International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning-Driven Real-Time Multimodal Healthcare Data Synthesis

M. Preetha[1], Raja Rao Budaraju[2], Jackulin. C, P. S. G. Aruna Sri[4], T. Padmapriya[5]

Submitted: 14/09/2023 **Revised:** 21/10/2023 **Accepted:** 13/11/2023

Abstract*:* In recent years, the healthcare sector has witnessed an exponential surge in data generation from various sources. This data influx has opened new avenues for researchers to construct models and analytics, enhancing patient healthcare. While research and applications in prediction and classification have prospered, numerous challenges persist in optimizing healthcare comprehensively. Challenges encompass improving physician performance, curbing healthcare costs, and uncovering novel disease treatments. Physicians often grapple with time-consuming tasks, resulting in fatigue and occasional misdiagnoses. Automating such tasks can save time, enabling healthcare professionals to concentrate on elevating care quality. Health datasets comprise multiple modalities, such as structured sequences, unstructured text, images, ECG, and EEG signals. Leveraging these diverse data types necessitates effective methods. Moreover, the healthcare landscape is hindered by limited treatment options reaching the market, with many potential solutions failing in clinical trials. Machine learning models can enhance clinical trial outcomes and consequently elevate patient treatment quality. This paper addresses these issues through the development of a multimodal deep learning framework. It generates text reports and aids physicians in clinical practice, offering a multifaceted approach to address the diverse challenges in healthcare. The intended objective is to create a generative model capable of generating chest X-ray images and their corresponding textual reports.

Keywords: *Deep learning, healthcare, clinical trials, X-ray images*

1. Introduction

In the healthcare sector, a number of intricate data sources that aid in decision-making include embedded gadgets, detectors, mobile apps, medical information, and online platforms. The details gathered by healthcare equipment help significantly in the early identification of disorders and the management of suitable therapies. Whenever we talk about "big data," we normally mean a wide range of smoothly sized types of data from heterogeneous sources that have been stacked up on memory units [1]. These types of information can be expressed in both petabytes and zeta-bytes. In the last few years, deep learning models have been increasing prominence and have been discovered effective in multiple fields such as object detection, speech recognition, and image processing. The DL frameworks have become commonly utilized in recognition of their inbuilt features. The attribute building can be accomplished by the DL model automatically. Unlike specifically stating so, the DL model finds features in the data that match, integrates them, and speeds up learning. The need for AI-powered development in the medical field is rising as a result of recent advancements in AI. The creation of data-driven models is a crucial first step in supporting behavioral interventions for prevention or rehabilitation, such as identifying and tracking risk variables related to chronic illnesses. Nevertheless, the creation of data-driven models usually requires access to medical information from multiple sources. Personal Health Records (PHRs) from individual smartwatches and cell phones are among these sources as well as Electronic Medical Records (EMRs).

The development of AI technology and data-driven models in the healthcare sector is hampered by three main issues. First, the General Data Protection Regulation (GDPR) has made public

[1]*Professor, Prince Shri Venkateshwara Padmavathy Engineering College,*
Email: preetha.m.cse@psvpec.in, smpreetha14@gmail.com
Orcid: 0000-0001-8483-9871
[2]*Senior Member of Technical staff, Oracle, 3990 Scottfield street Dublin 94568 CA USA*
Email: rajaraob@yahoo.com
[3]*Assistant Professor, Department of CSE, Panimalar Engineering College*
Email: chin.jackulin@gmail.com
Orcid: 0000-0001-7112-1553
[4]*Professor, Department of ECM, Koneru Lakshmaiah Education Foundation, Vaddeswaram*
Email: arunasri_2012@kluniversty.in
[5]*Melange Publications, Puducherry, India*
Email: padmapriyaa85@pec.edu

the personalized data sharing of people (data subjects) as a component of larger data samples. Time-consuming regulatory processes like ethical authorizations, privacy evaluations, and anonymization methods regulate this kind of data exchange, which can stifle creativity. The second problem is finding human volunteers who are available to represent the desired demographic in future study [2]. The third barrier is collecting data from these human participants, which is labor-intensive and vulnerable to interruptions from unforeseen events such as pandemics, which may result in a postponement or cancellation of the study.

