Journal of Chemical Health Risks

www.jchr.org

JCHR (2024) 14(3), 3366-3374 | ISSN:2251-6727

An Advanced Healthcare Imaging System for Analyzing X-Ray Image

Selvi S[1], Vanathi A[2], Kalpana B[3]

[1]Associate Professor, Dept. of Computer Science and Engg., Government College of Engineering, Bargur, Tamil Nadu, India
[2]Associate Professor, Dept. of Computer Science and Engg., Aditya University, Kakinada, Andra Pradesh, India
[3]Associate Professor, Dept. of Computer Science and Engg., RMD Engineering College, Tamil Nadu, India

(Received: 04 February 2024 Revised: 11 March 2024 Accepted: 08 April 2024)

KEYWORDS

Deep Learning, Medical Imaging,

X-ray Analysis, Transformer-based Model, Radiology Interpretation, Intelligent Assistant

ABSTRACT:

In the modern healthcare landscape, medical imaging has become an essential component, providing clinicians with detailed insights into the human body's complex internal structures. Traditionally, doctors bore the sole responsibility for interpreting X-rays, a process that was not only labour-intensive but also prone to occasional errors, leading to potential misdiagnoses or delays in treatment. This study aims to enhance the interpretation of medical X-ray images by employing deep learning techniques to autonomously generate captions. Utilizing a trained transformer-based model, the system analyses incoming X-ray images to generate descriptive textual summaries of observed findings. The proposed methodology integrates image preprocessing, model inference, and user interface development to facilitate seamless engagement for healthcare professionals. Through rigorous examination and refinement, the system achieves high accuracy and reliability in caption generation, thereby improving diagnostic efficiency and patient care. Ongoing monitoring and upkeep ensure the system remains robust and adaptable to evolving healthcare requirements. Ultimately, this initiative advances AI-driven solutions in medical imaging, addressing challenges in radiology interpretation and enhancing healthcare outcomes.

1. Introduction

Medical X-ray imaging plays a crucial role noninvasively and painlessly helping in diagnosing and monitoring a wide range of health conditions. Among the various imaging techniques available, X-rays have proven to be a critical instrument, producing detailed images of bones and organs that form the foundation of medical diagnostics. However, interpreting these images is a complex task that requires significant training and expertise. As a result, there is a pressing need for more efficient methods of interpreting X-ray images. This is where artificial intelligence (AI) and deep learning come into play. These state-of-the-art technologies have the potential to revolutionize X-ray analysis by significantly improving the speed and accuracy of interpretation.

Implementation of the system in clinical settings enables real-time interpretation of X-ray images, enhance the capabilities of medical professionals, offering healthcare practitioners immediate insights into medical findings, and improve patient care outcomes. It is a keen requirement of developing a system that can quickly analyzed X-ray images and provide concise yet comprehensive summaries of the findings. Hence, the resultant intelligent assistant will augment the capabilities of healthcare providers, streamline their workflow, and facilitate faster healthcare decisions, ultimately improving patient care. By integrating technology seamlessly into medical practice, it is proposed to raise the standard of care and improve the overall quality of healthcare delivery.

The literature survey provides an overview of recent advancements in medical image captioning, highlighting the introduction of sophisticated methodologies leveraging cutting-edge deep learning architectures and techniques. For instance, studies have proposed deep learning models augmented with LSTM networks and attention mechanisms tailored for chest X-ray image captioning, aiming to alleviate the workload of medical

Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(3), 3366-3374 | ISSN:2251-6727

professionals [1]. Another innovative fusion of deep learning and retrieval methods has shown promise in addressing previously unseen abnormalities and refining generated captions [2]. Transformer-based models like the Cross Encoder-Decoder Transformer (CEDT) have emerged to effectively attain both global and local image features, thereby enhancing the accuracy and comprehensiveness of generated reports [3].

Efforts have also been made to overcome specific challenges inherent in medical image captioning, such as integrating patient information and reconciling vocabulary discrepancies between models [4]. Techniques such as self-boosting frameworks and feature difference exploitation have been proposed to refine diagnostic report generation, emphasizing collaborative interactions between image and text branches to improve report quality [5, 6]. Furthermore, studies explore the adaptability of pre-trained language-image models to medical contexts, showcasing robust performance in producing best-quality captions for medical images [7].

