Quest Journals
Journal of Research in Applied Mathematics
Volume 9 ~ Issue 10 (2023) pp: 16-20
ISSN (Online): 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Review Paper

Fusion of Medical Imaging and Natural Language Processing Using Deep Learning

Lin Ma
Worcester Academy, Worcester, Massachusetts, USA

ABSTRACT: *Deep learning is having a profound impact on the medical field. Among them, medical imaging and natural language processing have been widely used for diagnosis and treatment decision making in recent years. This paper reviews the current research focus and some achievements of medical imaging, natural language processing and feature fusion technology. Since the feature fusion technology of deep learning can combine different types of information input with improving the performance of the system, after analyzing the feasibility, a hypothetical architecture combining medical images and natural language processing is proposed. Actual training and testing of the system are possible in the future.*

KEYWORDS: *Deep Learning, Multi-model Fusion, Medical Imaging, Natural Language Processing*

***R**eceived 20 Sep., 2023; **R**evised 02 Oct., 2023; **A**ccepted 04 Oct., 2023 © The author(s) 2023.*
P**ublished with open access at **www.questjournals.org

I. INTRODUCTION

According to the World Health Organization, Cancer is the main cause of death in the global population with nearly 10 million deaths in 2020. If cancer is detected early, patients will bear lower treatment costs and have a greater probability of survival. [1] Indeed, if patients can be diagnosed earlier, it is of positive significance both for the health of patients and for the saving of medical resources in society. In recent years, the medical community has tried to investigate techniques to help automatically diagnose tumors and predict the development of the disease. Deep learning architectures based on medical imaging and natural language processing are explored widely to settle the problem. Both approaches show great potential, but whether the two methods can be integrated into a system to improve the overall accuracy is an open question. Therefore, this paper mainly discusses the current trends of deep learning in medical imaging and natural language processing, as well as the feasibility of integrating the two systems, and proposes a possible framework.

II. BACKGROUND

Deep learning has been widely studied in recent years for its application in multiple fields. In short, deep learning is a learning system that simulates neural processing under machine learning. By setting multiple hidden layers, the weight and biases of hierarchical links are corrected in training so as to achieve automatic data processing and judgment. With the update of hardware devices, more computing power enables deep learning to run deeper architectures faster. Innovations in algorithms and practices have also led to the processing of larger data sets to train more accurate systems. Specifically, scholars are also gradually exploring the possibility of deep learning in medication. It has been applied in medical research, disease prevention, and auxiliary diagnosis. [2]

Computer vision is a research field in which computers perform specific tasks by extracting and analyzing pixel information from videos or images. This technology is widely studied in industrial and agricultural production, navigation, medicine and other fields. In medical imaging, deep learning has been used in recent decades to assist physicians in disease detection and diagnosis. Computer-aided detection/diagnosis (CAD) algorithms in the early days highly depended on feature observation. The system relied on the patterns discoveries by clinicians. That is, the input of the system is certain traits of the image, and the detection results are calculated by comparing the features of the disease (such as edges, shapes, lines, etc) that have been established. Although less training is required in such a way, there were limitations to the system that the algorithms could only detect the known feature of the image rather than recognize patterns that have not yet been identified, which is the ultimate purpose of machine learning. As a result, architecture such as convolutional neural networks (CNN) and massively trained artificial neural networks (MTANNs) were founded. These systems identify features by analyzing the pixel values of images directly. Although a lot of training data is required, the final model is generally more accurate. [2]

Natural Language Processing (NLP) is designed to enable computers to understand human language. It consists of functions of comprehension and generation of natural languages. Natural Language Understanding (NLU) extracts concepts and emotions from the text by parsing phonology, lexical, semantics, etc. Natural Language Generation (NLG) is where the computer identifies the target and generates paragraphs based on the understanding of certain languages. Both NLU and NLG have advanced rapidly in the recent decade due to the introduction of neural networks. In the medical field, particularly when applied to medical imaging, NLP can be used to analyze the radiology report to clarify the abnormal region of photographs. By extracting keywords and segmentation paragraphs, NLP can extract important information from unstructured texts for doctors' reference to reduce the misunderstanding of complex results. [3]