These three obstacles are lessened when artificial intelligence models are developed through the application of fake statistics. An alternate method of expressing the fundamental properties of real data about real human individuals, like dependencies and distributions, is the use of synthetic data. Furthermore, to protect patient privacy and lessen the possibility of re-identification, synthetic data is also used. The process of designing and developing artificial intelligence models for collecting data in medical fields setting (e.g., vital signs or electronic health records) was the subject of previous works. But with the introduction of biosensors and contemporary assistive technologies, such smartwatches, came new difficulties concerning the unique properties of the signals produced by this new category of gadgets. First of all, the signals that come from biosensors and wearables are multimodal. This indicates that variables are not routinely sampled and that human activities—such as monitoring exercise during the day and sleep during the night—are the only means of collecting them. Additionally, there is an erratic schedule for the collection of physiological indicators like blood pressure and heart rate. Multimodal data presents a very strong argument against current methods designed for traditional time-based electronic health records that are collected regularly because of these features.

Researchers have now brought out an extensive range of network flow categorization approaches. These technologies collapse into four primary groups: approaches based on ports, ML, deep packet inspection (DPI), and DL, which are presented specifically for a better understanding as they vary from standard machine learning techniques). On the one hand, classical network categorization methods like port recognition are no longer compatible with present-day network segmentation due to the progress of the network protocol itself [3]. This article delivers an outline of the different machine learning strategies and investigates the fundamental deep learning methods to recommend a multimodal framework that enhances the model's stability while concurrently improving its precision in classification [4].

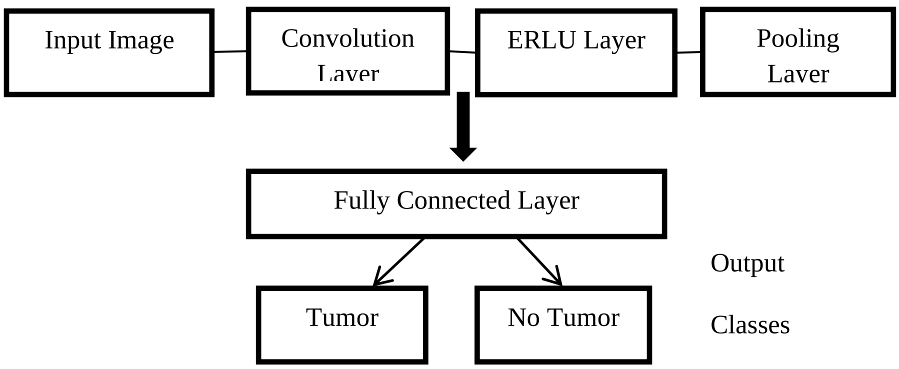

Fig. 1. Deep learning model usage for analytical medicine

A particular branch of machine learning called "deep learning" attempts to comprehend abstract information at a high level through the introduction of multiple processing layers with complex architectures or repeated nonlinear transformations. For cognitive schooling, it develops neural networks that may replicate the framework of the human brain. It is also frequently used in the context of clinical medical photographs, stimulating the brain's systems for analyzing data, which comprises texts, noises, and graphics. In the final analysis, there is a great deal of room for progress in the field of deep learning technology and medical image processing, which has turned into an important study issue currently. Figure 1 will indicate the deep learning model application in the computational field for clear details.