Initiatives focused on evaluating the effectiveness of AI-generated captions in clinical settings, including comparisons with radiologists' reports, highlight the potential for AI-assisted diagnosis [8]. Additionally, frameworks tailored for specific medical imaging modalities, such as knee X-rays, show promise in automating report generation and lay the foundation for future enhancements [9]. These endeavours emphasize the collaborative efforts between AI researchers and healthcare professionals, aiming to seamlessly integrate advanced technologies into clinical workflows to enhance patient care and diagnostic accuracy.

The surveys on automated generation of medical imaging reports [10] and biomedical image captioning [11] provide comprehensive overviews of deep learning techniques utilized in medical imaging analysis. They explore various architectures, datasets like IU X-Ray, PEIRGROSS, and ICLEF-CAPTION, and evaluation methods, acknowledging limitations such as dataset diversity and private data challenges. These papers set the stage for future research directions aimed at enhancing clinical applications through expanded datasets and improved methodologies.

The survey also includes references to key studies such as CheXpert [12], MIC [13], Multi-stage Medical Image Captioning using Classification and CLIP [14], Diagnostic captioning [15], and Image captioning model using attention and object features to mimic human image [16]. These studies collectively demonstrate the significant progress made in medical image captioning, addressing challenges such as dataset limitations, complex medical terminology, and the need for improved evaluation metrics. Furthermore, they highlight the potential of advanced AI techniques to enhance medical diagnostics and streamline clinical workflows, paving the way for improved patient care and diagnostic accuracy. Additional references, Hierarchical X-Ray Report Generation via Pathology tags and Multi Head Attention [17], Medical Image Captioning Using Optimized Deep Learning Model [18], and AEHRC CSIRO at ImageCLEFmed Caption [19], further enrich the literature survey by providing insights into specific methodologies and their applications in medical image captioning.

Related works in medical image captioning leverage deep learning techniques to automatically generate captions for various medical images, particularly chest X-rays. These systems utilize large datasets such as IU Chest X-ray and CheXpert, containing thousands of images with corresponding textual reports. They employ sophisticated models corresponding convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based architectures towards extract features from images and generate descriptive captions. Despite achieving promising results, limitations include challenges in handling uncertainties in image interpretation, maintaining dataset diversity, and ensuring accurate evaluation metrics. Future research aims to specify the abovesaid challenges and further enhance the capabilities of medical image captioning systems.

2. Objectives

The increasing demand for efficient interpretation methods in medical X-ray analysis drives the adoption of artificial intelligence (AI) and deep learning technologies. Traditional interpretation processes are time-consuming and require extensive expertise, leading to delays in diagnosis and treatment. By implementing AI systems, real-time interpretation can be achieved, empowering healthcare professionals with immediate insights and improving patient care outcomes.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2024) 14(3), 3366-3374 | ISSN:2251-6727

Developing intelligent assistants capable of quickly analysing X-ray images and providing concise summaries is essential to enhance the efficiency of medical workflows. Integrating technology seamlessly into clinical practice aims to streamline processes and improve overall patient care quality.

3. Methods

The proposed system architecture is shown in Figure 1 consist of following components Input Image, Preprocessing, Feature Vector Extraction, Attention Mechanism, Decoder, Generated Caption. The working process is described as next which is shown in Figure 2.