III. ARGUMENT

There have been many attempts to diagnose and predict tumors using deep learning in Medical Imaging and Natural Language Processing. These attempts indicate that the trained systems can reach certain accuracy on the prediction and can be improved with more training data or better architecture. In medical imaging, the convolution neural network (CNN) has been widely studied, and it has demonstrated remarkable image classification and segmentation since CNN demands less preprocessing compared to other models. [4] For instance, the study conducted by Amjad Rehman, Muhammad Attique Khan, etc. established an architecture using a 3D convolutional neural network to detect and classify the brain tumor. The system operated the BraTS datasets in 2015, 2017, and 2018, and reached an accuracy of 98.32, 96.97, and 92.67%, respectively. [5] Such a high accuracy represents that medical imaging systems based on CNN and other architectures can gradually become a reference for tumor detection and prediction.

Simultaneously, Natural Language Processing has been tested to contribute to the analysis of radiology reports and the detection of tumors. The unstructured radiology reports consist of detailed descriptions of the radiation images from the physicians. The idea of conducting a massive illustration is to notify all the information that the scanning might imply, avoiding missing any corner elements. As a result, There would be certain problems such as high cost, low efficiency of information transmission, and the information cannot be uniformly classified. [6] Although structured radiology reports with clear standards and formats are increasingly being used, the use of deep learning to examine massive data in unstructured radiology reports has been increasingly called up attention by researchers. For example, the research of Thomasian et al. trained multiple constructions of deep learning, including convolution neural network (CNN), recurrent neural network (RNN), etc, and tested their accuracy in identifying the severity of the tumors by intaking 10006 free-text radiology report from 1580 patients. (DATA) The result shows that the estimation of survival and hazard ratios of the system is similar to those of the physicians. [7] Thus, it can reasonable to infer that the model of deep learning is capable of analyzing the radiology report and giving results that are similar to manual judgment. In addition to the studies that have been introduced previously, many applications (Table 1) of deep-learning-based computer vision and natural language processing in oncology analysis have demonstrated the feasibility of these two methods for tumor identification and prediction. This, then, leads to the question of whether the two methods can be combined into a construction that further increase the accuracy of the system.

In order to integrate the Medical Imaging and Natural Language Processing and get more accurate results, one needs to consider different fusion strategies, their advantages, disadvantages, and feasibility. Currently, the widely used fusion strategies are early fusion, joint fusion, and late fusion. Early fusion, also known as feature level fusion, means that different data sets are integrated into one vector before importing them into a single deep learning system. Joint fusion, also known as intermediate fusion refers to the integration of other data or features in the middle layers of a neural network. Late fusion, or decision-level fusion, integrates and weights the outputs of different models to get the final result. [8]

Multiple applications of the fusion of neural networks have proven that compared to the modal that uses single data sets or sources, the construction of the multi-model fusion generally has better performance. For example, in the study conducted by D.Haritha and B.Sandhya, they trained and tested different models of deep learning to detect skin cancer. They compared the accuracy of the system of medical imaging and patient data fusion to that of models using every single feature. The result shows that the system of fusion generally operates better than the single model construction. In one of their measure, the fusion system has an accuracy of 68.96%

compared to that solely uses images or clinical data features, which is 66% and 61%, respectively. [9] Additionally, more studies on fusion models have proved that combining data sets with different features can improve the accuracy of the overall system. (Table 2)

