IJE TRANSACTIONS B: Applications Vol. 36, No. 08, (August 2023) 1569-1577

International Journal of Engineering

Journal Homepage: www.ije.ir

Cross-modal Deep Learning-based Clinical Recommendation System for Radiology Report Generation from Chest X-rays

S. Shetty*[a,b], V. S. Ananthanarayana[a], A. Mahale[c]

[a] *Department of Information Technology, National Institute of Technology Karnataka, Mangalore, Karnataka, India*
[b] *Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte (Deemed to be University), Udupi, Karnataka, India*
[c] *Deperatment of Radiology, Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, Karnataka, India*

PAPER INFO

Paper history:
Received 01 May 2023
Received in revised form 09 June 2023
Accepted 10 June 2023

Keywords:
Radiology Reports
Deep Learning
Encoder
Decoder
Clinical Recommendation System
Report Generation

ABSTRACT

Radiology report generation is a critical task for radiologists, and automating the process can significantly simplify their workload. However, creating accurate and reliable radiology reports requires radiologists to have sufficient experience and time to review medical images. Unfortunately, many radiology reports end with ambiguous conclusions, resulting in additional testing and diagnostic procedures for patients. To address this, we proposed an encoder-decoder-based deep learning framework that utilizes chest X-ray images to produce diagnostic radiology reports. In our study, we have introduced a novel text modelling and visual feature extraction strategy as part of our proposed encoder-decoder-based deep learning framework. Our approach aims to extract essential visual and textual information from chest X-ray images to generate more accurate and reliable radiology reports. Additionally, we have developed a dynamic web portal that accepts chest X-rays as input and generates a radiology report as output. We conducted an extensive analysis of our model and compared its performance with other state-of-the-art deep learning approaches. Our findings indicate significant improvement achieved by our proposed model compared to existing models, as evidenced by the higher BLEU scores (BLEU1 = 0.588, BLEU2 = 0.4325, BLEU3 = 0.4017, BLEU4 = 0.3860) attained on the Indiana University Dataset. These results underscore the potential of our deep learning framework to enhance the accuracy and reliability of radiology reports, leading to more efficient and effective medical treatment.

doi: *10.5829/ije.2023.36.08b.16*

1. INTRODUCTION

Hospitals around the world heavily rely on medical imaging, which provides valuable insights for disease diagnosis and treatment planning [1, 2]. However, it is crucial for the radiologist to thoroughly examine the medical images in order to provide comprehensive findings and interpretations [3]. The sample chest x-ray images with the associated unstructured medical text data is shown in Figure 1. The radiologists carefully analyze the chest X-rays, which include both frontal and lateral views. Their meticulous examination leads to the creation of comprehensive reports that confirm the diagnosis by documenting their detailed findings [4]. A diagnostic report is a comprehensive document that includes several sections to convey important information about a patient's condition. One of these sections is the indication, which describes the reason why the diagnostic test was ordered. The findings section presents the results of the test, including any abnormalities or other observations that were made. The impression section summarizes the radiologist's overall interpretation of the test results and may include a diagnosis or recommended next steps. Finally, the report may also include manual annotations, which are additional notes or comments that the radiologist has added to provide more context or clarification.

*Corresponding Author Email: shashankshetty.177it002@nitk.edu.in (S. Shetty)

Please cite this article as: S. Shetty, V. S. Ananthanarayana, A. Mahale, Cross-modal Deep Learning-based Clinical Recommendation System for Radiology Report Generation from Chest X-rays, ***International Journal of Engineering, Transactions B: Applications***, Vol. 36, No. 08, (2023), 1569-1577

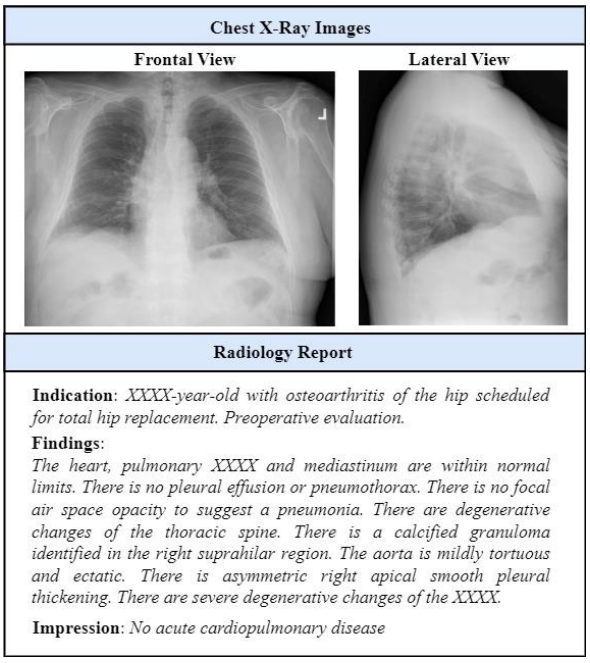

Figure 1. Sample chest X-ray images (i.e., frontal and lateral view) with associated unstructured medical report

In order to produce precise and reliable radiology reports, it is necessary for the radiologist to possess ample experience and devote a significant amount of time to scrutinizing the medical images [5]. A large number of radiology reports may end with inconclusive comments, resulting in patients undergoing additional tests, such as advanced imaging or pathology exams. The issue of the time required for a radiologist to create a detailed report is a significant concern, as on average, an experienced radiologist will need approximately 10-20 minutes to produce a thorough report. In situations such as overcrowded hospitals or during a pandemic [6-8], writing radiology reports can become challenging due to the ever-increasing number of cases [9]. These circumstances inspired our research into developing an automated radiology reporting system using a deep learning framework to faciliate Clinical Recommendation System (CRS) [10].