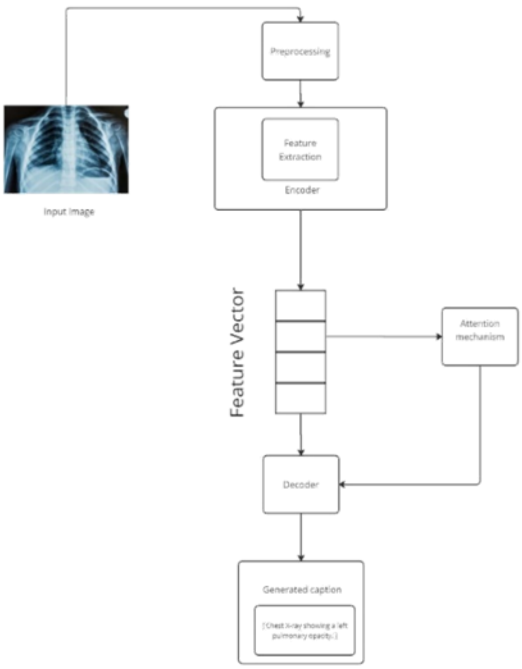

Fig. 1 System Architecture for Intelligent Healthcare Imaging System

The loading of pre-trained deep learning models marks a pivotal step in the application's workflow. The heart of the application lies wherein frames are captured from the webcam in real-time. The preprocessing step is crucial for ensuring uniformity in the orientation of captured images, which can be particularly important for certain image recognition tasks.

Defining the Region of Interest (ROI) within the frame is the subsequent step, with a bounding box drawn around it via cv2.rectangle(). This step visually indicates the region where data to be collected, helping users understand the area of focus within each frame.

Post-extraction, the ROI undergoes a series of sophisticated image processing techniques aimed at enhancing its quality and suitability for data collection. These techniques may include grayscale conversion, Gaussian blurring, adaptive thresholding, and thresholding using Otsu's method. Each of these steps contributes to improving the clarity and distinctiveness of features within the ROI, ultimately leading to more accurate data collection and analysis.

Algorithm: Intelligent healthcare imaging system

1. Importing the necessary libraries such as torch, transformers, matplotlib.pyplot, PIL, transforms from torchvision, and make_grid from torchvision.utils.

2. Generating Captions:

a. Use torch.no_grad() to disable gradient calculation, as it is not needed during inference.

b. Call encoder_decoder.generate() with the following arguments:

i. pixel_values: Preprocessed images tensor.

ii. bos_token_id: Beginning of sequence token ID in the tokenizer.

iii. eos_token_id: End of sequence token ID in the tokenizer.

iv. pad_token_id: Padding token ID in the tokenizer.

v. return_dict_in_generate: Whether to return a dictionary containing additional information in the generate method.

vi. use_cache: Whether to use cache when decoding sequences.

vii. max_length: Maximum length of the generated sequences.

viii. num_beams: Number of beams for beam search decoding.

c. Decode the generated sequences using tokenizer.batch_decode() to convert token IDs into

Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(3), 3366-3374 | ISSN:2251-6727

text captions. Set skip_special_tokens=True to remove special tokens like <bos> and <eos>.
3. Displaying images with generated captions.

Fig. 2 Working Process for Intelligent Healthcare Imaging System

This study makes use an encoder-decoder architecture which acts by pre-trained on pictures and then refining it using data from a specific domain. Here the performance is enhanced by employing sentence-LSTM and word-LSTM together as a decoder. Then, the attention-based caption generation which computes semantic based attention across extracted information.

4. Results

In this experiments, various CNN architectures have been used, purposely Inception-v4, DenseNet121, ResNet152, and analyzed that the DenseNet121 architecture shown better outcomes. Hence, it is used DenseNet121 for all test cases. Images are inputted with the size 320×320 pixels. Also, it is inferred that the Adam optimizer subjected with the parameters α of α_1 as 0.9, α_2 as 0.999 and learning rate to $1 \times 10-4$ which is agreed for the training process. Finally, batches are tested applying a stable batch size as 16. It is trained with 3 epochs and checkpoints each 4800 repetitions.

Here, Matplotlib facilitates the visualization of medical images, aiding in the interpretation and understanding of diagnostic results. Figure 3 shows the input images of X-rays of three patients.

Fig. 3 Sample Input Images of Three Patients

The output of the X-ray images is displayed using the Matplotlib library, providing visual representation and analysis. The system displays generated captions derived from the input X-ray images, providing textual summaries of observed findings as shown in Figure 4. This enables clinicians to quickly comprehend and interpret the complex internal structures depicted in the images. Such captions aid in improving diagnostic efficiency and patient care by facilitating faster and more accurate analysis.