It can be reasonably indicated that the fusion model of medical imaging and NLP will have higher accuracy than using a single feature. Therefore, a possible construction may be established. The whole architecture combines medical images, such as CT and MRI scans, and non-structured medical texts, such as radiology reports and patient records, into a set as input. The output is the system's judgment to predict the severity of the disease or the calculation of possible treatment methods and survival rate based on previous data. Since only two features are combined, the system can use either an early or intermediate fusion (the two approaches are similar in their logic). Firstly, for image feature extraction, the lesions in the image can be identified by using the trained model, and the diseased parts can be either segmented into parts or directly brought into the pixel value as the input value of the fusion (marked as a1 to an). On the other hand, trained NLP can be used to extract and classify textual descriptions from radiological reports and other diagnostic texts, including the physician's description of image features and other patient symptoms and signs that cannot be reflected on images, and then bring them into the input (marked as an to am) along with the previously mentioned image features. Finally, the combination of all the data is brought into the deep learning system to calculate the final judgment of the severity of the disease and the possible treatment and survival rate. (Figure 1) Such a combination of medical images and medical texts enables the system consider not only pixel information but more clinical data, which can enhance the accuracy of the overall judgment. At the same time, the analysis of unstructured medical texts can also improve the efficiency and accuracy of information transmission, and more information extracted from the texts will further improve the range of data the system input.

IV. FIGURES AND TABLES

Author	Purpose	Data Input	Architecture	Performance	Reference
Alkadi et al.	propose a novel 3D sliding window approach that perform detection, localization, and segmentation of prostate cancer	T2 MRI, I2CVB	Deep convolutional encoder-decoder	AUC 0.995, accuracy 0.894, recall 0.928.	[11][10]
Ha et al.	prediction of neoadjuvant chemotherapy (NAC) response prior to initiation of chemotherapy	Breast MRI	CNN	Overall accuracy 88%	[12][10]
Ranjbarzadeh et al.	localization and segmentation of brain tumor from MRI	Brain MRI	Cascade Convolutional Neural Network ((C-ConvNet/C-CNN)	whole tumor accuracy 0.9203	[13]
Jacobs et al.	Lung cancer detection based on low-dose CT scans	low-dose CT scans	grt123, Julian de Wit and Daniel Hammack (JWDH), and Aidence.	AUC for grt123 :0.877 (95% CI: 0.842, 0.910), JWDH: 0.902 (95% CI: 0.871, 0.932), Aidence : 0.900 (95% CI: 0.870, 0.928)	[14]
Wang et al.	Prediction of advanced colorectal cancer	EHR patient data	CNN	Sensitivity: 0.837, specificity: 0.867, PPV value: 0.532	[15][10]
Yoon et al.	Extract information from cancer pathology reports by apply Bayesian optimization	cancer pathology reports	CNN	F1-score: 0.915 & 0.902	[16][10]
Lee et al.	Fracture diagnosis by analyzing EMR text	EMR report	RNN	Accuracy: 0.982, Precision: 0.967, Recall: 0.967, F1 score: 0.967	[17]
D'Anniballe et al.	Classification of diseases based on CT report of chest, abdomen, and pelvis	CT report	RNN	AUC: > 0.95, Accuracy: 91–99%	[18]

Table 1: Applications of medical imaging and natural language processing using deep learning