Clinicians routinely collect data from a variety of sources, such as radiographic scans, histopathological studies, test results, body vital signs, and other clinical data, to obtain a deeper understanding of their patient's health and deliver customized medical care. Medicine is inherently multimodal because clinical decision-making depends on a variety of data sources. The term "data modality" refers to the type of data used, such as an X-ray, a histopathology image stained with hematoxylin and eosin (H&E), and patient demographic data. Due to the unique nature of each modality's data production, recording, or collection procedure, each modality in such data sets may have a different quality and scale. The data modalities may comprise the following: (i) -omics data from the genome, proteome, transcriptome, epigenome, and microbiome; (ii) radiological images from ultrasound machines, magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), immunohistochemistry, and immunofluorescence; (iii) digitized histopathology, immunohistochemistry, and immunofluorescence slides made using tissue samples and stored as gig pixel whole slide images (WSI); and (iv) electronic health record (EHR) that contains unstructured data like discharge notes or medical reports alongside structured data like demographics.

A unified, deeper picture of cancer that may be more comprehensive and informative than the sum of the parts can be produced by integrating data from several diverse modalities. Multimodal medical data has enormous potential to improve our

knowledge of intricate illnesses and support the creation of specialized, efficient therapies. The significance of gathering, organizing, and harmonizing multimodal data in cancer care is further highlighted by the recent development of machine learning algorithms that can learn from multimodal data [5]. An explosion of heterogeneous, multimodal data has resulted from the introduction of high-throughput multi-omics technologies such as next-generation sequencing (NGS), high-resolution imaging in radiology and histopathology, and the quick digitalization of medical records. The abundance of training data in machine learning has directly led to major advancements; hence this data flood has been beneficial. It is hoped that large-scale, representative, standardized multimodal datasets in the medical field would serve as a rich environment for the creation of cutting-edge translational machine-learning models. Large-scale, high-quality datasets are ideal for machine learning, but gathering such resources in the healthcare industry presents special difficulties. First, multimodal medical data, which includes both structured and unstructured data, is by nature noisy and heterogeneous. Such disparate data must be aggregated using a lot of manual processing and harmonization. Secondly, precision, robustness, and dependability are essential for any medical application. However, it can be difficult to create robust and trustworthy models since real-world clinical data is frequently scant, inaccurate, and missing.

2. Literature Review

As deep learning networks (DNNs) outperform human analysts in image categorization, they are highly valuable in the medical domain [6]. To classify images into beneficial and harmful situations based on CXR intensity analysis, Hemdan et al. [7] proposed a deep learning classification system that made use of seven distinct CNN models. Apostolopoulos et al. [8] developed a highly accurate method for identifying lung illnesses using a deep learning-based mobile net system with 3905 X-ray images. Pre-trained approaches were implemented to classify COVID-19 cases; however, this pre-trained algorithm resulted in an imbalanced dataset impact on training.

Image classification, object recognition, categorization, authorization, and other deep learning techniques have endured, according to Litjens et al.'s [9] analysis and study of numerous academic articles on deep learning techniques. The basis of multimodal data fusion in neuroimaging has been given by Zhang et al. [10]. They also described the benefits and drawbacks of various imaging modalities, as well as basic fusion guidelines, techniques for evaluating fusion quality, and present multimodal fusion difficulties. In addition, they provided an overview of recent advancements and uses of multimodal neuroimaging for neurological conditions and brain illnesses. The authors Bellemo et al. [11] presented a GAN-based classifier for creating images of the retinal fund which may be exploited with artificial datasets.