A CRS is a critical component of modern healthcare delivery systems that are necessary for providing high-quality healthcare. CRS is "*a health information system that assists clinicians in making well-informed decisions about patient care by utilizing patient data, including medical history, current medications, and symptoms, to provide enhanced evidence-based recommendations to clinicians in real-time".* We propose an artificial intelligence (AI)-based CRS framework for generating diagnostic reports from Chest X-Rays (CXR).

The organization of this paper is as follows: Section 2 introduces the problem statement and contribution, followed by section 3, which covers related work on deep learning-based report generation. Section 4 provides a comprehensive methodology, while section 5 focuses on the experimental setup and evaluation. The paper concludes with section 6, which includes the conclusion and discussion, and finally, section 7 presents the references.

2. PROBLEM STATEMENT AND CONTRIBUTION

The increasing demand for accurate and timely radiology reports, coupled with the challenges radiologists face in examining medical images and creating diagnostic notes, has led to burnout, errors, and delays in providing care [11]. While experts have turned to artificial intelligence and deep learning (DL) technologies to automate the generation of radiology reports, implementing and adopting these technologies, face several challenges [12, 13]. These include the need to address concerns regarding the accuracy and reliability of automated notes, integrating these technologies into existing clinical workflows, and ensuring that they are accessible and affordable to all healthcare facilities. Addressing these challenges will be crucial in realizing the potential of AI and DL technologies to improve the speed, accuracy, and efficiency of radiology diagnoses, ultimately enhancing patient care outcomes. The problem statement is defined as follows: "*Considering the multimodal medical cohort containing radiology images with associated diagnostic notes, design and develop an automatic diagnostic report generation by analyzing the visual features from the Chest X-ray scans*".

We propose a solution to the challenge by developing a deep encoder-decoder model that can automatically generate reports from chest X-rays. To achieve this, we have utilized a Multi-channel dilation layer with Depthwise Separable Convolution Neural Network to extract imaging features and knowledge-based text modelling for textual feature extraction. Finally, the Long Short-Term Memory (LSTM) model is used to fine-tune the generated report. We summarize the contributions of this study as follows:

- We propose an encoder-decoder-based deep learning framework to generate diagnostic radiology reports for given chest x-ray images.
- We have developed a dynamic web portal that can efficiently take in chest X-ray images as input and generate radiology reports as output, thereby providing an accessible and user-friendly solution.
- We conduct a comprehensive analysis and compare the performance of the proposed model with the state-of-the-art deep learning approaches.

3. RELATED WORKS

Considerable progress has been made in the field of generating medical descriptions. Yuan et al. [14] introduced an automatic report generation model that utilizes a multiview CNN encoder and a concept-

enriched hierarchical LSTM. The model leverages multi-view information in radiology by employing visual attention in a late fusion manner and enriches the semantics in the hierarchical LSTM decoder with medical concepts. Nguyen et al. [15], presented a set of three modules consisting of classification, generation, and interpretation. For the classification module, a multi-view encoder is employed to extract visual features from chest X-rays, while a text encoder converts reports into embeddings. The generation module utilizes both visual and textual features to create text on a word-by-word basis. Finally, the interpretation module fine-tunes the text generated. Tripathy et al. [16] showcased an automatic report generation model with following stages-NLP Pipeline: (Tokenization, Embedding, Removing special characters etc.); CNN: acts as an encoder in our model. A transfer Learning Model: ChexNet is used to extract the features of the image. Hierarchical LSTMs and Co-Attention mechanism: Hierarchical LSTMs are designed to enrich the representation ability of the LSTM, and the co-attention mechanism provides the context. The sentence and word LSTMs then generate the final reports required. Zhou et al. [17] presented a visual-textual attentive semantic model which uses DenseNet201 as a visual encoder model and BioSentVec as a text encoder. The LSTM model is utilized to generate the report.

A Knowledge Graph Auto-Encoder (KGAE) model is proposed by Liu et al. [18], which utilizes independent sets of Chest X-ray images and their associated reports during the training phase. KGAE consists of a pre-constructed knowledge graph, a knowledge-driven encoder and a knowledge-driven decoder. They have used the knowledge-driven encoder to project medical images and reports to the corresponding coordinates in latent space and the knowledge-driven decoder to generate a medical report on a given coordinate in that space. Sirshar et al. [19] propose an encoder-decoder model with CNN used as a visual encoder and an RNN decoder with attention used to produce the radiology reports. Nicolson et al. [20] presented the report generation framework, where the DenseNet pretrained on imageNet is used as an encoder for imaging feature extraction, and the Bidirectional Encoder Representations from Transformers (BERT) NLP encoder is utilized for textual feature extraction. The decoder model with attention is incorporated for report generation. The various general domain and domain-specific pre-trained checkpoints are evaluated and the best checkpoints are chosen for warm starting the encoder-decoder of a CXR report generator. These warm starting helps generate a diagnostically accurate report that can be used in a clinical setting. From the literature it is observed that there is a significant need for improving performance and the quality of the generated report.

The research paper introduced a deep learning model called CDGPT2, which aimed to generate radiology reports based on chest X-rays sourced from the Indiana University dataset [21]. To extract both visual and textual features, the model incorporated pre-trained Chexnet and ChatGPT2. However, the study identified limitations in the model's performance attributed to the small size of the available data. In a separate investigation, Babar et al. [22] introduced a novel metric known as Diagnostic Content Score (DCS). They initially created Diagnostic Tags for each report, leveraging them as external knowledge. By utilizing these tags, they developed a probabilistic model based on the training data. Subsequently, the model was employed to assess the diagnostic quality of the generated reports from the test data. However, the approach exhibited limitations, as indicated by a reduced Bleu4 score of 0.12 on the Indiana University dataset.