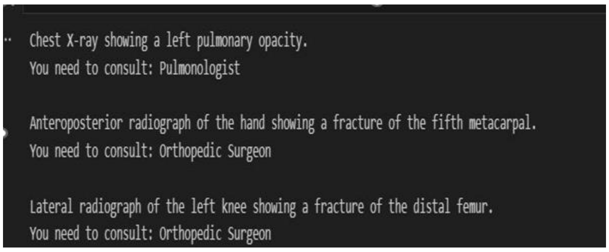

Fig. 4 Displaying the Generated Captions Output (for 3 input images)

Patient ID: 20PID404

Patient Name: John

Age: 25

Gender: Male

Date of Birth: 17-04-1999

Xray findings: Chest X-ray showing a left pulmonary opacity.

You need to consult: Pulmonologist

(a) Patient -1

Patient ID: 20PID407

Patient Name: Ram

Age: 22

Gender: Male

Date of Birth: 18-01-2002

Xray findings: Anteroposterior radiograph of the hand showing a fracture of the fifth metacarpal.

You need to consult: Orthopedic Surgeon

(b) Patient - 2

Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(3), 3366-3374 | ISSN:2251-6727

Patient ID: 20PID410

Patient Name: Kevin

Age: 19

Gender: Male

Date of Birth: 08-08-2004

Xray findings: Lateral radiograph of the left knee showing a fracture of the distal femur.

You need to consult: Orthopedic Surgeon

(c) Patient - 3

Fig. 5 Summary Report

Figure 5 shows the PDF report that includes details about the patient, findings from the X-ray, and advises medical staff to consult. This helps to ensure all necessary information provided for accurate diagnosis and treatment planning. The report acts as a comprehensive summary, making it easier for healthcare professionals to review and understand the patient's condition. By facilitating efficient communication between medical staff, it contributes to improved patient care and outcomes.

The BLEU score is presented as a quantitative measure of the similarity between the generated caption and the reference text, providing insight into the quality of the generated output. This evaluation metric helps assess the effectiveness of the caption generation model in accurately summarizing information from medical X-ray images, contributing to the overall reliability and utility of the system.

```
from nltk.translate.bleu_score import corpus_bleu

candidates=[[i] for i in medical_details]

bleu_score = corpus_bleu(candidates, references)

print("BLEU Score:", bleu_score)
[17]
... BLEU Score: 0.735773498736729
```

Fig. 6 Evaluation Metric – BLEU Score for Generated Caption

4.1 Evaluation Metrics

The BLEU score is calculated by combining precision and Brevity Penalty (BP) using Eq.1. Precision is obtained as the correspondence between the generated text and the reference text(s) using m-gram matches. BP adjusts the score to report for variations in length relating the generated and reference texts.

$$BLEU = BP \times \exp\left(\sum_{m=1}^{N} \omega_m \,.\log(p_m)\right) \qquad [1]$$

BP penalizes short translations compared to the reference. It is calculated as in Eq. 2.

$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ \exp\left(1 - \frac{r}{c}\right) & \text{if } c \leq r \end{cases} \qquad [2]$$

here c is the length of the generated text; r is the effective reference length, which is the closest reference length to c. This paper corpus_bleu() is utilized from the NLTK library to compute the BLEU score. This function takes two parameters: references, which is a list of reference texts, and candidates which is a list of generated texts. By applying the BLEU score using Eq.1, it is quantitatively assessing the performance of the text generation model as follows.

Reference captions:

[Chest X-ray showing a left pulmonary opacity,

Anteroposterior radiograph of the hand showing a fracture of the fifth metacarpal,

Lateral radiograph of the left knee showing a fracture of the distal femur].

