Exemplifying the Automated Decision-Making Process of Medical Image Annotation Systems through Explainable AI Tools*

Md. Rakibul Hasan, Md Mahmudur Rahman, *Member, IEEE*

***Abstract*—The field of medical imaging analysis has undergone a revolution thanks to the incorporation of automated methods for medical image annotation that are powered by artificial intelligence (AI). The lack of interpretability and transparency in these systems, however, raises questions about how they make decisions and restricts their use in therapeutic settings. Using explainable AI (XAI) tools to illustrate the decision-making procedure of medical picture annotation systems, we propose to address this difficulty in this study. We investigate numerous XAI methods, including Grad-CAM, LRP, and attention mechanisms, to shed light on the key elements and justifications for the system's annotations. We seek to improve the interpretability, reliability, and therapeutic utility of these automated systems by the visualization and explanation of the decision-making process.**

Keywords— ***Explainable AI (XAI), Medical Image Annotation, Artificial Intelligence, Caption Generation, Multi-labeling***

I. Introduction

Artificial intelligence (AI) integration in medical image analysis has transformed the area in recent years, enabling improved diagnosis, treatment planning, and academic study. But as AI algorithms grow more complicated and opaquer, questions about their interpretability and transparency have surfaced. The idea of Explainable AI (XAI), which aims to give consumers insight into the decision-making process of AI systems and promote trust and comprehension among users, has gained popularity in response to these worries.

The objective of this paper is to investigate and demonstrate the use of explainable AI techniques for automated medical image annotation, more specifically multi-labeling and caption generation tasks. Efficacious and precise processing and interpretation are of paramount importance given the exponential rise of digital medical imaging data. By utilizing XAI approaches, we may improve the interpretability of AI models and help medical practitioners comprehend the reasoning behind the automatically generated annotations and captions. This study will offer a thorough examination of the XAI methodology and approaches currently in use for biomedical image analysis. We will explore several XAI methodologies, LIME, SHAP, saliency mapping, attention mechanisms, and model-agnostic strategies, etc. This paper will also offer useful illustrations and case studies that show how XAI tools can be incorporated into the automated biomedical image annotation and caption production procedure. We will seek to enhance the clinical value, trustworthiness, and transparency of AI systems by the incorporation of visuals and explanations, ultimately resulting in better patient care and biological decision-making. Moreover, we are using a wide range of scientific articles, research papers, and other publications from reliable sources to support our conclusions and discussions.

To accomplish the objective of this research paper, we are utilizing the medical image dataset derived from the ImageCLEFmedical Caption Task from both 2022 [1] and 2023 [2]. The dataset from 2022 consisted of 83,275 training images and 7,645 validation images. Similarly, the dataset from 2023 comprised 60,918 training images and 10,437 validation images. It is worth noting that both datasets are subsets of the ROCO (Radiology Objects in Context) image dataset [3]. To meet the goals of this paper, we merged the images from these two consecutive years, resulting in a combined training set of 144,193 images, which is then split into an 80% training set and a 20% validation set. Additionally, we formed a combined validation set consisting of 18,082 images, which was used as the test dataset.

II. Methods and Approaches

Medical image annotation is the process of labeling medical images using text or unique identifiers or both to provide meaningful information for deep learning models both for training and prediction purposes [4]. Moreover, the annotation task involves the practice of automatically creating evocative concept identifiers or captions or textual summaries for medical images [5].

Numerous deep learning and AI architectures have been created and used in automated medical image multi-labeling and caption generation. Convolutional Neural Networks (CNNs) are a well-known design that are excellent at extracting pertinent characteristics from medical images using convolutional layers for image classification, object recognition, and segmentation tasks [6]. Recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks are RNNs that can capture temporal relationships and produce coherent and contextually pertinent captions for medical images [7]. Attention-based architectures can efficiently highlight important elements and enhance the overall performance of the multi-labeling and caption

*Resrach supported by NSF.

Md. Rakibul Hasan is graduate student of the Department of Computer Science, Morgan State University, Baltimore, MD 21251 USA (corresponding author to provide phone: 667-464-1855; e-mail: mdhas1@ morgan.edu).

Md Mahmudur Rahman is with the Computer Science Department, Morgan State University, Baltimore, MD, 21251 USA (phone: 443-885-1056; e-mail: md.rahman@ morgan.edu).

creation process by dynamically allocating attention to various image areas [8]. Additionally, Transformers (i.e., Vision Transformer) make use of self-attention processes to identify global and local dependencies in images, permitting accurate feature extraction and caption and multi-label generation [4]. Other Transformer architectures like GPT (Generative Pre-trained Transformer) and T5 (Text-To-Text Transfer Transformer) can get promising results in our intended tasks. Furthermore, ensemble models, which pool the results of various distinct models can improve performance and increase the robustness of the annotation and caption creation process by utilizing the collective intelligence of various models [9].