The accurate interpretation and summary of medical images, particularly those generated by radiology tests such as X-rays, CT scans, and MRIs, are crucial components of clinical diagnosis. Generating a diagnosis report from radiology images is an essential step in clinical diagnosis, and highly skilled radiologists are required for this task. However, the process can be time-consuming and mentally taxing for radiologists, especially in busy and overcrowded situations. To alleviate this burden and speed up the diagnosis process, there is a growing need for automated and reliable diagnostic report generation frameworks. Existing deep learning techniques for report generation have shown promise, but there is still room for improvement, particularly in terms of the BLEU score [14-22]. One promising approach is to develop a cross-modal framework that combines textual and imaging features to assist radiologists in automatically generating accurate reports from medical images. The proposed cross-modal framework leverages the knowledge base to extract textual features and incorporates multi-scale feature extraction from chest X-ray images. This approach facilitates the extraction of highly discriminative features, resulting in enhanced performance compared to existing methodologies. By using such frameworks, healthcare providers can reduce the workload on radiologists, speed up the diagnosis process, and provide better patient care. Additionally, these frameworks can ensure consistency and accuracy in diagnosis reports, minimizing the risk of errors and improving the overall quality of patient care.

4. METHODOLOGY

The proposed encoder-decoder framework aims to generate radiology reports from chest X-rays, which include both frontal and lateral images. During the training process, both the chest X-rays and the corresponding reports are provided as input to the encoder. The encoder consists of two components: the Unimodal Medical Visual Encoding Subnetwork (UM-

VES) for extracting visual features and the Unimodal Medical Text Embedding Subnetwork (UM-TES) for extracting textual features. These features are then used by the LSTM-based decoder to generate the reports. The encoder operates by processing each item in the input sequence and aggregating the captured information into a context vector. Once the entire input sequence has been processed, the encoder transfers the context vector to the decoder, which generates the output sequence item by item. This process allows the model to effectively combine the visual and textual information and generate contextually relevant reports.

The detailed architecture of the proposed cross-modal retrieval is shown in Figure 2. During the training phase, the model aims to establish connections between the textual information in the reports and the visual features extracted from the chest X-ray images. The UM-TES approach is employed to encode the textual information, while the UM-VES technique is used to extract visual features. These modalities are then integrated into a joint representation, enabling the model to learn the correlations between the input chest X-ray images and the associated textual information in the reports. By iteratively optimizing the model's parameters, it gradually acquires the capability to generate coherent and contextually relevant reports.

During the testing phase, only the chest X-ray images are provided as input to the trained model. Drawing upon the learned associations between the image and the textual information from the training phase, the model generates a report based solely on the input image. This process is achieved by utilizing the decoding mechanism of the trained model, such as a Long Short-Term Memory (LSTM), to generate the text-based output.

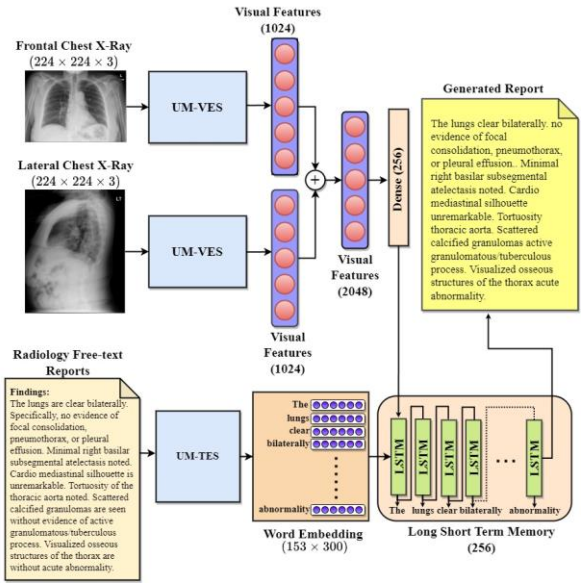

Figure 2. Overall architecture of the proposed cross-modal deep learning-based model for automatic report generation

4. 1. Unimodal Medical Visual Encoding Subnetwork (UM-VES) The UM-VES technique suggested employs a depthwise separable convolution neural network with a multichannel dilation layer to extract imaging features. The suggested multichannel dilation convolution layer provides more comprehensive imaging data by generating a larger receptive field, while keeping the network parameters constant, in contrast to the traditional convolutional layer. Moreover, to ensure an even distribution of computational workload across each layer, the Depthwise Separable convolution network is utilized instead of the conventional convolution network [23]. The UM-VES framework is used to extract visual features from both the frontal and lateral CXR images independently, and the resulting features are combined by concatenation. The overall architecture of the proposed UM-VES model is shown in Figure 3. The UM-VES model is composed of three parallel dilation channels that capture imaging features with a wider receptive field. The resulting features are then concatenated and passed through 13 depthwise separable layers to learn and extract additional features. For a more comprehensive understanding of the model, readers can refer to our previous paper [23], where we provide a detailed overview and description of the UM-VES model's architecture and components.

4. 2. Unimodal Medical Text Embedding Subnetwork (UM-TES) The radiology findings are subjected to pre-processing, which involves removing stop words and punctuation, as well as performing stemming to retain root words. Additionally, tokenization is applied to extract important latent medical concepts. Customized Clinical Knowledge-based Text Modelling is utilized to learn word embeddings from medical terminology. During the training of the text model, the glove word embeddings [24] are combined with the word embeddings obtained from a knowledge base of 4.5 million Stanford reports [25, 26]. This combination enhances the effectiveness of the text model by leveraging the information contained in the knowledge base. The dense word embeddings generated are then mapped to medical terms from the findings using the Embedding Layer. The detailed architecture of the proposed UM-TES model is shown in Figure 4. To gain a more thorough understanding of the UM-TES model's architecture and components, readers can refer to our previous paper [23], where we offer a detailed overview and description.