The precision for unigrams, bigrams, trigrams, and so on, up to the desired m-gram size can be calculated as shown in Eq. 3. Precision for each m-gram is the proportion of the number of m-grams in the generated caption that appear in the reference captions to the total number of m-grams in the generated caption. Precision separately for each generated caption by using Eq. 3 is computed as given below.

$$\text{Precision} = \sqrt[3]{\text{Unigram Prec.} \times \text{Bigram Prec.} \times \text{Trigram Prec.}} \qquad [3]$$

(i) Precision Calculation:

For the given generated captions;

Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(3), 3366-3374 | ISSN:2251-6727

Caption 1: "Chest X-ray showing a left pulmonary opacity"

Caption 2: "Anteroposterior radiograph of the hand showing a fracture of the fifth metacarpal"

Caption 3: "Lateral radiograph of the left knee showing a fracture of the distal femur"

Unigram Precision: No. of unigrams in the generated caption that appear in the reference captions: 5 ("Chest", "X-ray", "showing", "a", "left")

Total no. of unigrams in the generated caption: 6

Unigram Precision = 5/6

Bigram Precision: No. of bigrams in the generated caption that appear in the reference captions: 4 ("Chest X-ray", "X-ray showing", "showing a", "a left")

Total no. of bigrams in the generated caption: 5

Bigram Precision = 4/5

Trigram Precision: No. of trigrams in the generated caption that appear in the reference captions: 3 ("Chest X-ray showing", "X-ray showing a", "showing a left")

Total no. of trigrams in the generated caption: 4

Trigram Precision = ¾

Calculate geometric mean of precision scores as shown in figure 7.

$$\text{Precision} = \sqrt[3]{\text{Unigram Precision} \times \text{Bigram Precision} \times \text{Trigram Precision}}$$

$$\text{Precision} = \sqrt[3]{(5/6) \times (4/5) \times (3/4)}$$

$$\text{Precision} = \sqrt[3]{0.4167 \times 0.8 \times 0.75}$$

$$\text{Precision} = \sqrt[3]{0.25}$$

$$\text{Precision} \approx 0.6299$$

Fig. 7 Evaluation Metric Calculation – Precision for the Generated Caption

(ii) Brevity Penalty Calculation: Calculate the brevity penalty for each generated caption based on its length compared to the reference captions.

Generated caption length:7 words

Closest reference caption length :6 words

Brevity penalty is calculated as shown in figure 8.

$$\text{Brevity Penalty} = \exp\left(1 - \tfrac{r}{c}\right)$$

$$\text{Brevity Penalty} = \exp\left(1 - \tfrac{6}{7}\right)$$

$$\text{Brevity Penalty} = \exp\left(\tfrac{-1}{7}\right)$$

$$\text{Brevity Penalty} \approx 0.8776$$

Fig. 8 Evaluation Metric Calculation – Brevity Penalty for the Generated Caption

Now, the BLEU score is evaluated as follows.

(iii) Final BLEU Score Calculation: Combine the precision and brevity penalty using the BLEU score formula to get the overall BLEU score. BLEU score is calculated as shown in figure 9.

$$\text{BLEU} = \text{BP} \times \exp\left(\sum_{n=1}^{N} w_n \cdot \log(p_n)\right)$$

$$\text{BLEU} \approx 0.8776 \times \exp(0.25 \times \log(0.6299))$$

$$\text{BLEU} \approx 0.8776 \times \exp(0.25 \times (-0.1543))$$

$$\text{BLEU} \approx 0.8776 \times \exp(-0.0386)$$

$$\text{BLEU} \approx 0.8776 \times 0.9628$$

$$\text{BLEU} \approx 0.8462$$

Fig. 9 Evaluation Metric Calculation – BLEU Score for the Generated Caption

4.2 Performance Comparison of the Intelligent Healthcare Imaging System

Table 1 shows the collection of BLEU score obtained by applying various techniques and performance comparison made against with other research papers.