On the other hand, XAI tools SHAP (SHapley Additive exPlanations), LRP (Layer-wise Relevance Propagation), LIME (Local Interpretable Model-agnostic Explanations), Grad-CAM (Gradient-weighted Class Activation Mapping), Attention Visualization, Layer Activation Visualization, and Integrated Gradients will be employed to explain the outputs of the above mentioned models by using diverse approaches of feature relevance analysis, mapping, visualization, and weight improvement tracking techniques [10-13].

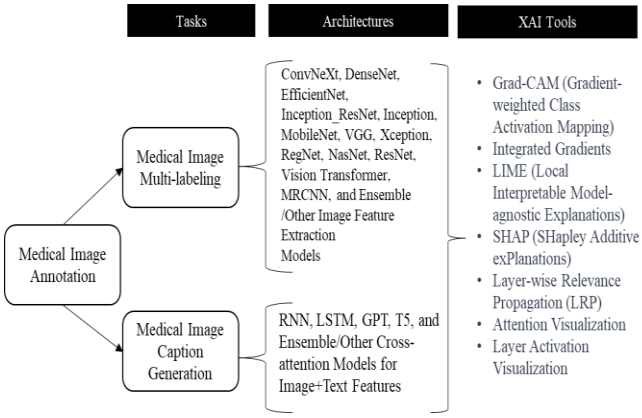

Figure 1. Holistic view of methodological approach.

Fig. 1 illustrates our generic approach to addressing the intended objective of our paper.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation (NSF) grant (ID. 2131307) under CISE-MSI program.

REFERENCES

[1] B. Ionescu, H. Müller, R. Peteri, et al., "Overview of the ImageCLEF 2022: Multimedia Retrieval in Medical, Social Media and Nature Applications," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction,* ser. Proceedings of the 13th International Conference of the CLEF Association (CLEF 2022), Bologna, Italy: LNCS Lecture Notes in Computer Science, Springer, Sep. 2022.

[2] B. Ionescu, H. Müller, A. Drăgulinescu, et al., "Overview of ImageCLEF 2023: Multimedia retrieval in medical, socialmedia and recommender systems applications," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction,* ser. Proceedings of the 14th International Conference of the CLEF Association (CLEF 2023), Thessaloniki, Greece: Springer Lecture Notes in Computer Science LNCS, Sep. 2023.

[3] O. Pelka, S. Koitka, J. Rückert, F. Nensa, and C. M. Friedrich, "Radiology objects in context (roco): A multimodal image dataset," in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis: 7th Joint International Workshop, CVII-STENT 2018 and Third International Workshop, LABELS 2018*, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3, Springer, 2018, pp. 180–189.

[4] M. Aljabri, M. AlAmir, M. AlGhamdi, M. Abdel-Mottaleb, and F. Collado-Mesa, "Towards a better understanding of annotation tools for medical imaging: A survey," *Multimedia tools and applications*, vol. 81, no. 18, pp. 25 877–25 911, 2022.

[5] J.-G. Lee, S. Jun, Y.-W. Cho, et al., "Deep learning in medical imaging: General overview," *Korean journal of radiology*, vol. 18, no. 4, pp. 570–584, 2017. C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[6] H. Ayesha, S. Iqbal, M. Tariq, et al., "Automatic medical image interpretation: State of the art and future directions," *Pattern Recognition*, vol. 114, p. 107 856, 2021.

[7] T. Pang, P. Li, and L. Zhao, "A survey on automatic generation of medical imaging reports based on deep learning," *BioMedical Engineering OnLine*, vol. 22, no. 1, pp. 1–16, 2023.

[8] M. M. A. Monshi, J. Poon, and V. Chung, "Deep learning in generating radiology reports: A survey," *Artificial Intelligence in Medicine*, vol. 106, p. 101 878, 2020.

[9] B. Jing, P. Xie, and E. Xing, "On the automatic generation of medical imaging reports," arXiv preprint arXiv:1711.08195, 2017.

[10] R.-K. Sheu and M. S. Pardeshi, "A survey on medical explainable ai (xai): Recent progress, explainability approach, human interaction and scoring system," *Sensors*, vol. 22, no. 20, p. 8068, 2022.

[11] S. Masıs, Interpretable machine learning with Python: Learn to build interpretable high-performance models with hands-on real-world examples. Packt Publishing Ltd, 2021.

[12] S. Barocas, M. Hardt, and A. Narayanan, "Fairness in machine learning," *Nips tutorial*, vol. 1, p. 2017, 2017.

[13] D. Rothman, *Hands-On Explainable AI (XAI) with Python: Interpret, visualize, explain, and integrate reliable AI for fair, secure, and trustworthy AI apps*. Packt Publishing Ltd, 2020