Table 2: Applications of Multi-modal fusion deep learning

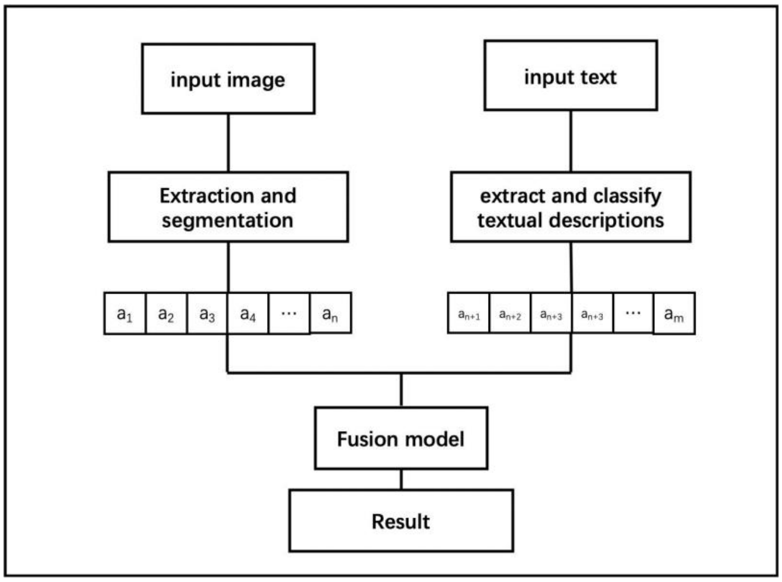

Figure 2: Possible construction of fusion model

V. DISCUSSION

If the architecture works properly and deep learning based medical imaging and natural language processing can be effectively combined, the accuracy of tumor diagnosis and prediction can be further increased. There have been some previous studies on medical image fusion such as the research by Huang et al. In the study, the researchers combine CT scans and patient information on the Electronic Medical Record (EMR) to diagnose Pulmonary Embolism. By testing different fusion strategies, the team created several architectures that made well performance. [23] The model presented in this paper does not coincide exactly with the model studied by the team. The research by Huang et al. primarily work on structured data presented in the EMR such as vitals, inpatient and outpatient medications, and so on; whereas, this article discusses the use of NLP to analyze unstructured medical texts such as radiology report.

Furthermore, if the construction described in this paper can be implemented, on the one hand, the fusion of medical imaging and natural language processing can be further studied, and at the same time, the workload of clinical records of medical workers can be reduced to a certain extent, because they can no longer be limited to filling in tables and reports with standard formats.

VI. CONCLUSION

By discussing the current progress of medical imaging and natural language processing and part of the research carried out, after analyzing the feasibility of combining the two using deep learning and other feature fusion cases, a possible architecture is proposed. If this architecture is confirmed, it will open up more possibilities for convergence models in deep learning in the medical field and will improve tumor diagnosis and prediction to benefit more patients.

REFERENCES

[1] Cancer. World Health Organization. https://www.who.int/en/news-room/fact-sheets/detail/cancer#:~:text=The%20problem.%20Cancer%20is%20a%20leading%20cause%20of,million%20cases%29%20Skin%20cancer%20%28non-melanoma%29%20%281.04%20million%20cases%29

[2] Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML. Deep learning in medical imaging and radiation therapy. Med Phys. 2019 Jan;46(1):e1-e36. doi: 10.1002/mp.13264. Epub 2018 Nov 20. PMID: 30367497.

[3] Khurana, D., Koli, A., Khatter, K. et al. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl (2022). https://doi.org/10.1007/s11042-022-13428-4

[4] Bhatt D, Patel C, Talsania H, Patel J, Vaghela R, Pandya S, Modi K, Ghayvat H. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. Electronics. 2021; 10(20):2470. https://doi.org/10.3390/electronics1020247

[5] Rehman, A, Khan, MA, Saba, T, Mehmood, Z, Tariq, U, Ayesha, N. Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture. Microsc Res Tech. 2021; 84: 133– 149. https://doi.org/10.1002/jemt.23597

[6] Sinha R. Structured vs Unstructured Radiology Reports. collaborative imaging. https://collaborativeimaging.com/2020/01/27/structured-vs-unstructured-radiology-reports/

[7] Thomasian, Nicole M., Imami, Maliha R. et al. Deep Learning for Natural Language Processing of Neuro-Oncology Imaging Reports. Available at SSRN: https://ssrn.com/abstract=4167846 or http://dx.doi.org/10.2139/ssrn.4167846

[8] Huang, SC., Pareek, A., Seyyedi, S. et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. npj Digit. Med. 3, 136 (2020). https://doi.org/10.1038/s41746-020-00341-z

[9] D. Haritha and B. Sandhya, "Multi-modal Medical Data Fusion using Deep Learning," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), 2022, pp. 500-505, doi: 10.23919/INDIACom54597.2022.9763296.