Table 1. BLEU Score Comparison with the Existing Solutions

Methodology Used	BLEU: 1 Score	Limitations
LSTM with visual and medical concepts features [6]	0.42	Limited in generating long captions, model performance heavily dependent on multi-label classifier performance

Encoder-decoder model with CNN and LSTM, utilizing generative pretrained transformer (GPT-3) [2]	0.63	Challenges in integrating visual and textual data to generate medically accurate and contextually relevant captions.
Feature extraction by CNN and attention mechanism using attention mechanism for report generation [5]	0.72	Struggles with understanding diseases from CXR images and applying correct languages semantics to describe them
ALBEF language-image model pre-trained on MIMIC-CXR dataset and fine-tuned for image-text matching [4]	0.62	Overfitting to training data and difficulties in handling diverse clinical cases
Global-local visual extractor following cross encoder-decoder transformer [10]	0.72	Limited text generating from global features
Two LSTM layers with bottom-up and top-down attention mechanisms [11]	0.56	Managing the complexity of integrating two attention mechanism
Deep learning which includes encoder-decoder models [7]	0.74	Limited by the quality and diversity of training datasets, which can affect the generalizability and accuracy of generated reports
Proposed Model	0.73	Potential limitations which reliance on data quality and interpretability of generated captions

It is evident that the proposed technique shows better BLEU score than the existing solutions. Fig. 10 shows the BLEU score values comparison of different techniques applied.

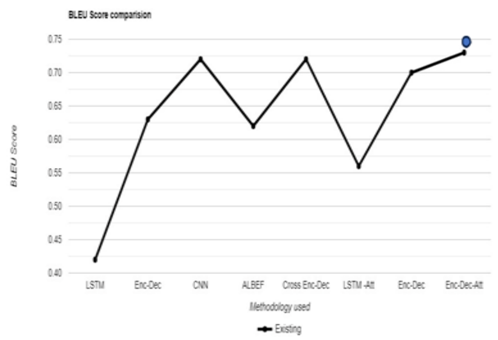

Fig. 10 BLEU Score Values Comparison for the Generated Caption

One of the major limitations in this radiograph interpretation models is that lack of datasets, only a few radiographic datasets are available as public.

5. Conclusion

This paper marks a significant milestone in the domain of medical image analysis, showcasing the transformative capacity of DL and AI in revolutionizing radiology interpretation processes. Through the integration of cutting-edge transformer-based models with sophisticated image processing techniques, the developed system has successfully demonstrated its capability to automate and enhance the generation of captions from medical X-ray images. This achievement underscores the power of AI-driven approaches in addressing longstanding challenges in medical imaging, particularly in the domain of radiology, where the interpretation of complex images is traditionally labour-intensive and prone to human error.

Furthermore, the fusion of transformer-based models with advanced image processing techniques amplifies the system's effectiveness in extracting pertinent information from medical X-ray images. Through meticulous preprocessing and feature extraction, the system enhances its ability to discern clinically relevant details and nuances within the images, thereby contributing to the generation of comprehensive and clinically meaningful captions. This holistic approach to medical image analysis not only enhances the efficiency of radiology workflows but also enhances the overall quality and accuracy of diagnostic reporting, laying the groundwork for future advancements in AI-driven healthcare technologies, hence, holds immense promise for improving patient care, driving innovation, and advancing the frontiers of medical science.

Declaration of competing interest

The authors declare that they have no competing personal relationships or financial interests and did not influence the interpretation of the results.

Acknowledgment

This work did not receive any funds.

