to-text radiology report generation. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 5288–5304 (2021)
14. Najdenkoska, I., Zhen, X., Worring, M., Shao, L.: Uncertainty-aware report generation for chest x-rays by variational topic inference. Medical Image Analysis **82**, 102603 (2022)
15. Nooralahzadeh, F., Perez Gonzalez, N., Frauenfelder, T., Fujimoto, K., Krauthammer, M.: Progressive transformer-based generation of radiology reports. In: Findings of the Association for Computational Linguistics: EMNLP 2021. pp. 2824–2832 (2021). https://doi.org/10.18653/v1/2021.findings-emnlp.241, https://aclanthology.org/2021.findings-emnlp.241
16. Reid, D., Samangooei, S., Chen, C., Nixon, M., Ross, A.: Chapter 13 - soft biometrics for surveillance: An overview. In: Rao, C., Govindaraju, V. (eds.) Handbook of Statistics, Handbook of Statistics, vol. 31, pp. 327–352 (2013). https://doi.org/https://doi.org/10.1016/B978-0-444-53859-8.00013-8, https://www.sciencedirect.com/science/article/pii/B9780444538598000138
17. Reiner, B.I.: Quantifying analysis of uncertainty in medical reporting: creation of user and context-specific uncertainty profiles. Journal of Digital Imaging **31**(4), 379–382 (2018)
18. Ren, K., Zou, K., Liu, X., Chen, Y., Yuan, X., Shen, X., Wang, M., Fu, H.: Uncertainty-informed mutual learning for joint medical image classification and segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 35–45 (2023)
19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems. pp. 5998–6008 (2017)
20. Wang, J., Yang, Z., Hu, X., Li, L., Lin, K., Gan, Z., Liu, Z., Liu, C., Wang, L.: GIT: A generative image-to-text transformer for vision and language. Transactions on Machine Learning Research (2022)
21. Wang, J., Bhalerao, A., He, Y.: Cross-modal prototype driven network for radiology report generation. In: European Conference on Computer Vision. pp. 563–579. Springer (2022)
22. Wang, L., Ning, M., Lu, D., Wei, D., Zheng, Y., Chen, J.: An inclusive task-aware framework for radiology report generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 568–577. Springer (2022)
23. Wang, M., Lin, T., Wang, L., Lin, A., Zou, K., Xu, X., Zhou, Y., Peng, Y., Meng, Q., Qian, Y., et al.: Uncertainty-inspired open set learning for retinal anomaly identification. Nature Communications **14**(1), 6757 (2023)

24. Wang, Y., Lin, Z., Xu, Z., Dong, H., Luo, J., Tian, J., Shi, Z., Huang, L., Zhang, Y., Fan, J., et al.: Trust it or not: Confidence-guided automatic radiology report generation. Neurocomputing **578**, 127374 (2024)
25. Wu, J.T., Agu, N.N., Lourentzou, I., Sharma, A., Paguio, J.A., Yao, J.S., Dee, E.C., Mitchell, W., Kashyap, S., Giovannini, A., et al.: Chest imagenome dataset (version 1.0. 0). PhysioNet **5**, 18 (2021)
26. Xu, Z., Luo, J., Lu, D., Yan, J., Frisken, S., Jagadeesan, J., Wells III, W.M., Li, X., Zheng, Y., Tong, R.K.y.: Double-uncertainty guided spatial and temporal consistency regularization weighting for learning-based abdominal registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 14–24. Springer (2022)
27. Yan, A., He, Z., Lu, X., Du, J., Chang, E., Gentili, A., McAuley, J., Hsu, C.N.: Weakly supervised contrastive learning for chest X-ray report generation. In: Findings of the Association for Computational Linguistics: EMNLP 2021. pp. 4009–4015 (2021). https://doi.org/10.18653/v1/2021.findings-emnlp.336, https://aclanthology.org/2021.findings-emnlp.336
28. Yan, S., Cheung, W.K., Chiu, K., Tong, T.M., Cheung, K.C., See, S.: Attributed abnormality graph embedding for clinically accurate x-ray report generation. IEEE Transactions on Medical Imaging **42**(8), 2211–2222 (2023). https://doi.org/10.1109/TMI.2023.3245608
29. You, D., Liu, F., Ge, S., Xie, X., Zhang, J., Wu, X.: Aligntransformer: Hierarchical alignment of visual regions and disease tags for medical report generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 72–82 (2021)
30. Zhang, Y., Wang, X., Xu, Z., Yu, Q., Yuille, A., Xu, D.: When radiology report generation meets knowledge graph. In: Proceedings of the Conference of Association for the Advance of Artificial Intelligence. pp. 12910–12917 (2020)