[10] Babita Pandey, Devendra Kumar Pandey, Brijendra Pratap Mishra, Wasiur Rhmann, A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions, Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 8, Part A,2022,Pages 5083-5099,ISSN 1319-1578,https://doi.org/10.1016/j.jksuci.2021.01.007.

[11] Alkadi, R., Taher, F., El-baz, A. et al. A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images. J Digit Imaging 32, 793–807 (2019). https://doi.org/10.1007/s10278-018-0160-1

[12] Ha, R., Chin, C., Karcich, J. et al. Prior to Initiation of Chemotherapy, Can We Predict Breast Tumor Response? Deep Learning Convolutional Neural Networks Approach Using a Breast MRI Tumor Dataset. J Digit Imaging 32, 693–701 (2019). https://doi.org/10.1007/s10278-018-0144-1

[13] Ranjbarzadeh, R., Bagherian Kasgari, A., Jafarzadeh Ghoushchi, S. et al. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. Sci Rep 11, 10930 (2021). https://doi.org/10.1038/s41598-021-90428-8

[14] Jacobs C. et al. Deep Learning for Lung Cancer Detection on Screening CT Scans: Results of a Large-Scale Public Competition and an Observer Study with 11 Radiologists. Radiology: Artificial Intelligence 2021 3:6. https://pubs.rsna.org/doi/abs/10.1148/ryai.2021210027

[15] Y.-H. Wang, P.-A. Nguyen, M.M. Islam, Y.-C. Li, H.-C. Yang. Development of deep learning algorithm for detection of colorectal cancer in EHR data. Stud. Health Technol. Inform., 264 (2019), pp. 438-441, 10.3233/SHTI190259

[16] H. -J. Yoon, J. Gounley, S. Gao, M. Alawad, A. Ramanathan and G. Tourassi, "Model-based Hyperparameter Optimization of Convolutional Neural Networks for Information Extraction from Cancer Pathology Reports on HPC," 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2019, pp. 1-4, doi: 10.1109/BHI.2019.8834674.

[17] Lee C., Kim Y. et al. Automatic Disease Annotation From Radiology Reports Using Artificial Intelligence Implemented by a Recurrent Neural Network. American Journal of Roentgenology 2019 212:4, 734-740. https://www.ajronline.org/doi/10.2214/AJR.18.19869

[18] D'Anniballe, V.M., Tushar, F.I., Faryna, K. et al. Multi-label annotation of text reports from computed tomography of the chest, abdomen, and pelvis using deep learning. BMC Med Inform Decis Mak 22, 102 (2022). https://doi.org/10.1186/s12911-022-01843-4

[19] Yap, J, Yolland, W, Tschandl, P. Multimodal skin lesion classification using deep learning. Exp Dermatol. 2018; 27: 1261– 1267. https://doi.org/10.1111/exd.13777

[20] Nie, D., Lu, J., Zhang, H. et al. Multi-Channel 3D Deep Feature Learning for Survival Time Prediction of Brain Tumor Patients Using Multi-Modal Neuroimages. Sci Rep 9, 1103 (2019). https://doi.org/10.1038/s41598-018-37387-9

[21] Ghulam Muhammad, M. Shamim Hossain, COVID-19 and Non-COVID-19 Classification using Multi-layers Fusion From Lung Ultrasound Images, Information Fusion, Volume 72, 2021, Pages 80-88, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2021.02.013.

[22] Zhuang Z. et al. Breast ultrasound tumor image classification using image decomposition and fusion based on adaptive multi-model spatial feature fusion, Computer Methods and Programs in Biomedicine, Volume 208, 2021, 106221, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2021.106221.

[23] Huang, SC., Pareek, A., Zamanian, R. et al. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. Sci Rep 10, 22147 (2020). https://doi.org/10.1038/s41598-020-78888-w