References

1. Ayesha H., Tariq M., Israr S., 2023. Computer Aided Deep Image Captioning for Medical Images. Machines and Algorithms. 2(1).
2. Beddiar D.R., Oussalah M., Seppänen T., 2023. Retrieved Generative Captioning for Medical Images. CBMI '23: Proceedings of the 20th International Conference on Content-based Multimedia Indexing, Pages 48–54. DOI:10.1145/3617233.3617246.
3. Lee H., Cho H., Park J., Jihie Kim J.C., 2022. Cross Encoder-Decoder Transformer with Global-Local Visual Extractor for Medical Image Captioning. Sensors. 22(4), 1429. DOI:10.3390/s22041429
4. Selivanov A., Daniil Chesakov O.Y.R, Shelmanov A., Fedulova I., Dylov D.V., 2023. Medical image captioning via generative pretrained transformers. Scie. Reports. 13, 4171. DOI:10.1038/s41598-023-31223-5
5. Park H., Kim K., Yoon J., Park S., Choi J., 2020. Feature Difference Makes Sense: A medical image captioning model exploiting feature difference and tag information. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 95–102, Online. Association for Computational Linguistics. https://aclanthology.org/2020.acl-srw.14.pdf
6. Wang Z., Zhou L., Wang L., Li X., 2021. A Self-boosting Framework for Automated Radiographic Report Generation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA,pp. 2433-2442. DOI: 10.1109/CVPR46437.2021.00246.
7. Zhou W., Ye Z., Wang Y.Y.S, Huang H., Wang R., Yang D., 2023. Transferring Pre-Trained Large Language-Image Model for Medical Image Captioning. CLEF2023 Working Notes, CEUR Workshop Proceedings.
8. Gajbhiye G.O., Nandedkar A.V., Faye I., 2020. Automatic Report Generation for Chest X-Ray Images: A Multilevel Multi-attention Approach: In book: Computer Vision and Image Processing, Nain N., et al. (Eds.): CVIP 2019, CCIS 1147, pp. 174–182. DOI:10.1007/978-981-15-4015-8_15.
9. Gasimova A., Montana G., Rueckert D. 2017. Automated Knee X-ray Report Generation. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017. https://arxiv.org/pdf/2105.10702.pdf
10. Pang T., Li P., Zhao L., 2023. A survey on automatic generation of medical imaging reports based on deep learning. BioMed Eng. OnLine 22, 48. DOI:10.1186/s12938-023-01113-y
11. Pavlopoulos J., Kougia V., Androutsopoulos I., 2019. A Survey on Biomedical Image Captioning. In Proceedings of the Second Workshop on Shortcomings in Vision and Language, pages 26–36, Minneapolis, Minnesota. Association for Computational Linguistics.
12. Irvin J., Ko P.R.M., Yu Y., Ciurea-Ilcus S., Chute C., Marklund H., Robyn Ball B.H., Shpanskaya K., Seekins J., Mong Safwan S. Halabi D.A., Sandberg J.K., Jones R., Larson D.B., Langlotz C.P., Patel B.N., Lungren M.P., Ng A.Y., 2022. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), pp.590-597.
13. Beddiar D.R., Oussalah M., Seppänen T., 2023. Automatic captioning for medical imaging (MIC): a rapid review of literature. Artif Intell Rev. 56, 4019–4076. DOI:10.1007/s10462-022-10270-w
14. Aono M., Shinoda H., Asakawa T., Shimizu K., Togawa T., Komoda T., 2023. Multi-stage medical

image captioning using classification and CLIP. In CLEF2023 Working Notes, CEUR Workshop Proceedings, Thessaloniki, Greece.

15. Pavlopoulos J., Kougia V., Androutsopoulos I. Papamichail D., 2022. Diagnostic captioning: a survey. Knowl Inf Syst. 64, 1691–1722. DOI:10.1007/s10115-022-01684-7
16. Al-Malla M.A., Jafar A. Ghneim N., 2022. Image captioning model using attention and object features to mimic human image understanding. J Big Data. 9, 20. DOI:10.1186/s40537-022-00571-w
17. Srinivasan P., Thapar D., Bhavsar A., Nigam A., 2021. Hierarchical X-Ray Report Generation via Pathology Tags and Multi Head Attention. In: Ishikawa H., Liu C.L., Pajdla T., Shi J. (eds) Computer Vision – ACCV 2020. Lecture Notes in Computer Science, vol. 12626. Springer, Cham. DOI:10.1007/978-3-030-69541-5_36
18. Tiwari A.S.P.K, Raguru J.K., Prasad G., Chauhan S., Zaguia A., Mohammad Aman Ullah M.A., 2023. Medical Image Captioning Using Optimized Deep Learning Model. Comput Intell Neurosci. 9638438. DOI: 10.1155/2022/9638438
19. Aaron Nicolson A., Dowling J., Koopman B., 2021. AEHRC CSIRO at ImageCLEF med 2021Caption. https://ceur-ws.org/Vol-2936/paper-109.pdf