4. 3. Long Short-term Memory-based Report Generation The fundamental concept of utilizing LSTM for report generation centers around the memory cell, denoted as c, which primarily stores the information on the input received at any given moment. The function of these cells is controlled by layers or gates that are inserted in a multiplicative manner and can maintain values of either 0 or 1 which are determined by the gates.

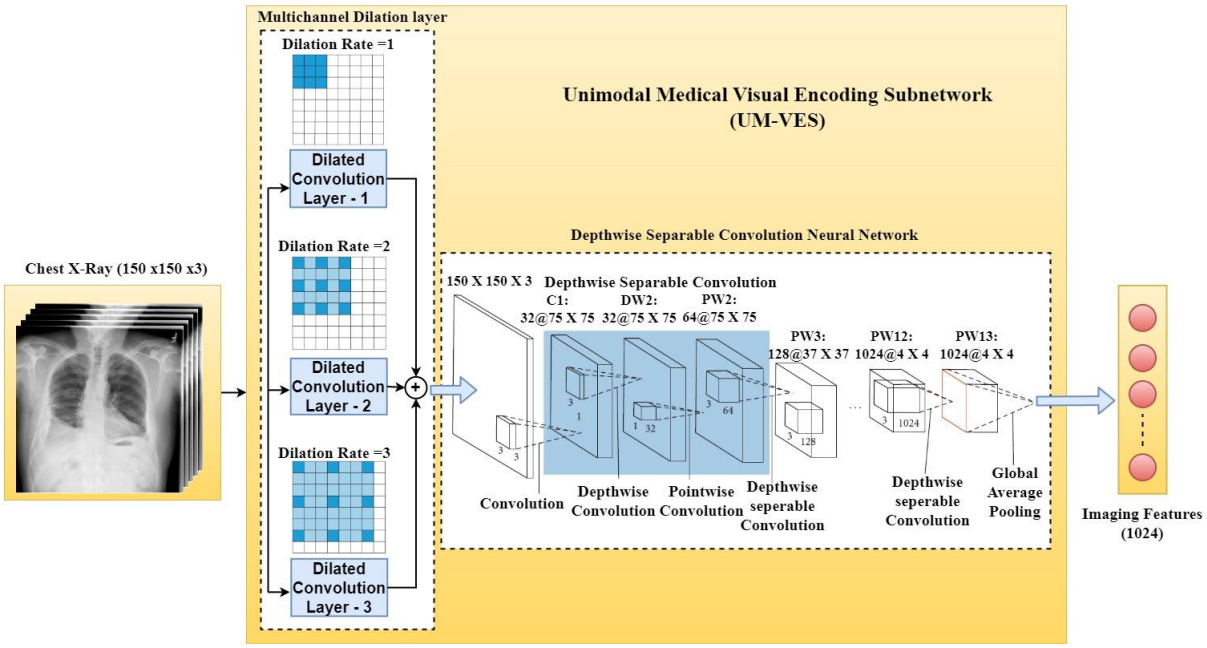

Figure 3. Overall architecture of the UM-VES model

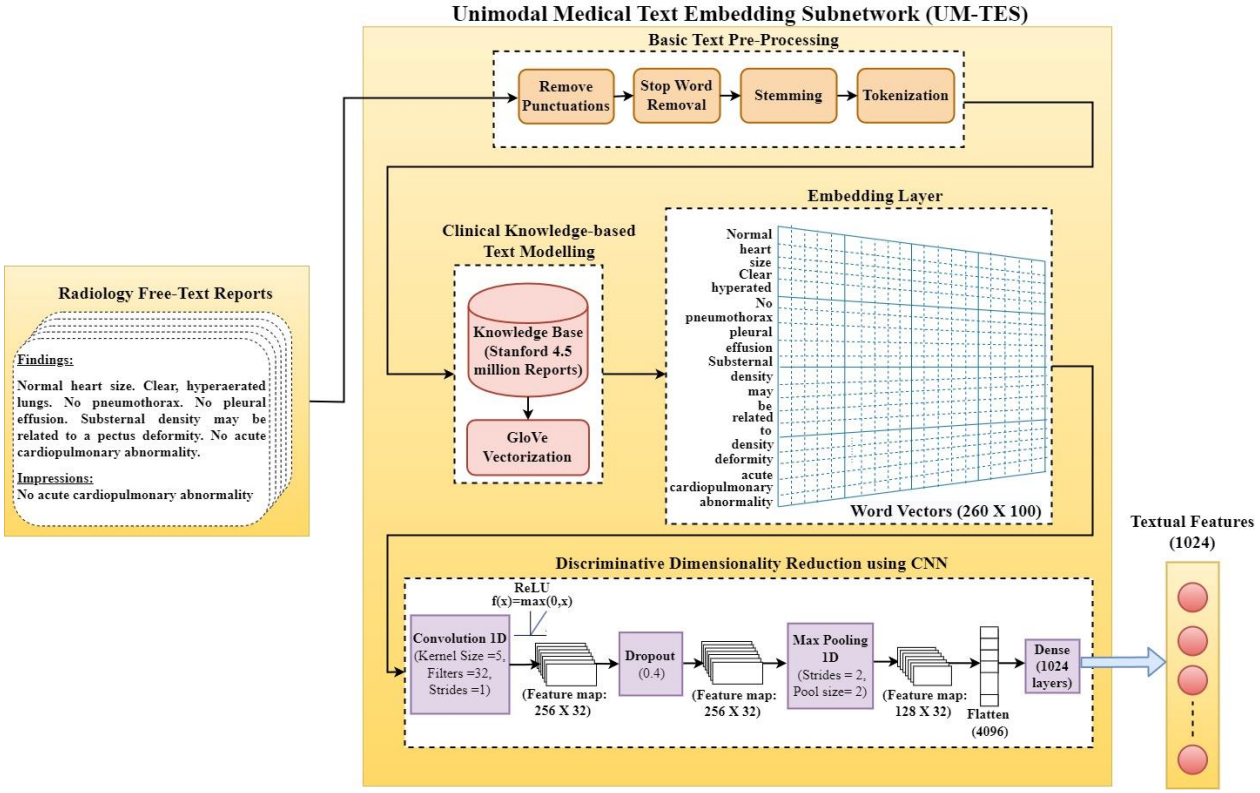

Figure 4. Overall architecture of the UM-TES model

Specifically, three gates are employed to monitor whether the present value of the cell should be disregarded, if the new cell value should be generated (output gate 0), or if it should be interpreted as input, as illustrated in Figure 5.