Utilizing data from the original field to support the target learner in achieving enhanced efficiency is the objective of transfer learning. Transfer learning may be accomplished more effectively the closer the relationship is between the source and destination domains. If not, it could be more challenging and might have unfavorable consequences. After demonstrating exceptional generalization efficiency in the source field, the previously trained system is further optimized by small-scale data from the target area [12]. Several pre-trained deep learning algorithms are being used for COVID-19 diagnosis because it is costly to obtain CT or X-ray pictures of COVID-19 patients. A unique domain transfer learning approach called feature fusion, deconstruct, and transfer (FFDT) is proposed by Kabe et al. [13] for the classification of COVID-19 cases. To improve features, their suggested FFDT combines distant features from distant domains into only one attribute set where the distribution mismatch is minimized. To extract features, they also employ modified convolutional neural networks (MCNNs), and class reconstruction is applied to reveal the local structure of the data distribution. Their methodology obtains 94.5 percent classification accuracy. Since training the CNN algorithm from scratch is rather complex, characteristics from chest CT scans may be extracted using transfer learning. The foundation for extracting features from the CT images is made up of the pre-trained ResNet-18 and ResNet-50 algorithms. Discriminant correlation evaluation is used for integrating the obtained characteristics into improved image characteristics. Finally, several randomized neural networks are developed with enhanced features, and their forecasts are merged to provide more dependable classification accuracy.

A summary of current trends was provided by this article [14], which conducted an in-depth analysis of the literature on MML and its problems. The PRISMA technique was used to perform the study, and a detailed analysis of its selection process was provided. From a total of 1032 researched documents, 350 documents were ultimately chosen based on how well they addressed each study question that was formulated. The results of this review show that MML focuses on ML objectives and methods in addition to senses. The results of the research indicate that the majority of commonly used data sets are images and methods on the basis of neural networks. The various facets of multimodal representation, interpreting, symmetry, integration, and co-learning are also highlighted in the above article, along with any potential weaknesses. The various facets of multimodal participation, transcription, cooperation, fusion, and collaborative learning are also highlighted in this paper, along with any potential weaknesses. On the other hand, this SLR presented an overview of the MML research that has to be further developed to give academics interested in this multidisciplinary subject a current option for future work.

3. Methodology

Many current approaches aim to produce complete reports directly from the raw input but face several challenges. These challenges include the presence of errors in the generated reports, which require manual review and correction. Additionally, this method does not streamline the report-writing process when doctors wish to include supplementary information. Furthermore, the generated reports lack customization to align with the individual preferences of each doctor. In this paper, a deep learning-driven approach for clinical report auto-completion is introduced. This method operates interactively, constructing reports one sentence at a time. It leverages doctors' anchor words and partially completed sentences to generate each sentence of the report. The system, known as CRAC, seeks out the most pertinent sentences from existing reports to serve as templates for the current report. These recovered sentences undergo sequential adjustments, combining them with input feature representations to produce the ultimate report.

Diagnostic imaging and neural files, like X-rays and EEG scans, are essential for diagnosis and therapy in clinical settings. To detect important disease traits, clinical specialists have historically actively examined these images and signals. They then painstakingly craft textual reports to describe any anomalies and provide detailed explanations of their observations. However, the current process of composing clinical reports is arduous and time-consuming. Furthermore, it demands a deep understanding of image and signal patterns, as well as extensive experience in correlating these patterns with specific medical conditions.

Enhancing the caliber and efficiency of medical report composition can significantly influence telemedicine and web-based healthcare. In addressing the limitations of manual report writing, numerous methods for generating medical image reports have been put forward [15]. However, neither of the current methods concurrently achieves the further necessary requirements for the creation of medical reports:

Conformance with Disease Phenotypes: Test results and diagnostics obtained from medical pictures or brain activities are communicated through medical reports, which are essential. These reports must be in line with certain illness characteristics and relate to correct terminology for medical usage.

Adaptive Report Generation: The generated reports must possess the capability to adapt to the preferences of end-users, particularly clinicians, to enhance their acceptance and usability.

In response to this need, the paper introduces an interactive approach, referred to as DLCRAC, which aids in composing medical reports on a sentence-by-sentence basis. This method relies on anchor words (corresponding to disease phenotypes) and half of the completed sentences (termed prefix text) determined by medical professionals. CRAC, the system behind DLCRAC, employs an adaptive recover-and-edit framework, allowing for the step-by-step generation of reports under the guidance of doctors.