Equations (1), (2) and (3) depicts the input, forget, and output layers, respectively.

$$input_t = \sigma\left(W_{iy}y_t + W_{im}m_{t-1}\right) \quad (1)$$

$$forget_t = \sigma(W_{fy}y_t + W_{fm}m_{t-1}) \quad (2)$$

$$output_t = \sigma(W_{oy}y_t + W_{om}m_{t-1}) \quad (3)$$

Equations (4), (5) and (6) represent the other operation of the LSTM model.

$$cell_t = forget_t \odot cell_{t-1} + input_t \odot h(W_{cy}y_t + W_{cm}m_{t-1}) \quad (4)$$

$$cell_t = output_t \odot cell_t \quad (5)$$

$$P_{t+1} = Softmax(m_t) \quad (6)$$

where, $input_t$, $forget_t$ and $output_t$ denotes the output of the input, forget and output gates, respectively at time t; y_t represents the input vector at time t; $m_t - 1$ is the hidden state of the LSTM at time $t - 1$; W_{iy}, W_{fy}, W_{oy}, W_{cy}, W_{im}, W_{fm}, W_{om} and W_{cm} indicate the weight matrices that manage that manage the input and hidden connections between the input, forget and output gates and cells; $cell_t$ represents the state of the cell at time t and P_{t+1} represents a probability distribution over a set of possible outcomes at time $t + 1$.

4. 4. Web-based Framework for Report Generation We utilized the Flask web framework to create a user-friendly web interface for our model. By uploading both frontal and lateral X-ray images through this interface, users can obtain reports with ease. To streamline the user experience, we implemented Ajax, a technique that enables data to be sent and retrieved asynchronously in the background of the application without requiring the entire page to be reloaded. This approach is particularly useful when we want to update specific portions of an existing page without redirecting or reloading the page for the user. As depicted in Figure 6, in order to obtain a report, users are required to upload both frontal and lateral X-ray images. After clicking on the 'Generate Report' button, an Ajax request is sent to the Flask App hosted on the server. The Flask application

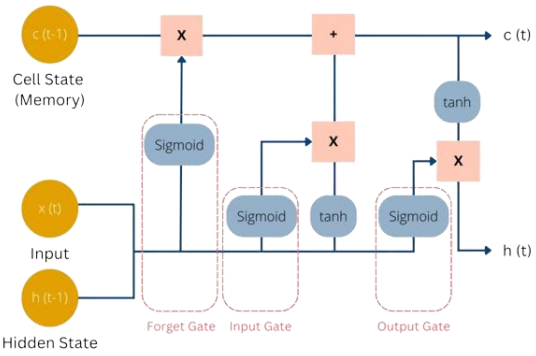

Figure 5. Long Short-term Memory Architecture

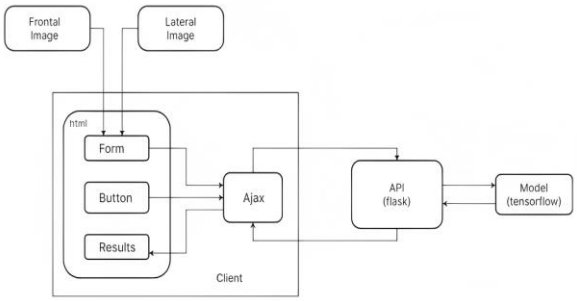

Figure 6. Client-Server interaction used for predicting reports

utilizes the uploaded images to generate predictions for the report, which are then transmitted back to the client side. Upon receipt, the predicted report is displayed to the users.

5. EXPERIMENTAL SETUP AND EVALUATION

For model training, we utilized the IU Chest X-Ray collection [27], which includes a comprehensive set of chest x-ray images accompanied by their corresponding diagnostic reports. The cohort of multimodal medical data consisted of 7,470 pairs of images and reports, with a total of 3,996 cases. The reports contained two main sections, impressions and findings. In our investigation, we selected frontal and lateral images and the content of the findings section as the target captions to be generated. To conduct our experiment, we removed cases without reports and frontal/lateral images, ultimately working with 3,638 cases. Two methods were used to generate text reports: greedy search [28] and beam search [29]. Greedy search is an algorithmic approach that incrementally constructs a solution by selecting the next piece that seems to provide the most immediate benefit. In contrast, beam search expands on the greedy search technique by generating a list of the most likely output sequences, each with its own score. The sequence with the highest score is then chosen as the final result.

To evaluate the performance of the generated reports, we incorporated the BLEU score [30]. The Bilingual Evaluation Understudy (BLEU) Score is a method used to evaluate the similarity between a generated sentence and a reference sentence. The score ranges from 0.0, indicating a total mismatch, to 1.0, indicating a perfect match. This approach involves tallying the number of matching n-grams in the candidate text with those in the reference text. For instance, a uni-gram or 1-gram would correspond to each token, whereas a bi-gram comparison would correspond to each pair of words. Achieving a perfect score is not practical, as it necessitates an exact match with the reference, which even human translators cannot achieve. Furthermore,

comparing scores across datasets can be difficult due to the number and quality of the references used to determine the BLEU score.

We computed the BLEU score for an automatic report generated using beam and greedy search. It is observed that beam search produces a superior BLEU score compared to the greedy search algorithm. The qualitative analysis of the proposed deep learning-based model using a beam and greedy search algorithm is shown in the Table 1. The BLEU score of 0.5459, 0.4131, 0.386 and 0.3552 is obtained for different n-grams in the greedy search approach. The beam search approach produces a BLEU score of 0.5881, 0.4325, 0.4017 and 0.3860. We have also compared the results with the existing automatic diagnostic report generation work. Most of the existing work has shown lesser BLEU4 as it compares the four words together with the ground truth. Our proposed model outperforms the existing models while generating robust diagnostic reports. This may be due to the multi-channel visual features and knowledge-based discriminate text features extracted in the encoder of the proposed network. The detailed analysis with the various existing model is summarized in Table 2.

We designed and developed a flask web application interface for quantitative analysis of the model. Figure 7 shows the web interface to upload the chest x-ray images and produce the diagnostic report. The user has to input frontal and lateral chest X-ray images to the web interface. When the user clicks the "generate report", an Ajax request will be sent to the Flask App on the server, where the Flask application uses the uploaded images to predict reports. The predicted reports will be sent back to the client, where they are displayed to the users with the BLEU score. Figures 8 and 9 present two samples of reports generated using the proposed framework.

TABLE 1. Performance analysis of the proposed model

Method	Bleu1	Bleu2	Bleu3	Bleu4
Greedy Search	0.5459	0.4131	0.3864	0.3552
Beam Search	0.5881	0.4325	0.4017	0.3860

TABLE 2. Performance analysis compared with existing work of report generation

Method	Bleu1	Bleu2	Bleu3	Bleu4
Tripathy et al. [16], 2021	0.213	0.258	0.325	0.381
Nguyen et al. [15], 2021	0.515	0.378	0.293	0.235
Liu et al. [18], 2021	0.417	0.263	0.181	0.126
Zhou et al. [17], 2021	0.536	0.392	0.314	0.339
Sirshar et al. [19], 2022	0.58	0.342	0.263	0.155
Nicolson et al. [20], 2022	0.4777	0.308	0.2274	0.1773
Proposed Model	**0.5881**	**0.4325**	**0.4017**	**0.3860**

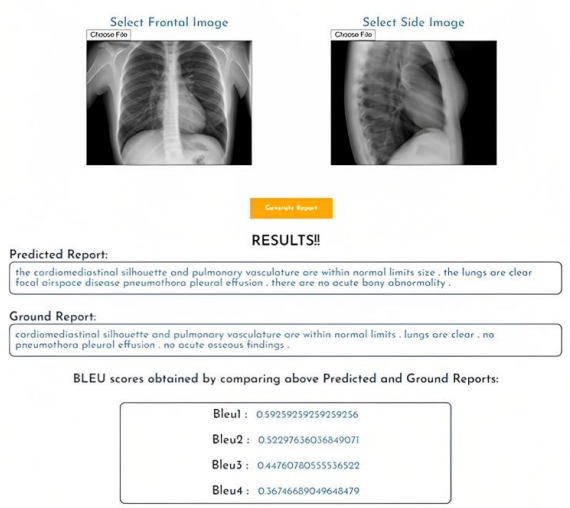

Figure 7. The dynamic web portal for automatic diagnostic report generation

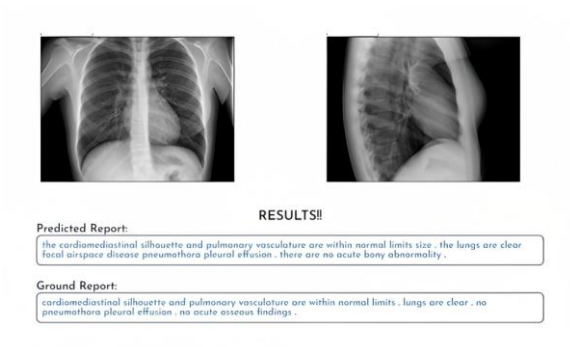

Figure 8. Generated Report (Sample 1)

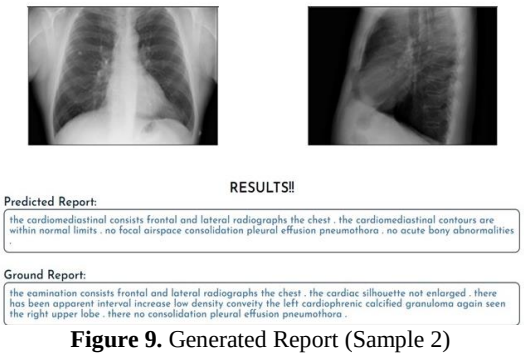

Figure 9. Generated Report (Sample 2)

In summary, an automated framework that employs a deep learning-based encoder-decoder approach to generate reports from chest X-ray scans. The modules used in the framework, such as UM-VES, UM-TES, and LSTM, are discussed in detail. In addition, a dynamic web framework was developed and implemented that accepts chest X-ray images as input and generates diagnostic reports as output. To evaluate the proposed

framework, a comprehensive set of experiments was conducted, and the results were compared with those of state-of-the-art report generation frameworks. The proposed framework yielded better performance, as evidenced by an improved BLEU score compared to existing models.

6. CONCLUSION AND FUTURE WORK

In this paper, we aimed to develop a deep learning-based model that can accurately and automatically generate diagnostic reports from CXR images. To achieve this, we employed a cross-modal retrieval technique that retrieves radiology reports from the image. Our approach, which utilized the beam search method, outperformed existing models in generating robust diagnostic reports. This can be attributed to the encoder of our proposed network, which extracted multi-channel visual features and discriminative text features based on knowledge. Compared to existing models, our approach showed superior results in terms of BLEU4 scores, which is a standard metric used to compare the accuracy of generated text to the ground truth. In addition, we created a dynamic web portal that allows for the easy uploading of frontal and lateral CXR images, and provides the corresponding diagnostic reports as output. This feature greatly simplifies the report writing process for radiologists, as it automates the process and saves time.

One potential limitation of the proposed work is the need to assess the generalizability of the model. While the deep learning framework exhibited notable enhancements in generating accurate and reliable radiology reports compared to existing models, it is crucial to recognize that the evaluation and analysis were restricted to the Indiana University Dataset. The performance of the model may differ when applied to alternative datasets or diverse clinical settings. Therefore, additional research and validation on varied datasets and real-world scenarios are imperative to determine the generalizability and robustness of the proposed approach.

Furthermore, we plan to expand the scope of our model by applying it to other types of diagnostic images such as MRI, ultrasound and CT scans. This will allow us to further evaluate the effectiveness and robustness of our proposed model across different modalities and ultimately improve its overall utility in clinical settings.

7. REFERENCES

1. Shetty, S. and Mahale, A., "Comprehensive review of multimodal medical data analysis: Open issues and future research directions", ***Acta Informatica Pragensia***, Vol. 11, No. 3, (2022), 423-457. doi: 10.18267/j.aip.202.
2. Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G. and Murphy, K., "Deep learning for chest x-ray analysis: A survey", ***Medical Image Analysis***, Vol. 72, (2021), 102125. https://doi.org10.1016/j.media.2021.102125
3. Ramirez-Alonso, G., Prieto-Ordaz, O., López-Santillan, R. and Montes-Y-Gómez, M., "Medical report generation through radiology images: An overview", ***IEEE Latin America Transactions***, Vol. 20, No. 6, (2022), 986-999. doi: 10.1109/TLA.2022.9757742.
4. Monshi, M.M.A., Poon, J. and Chung, V., "Deep learning in generating radiology reports: A survey", ***Artificial Intelligence in Medicine***, Vol. 106, (2020), 101878. https://doi.org/10.1016/j.artmed.2020.101878
5. Jing, B., Xie, P. and Xing, E., "On the automatic generation of medical imaging reports", ***arXiv preprint arXiv:1711.08195***, (2017). doi: 10.18653/v1/P18-1240.
6. Dey, A., "Cov-xdcnn: Deep learning model with external filter for detecting covid-19 on chest x-rays", in Computer, Communication, and Signal Processing: 6th IFIP TC 5 International Conference, ICCCSP 2022, Chennai, India, February 24–25, 2022, Revised Selected Papers, Springer. (2022), 174-189.
7. Shetty, S. and Mahale, A., "Ms-chexnet: An explainable and lightweight multi-scale dilated network with depthwise separable convolution for prediction of pulmonary abnormalities in chest radiographs", ***Mathematics***, Vol. 10, No. 19, (2022), 3646. https://doi.org/10.3390/math10193646
8. Alahmari, S.S., Altazi, B., Hwang, J., Hawkins, S. and Salem, T., "A comprehensive review of deep learning-based methods for covid-19 detection using chest x-ray images", ***IEEE Access***, (2022). doi: 10.1109/ACCESS.2022.3208138.
9. Yang, S., Wu, X., Ge, S., Zhou, S.K. and Xiao, L., "Knowledge matters: Chest radiology report generation with general and specific knowledge", ***Medical Image Analysis***, Vol. 80, (2022), 102510. https://doi.org/10.1016/j.media.2022.102510
10. Carson, E.R., Cramp, D.G., Morgan, A. and Roudsari, A.V., "Clinical decision support, systems methodology, and telemedicine: Their role in the management of chronic disease", ***IEEE Transactions on Information Technology in Biomedicine***, Vol. 2, No. 2, (1998), 80-88. doi: 10.1109/4233.720526.
11. Abtahi, Z., Sahraeian, R. and Rahmani, D., "A stochastic model for prioritized outpatient scheduling in a radiology center", ***International Journal of Engineering Transactions A: Basics***, Vol. 33, No. 4, (2020). doi: 10.5829/ije.2020.33.04a.11.
12. Khatami, A., Babaie, M., Tizhoosh, H., Nazari, A., Khosravi, A. and Nahavandi, S., "A radon-based convolutional neural network for medical image retrieval", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 31, No. 6, (2018), 910-915. doi: 10.5829/ije.2018.31.06c.07.
13. Gheitasi, A., Farsi, H. and Mohamadzadeh, S., "Estimation of hand skeletal postures by using deep convolutional neural networks", ***International Journal of Engineering, Transactions A: Basics***, Vol. 33, No. 4, (2020), 552-559. doi: 10.5829/ije.2020.33.04a.06.
14. Yuan, J., Liao, H., Luo, R. and Luo, J., "Automatic radiology report generation based on multi-view image fusion and medical concept enrichment", in Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22, Springer. (2019), 721-729.
15. Nguyen, H.T., Nie, D., Badamdorj, T., Liu, Y., Zhu, Y., Truong, J. and Cheng, L., "Automated generation of accurate\& fluent medical x-ray reports", arXiv preprint arXiv:2108.12126, (2021). doi: 10.18653/v1/2021.emnlp-main.288.
16. Tripathy, B., Sai, R.R. and Banu, K.S., Automated medical report generation on chest x-ray: Images using co-attention mechanism,

in Hybrid computational intelligent systems. 2023, CRC Press.111-122.

17. Zhou, X., Li, Y. and Liang, W., "Cnn-rnn based intelligent recommendation for online medical pre-diagnosis support", ***IEEE/ACM Transactions on Computational Biology and Bioinformatics***, Vol. 18, No. 3, (2020), 912-921. doi: 10.1109/TCBB.2020.2994780.
18. Liu, F., You, C., Wu, X., Ge, S. and Sun, X., "Auto-encoding knowledge graph for unsupervised medical report generation", ***Advances in Neural Information Processing Systems***, Vol. 34, (2021), 16266-16279. https://arxiv.org/abs/2111.04318
19. Sirshar, M., Paracha, M.F.K., Akram, M.U., Alghamdi, N.S., Zaidi, S.Z.Y. and Fatima, T., "Attention based automated radiology report generation using cnn and lstm", ***Plos one***, Vol. 17, No. 1, (2022), e0262209. doi: 10.1371/journal.pone.0262209.
20. Nicolson, A., Dowling, J. and Koopman, B., "Improving chest x-ray report generation by leveraging warm-starting", arXiv preprint arXiv:2201.09405, (2022). https://arxiv.org/abs/2201.09405
21. Alfarghaly, O., Khaled, R., Elkorany, A., Helal, M. and Fahmy, A., "Automated radiology report generation using conditioned transformers", ***Informatics in Medicine Unlocked***, Vol. 24, (2021), 100557. doi: 10.1016/j.imu.2021.100557.
22. Babar, Z., van Laarhoven, T., Zanzotto, F.M. and Marchiori, E., "Evaluating diagnostic content of ai-generated radiology reports of chest x-rays", ***Artificial Intelligence in Medicine***, Vol. 116, (2021), 102075. https://doi.org10.1016/j.artmed.2021.102075
23. Shetty, S. and Mahale, A., "Multimodal medical tensor fusion network-based dl framework for abnormality prediction from the radiology cxrs and clinical text reports", ***Multimedia Tools and Applications***, (2023), 1-48. https://doi.org/10.1007/s11042-023-14940-x.
24. Pennington, J., Socher, R. and Manning, C.D., "Glove: Global vectors for word representation", in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). (2014), 1532-1543.
25. Zhang, Y., Ding, D.Y., Qian, T., Manning, C.D. and Langlotz, C.P., "Learning to summarize radiology findings", arXiv preprint arXiv:1809.04698, (2018). http://arxiv.org/abs/1809.04698
26. Shetty, S., Ananthanarayana, V. and Mahale, A., "Medical knowledge-based deep learning framework for disease prediction on unstructured radiology free-text reports under low data condition", in Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference: Proceedings of the EANN 2020 21, Springer. (2020), 352-364.
27. Demner-Fushman, D., Kohli, M.D., Rosenman, M.B., Shooshan, S.E., Rodriguez, L., Antani, S., Thoma, G.R. and McDonald, C.J., "Preparing a collection of radiology examinations for distribution and retrieval", ***Journal of the American Medical Informatics Association***, Vol. 23, No. 2, (2016), 304-310. doi: 10.1093/jamia/ocv080.
28. Gu, J., Cho, K. and Li, V.O., "Trainable greedy decoding for neural machine translation", arXiv preprint arXiv:1702.02429, (2017). doi: 10.18653/v1/D17-1210.
29. Freitag, M. and Al-Onaizan, Y., "Beam search strategies for neural machine translation", arXiv preprint arXiv:1702.01806, (2017). https://doi.org/10.48550/arXiv.1702.01806
30. Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J., "Bleu: A method for automatic evaluation of machine translation", in Proceedings of the 40th annual meeting of the Association for Computational Linguistics. (2002), 311-318.

Persian Abstract

چکیده

تولید گزارش رادیولوژی یک وظیفه حیاتی برای رادیولوژیست ها است و خودکار کردن این فرآیند می تواند حجم کار آنها را به میزان قابل توجهی ساده کند. با این حال، ایجاد گزارش های رادیولوژی دقیق و قابل اعتماد مستلزم آن است که رادیولوژیست ها تجربه و زمان کافی برای بررسی تصاویر پزشکی داشته باشند. متأسفانه، بسیاری از گزارش‌های رادیولوژی با نتیجه‌گیری‌های مبهم خاتمه می‌یابند که منجر به آزمایش‌های اضافی و روش‌های تشخیصی برای بیماران می‌شود. برای پرداختن به این موضوع، ما یک چارچوب یادگیری عمیق مبتنی بر رمزگذار–رمزگشا پیشنهاد کردیم که از تصاویر اشعه ایکس قفسه سینه برای تولید گزارش‌های رادیولوژی تشخیصی استفاده می‌کند. در مطالعه خود، یک مدل سازی متن جدید و استراتژی استخراج ویژگی بصری را به عنوان بخشی از چارچوب یادگیری عمیق مبتنی بر رمزگذار–رمزگشای پیشنهادی خود معرفی کرده‌ایم. هدف رویکرد ما استخراج اطلاعات بصری و متنی ضروری از تصاویر اشعه ایکس قفسه سینه برای تولید گزارش‌های رادیولوژی دقیق‌تر و قابل اعتمادتر است. علاوه بر این، ما یک پورتال وب پویا ایجاد کرده ایم که اشعه ایکس قفسه سینه را به عنوان ورودی می پذیرد و گزارش رادیولوژی را به عنوان خروجی تولید می کند. ما تجزیه و تحلیل گسترده ای از مدل خود انجام دادیم و عملکرد آن را با دیگر رویکردهای پیشرفته یادگیری عمیق مقایسه کردیم. یافته‌های ما نشان‌دهنده بهبود قابل‌توجهی است که توسط مدل پیشنهادی ما در مقایسه با مدل‌های موجود به دست آمده است، همانطور که با نمرات BLEU بالاتر (BLEU1 = 0.588، BLEU2 = 0.4325، BLEU3 = 0.4017، BLEU4 = 0.3860) به دست آمده در مجموعه داده‌های دانشگاه ایندیانا مشهود است. این نتایج بر پتانسیل چارچوب یادگیری عمیق ما برای افزایش دقت و قابلیت اطمینان گزارش های رادیولوژی تاکید می کند که منجر به درمان پزشکی کارآمدتر و موثرتر می شود.