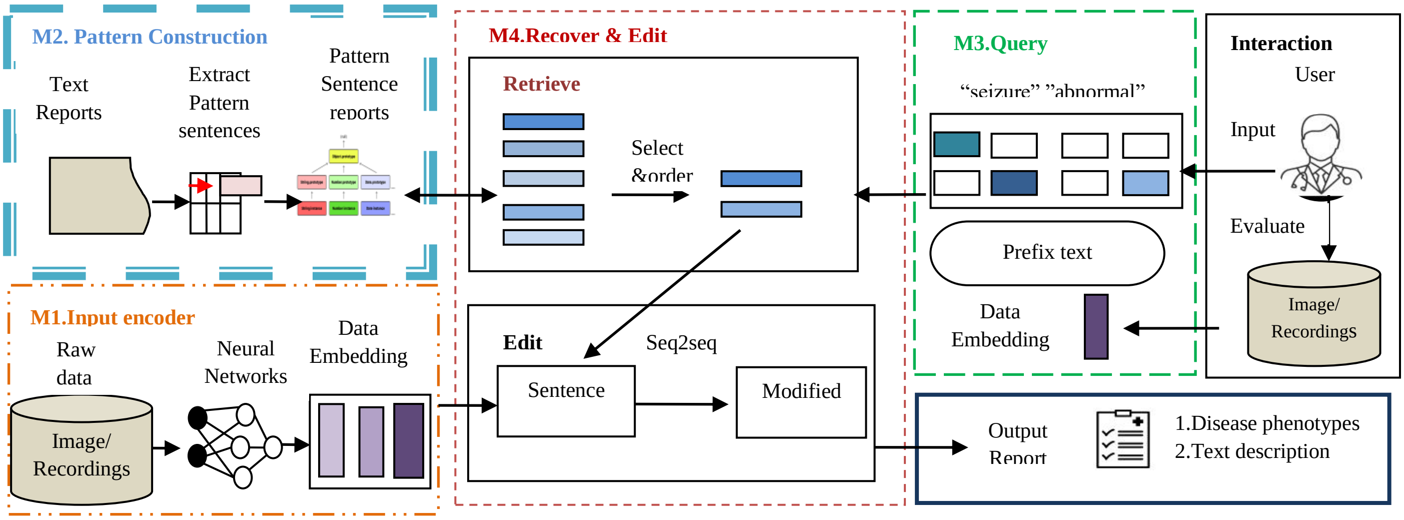

Fig. 2. DLCRAC Framework

Figure 2 shows the deep learning based clinical report auto-completion framework that comprises modules M1, M2, M3 and M4. DLCRAC comprises an input encoder module, which learns embeddings from medical images and neural recordings. Additionally, it constructs a prototype repository by indexing unique sentences from various medical reports. Users provide anchor words and prefix text as queries to retrieve the most relevant sentence templates. These templates undergo modifications in the edit module using a seq2seq model, ultimately generating new sentences for the current report. This process iteratively generates all sentences within the report description and their associated disease phenotypes, enhancing the overall report-writing process.

M1 serves as the Input Encoder module, responsible for converting medical data, like images or EEG time series, into compact feature representations, facilitating efficient data processing and analysis. This transformation enables the subsequent stages of the system to work with compressed yet informative data representations.

This module serves the purpose of extracting data embeddings from the input, providing guidance for report completion. The input may consist of raw measurements from X-ray or EEG sources. In the case of images X^i or EEG time series TS represented as a sequence of EEG epochs TS = x1, x2, ..., xT, we employ a convolutional neural network (CNN) for encoding. This process yields image embedding $E^{(i)}$ or EEG embedding E_n^i for each epoch t, facilitating subsequent report generation.

$$E^i = CNN[X^i] \tag{1}$$

$$E_n^{(i)} = CNN[X_i] \tag{2}$$

CNN uses the DenseNet [16] framework for X-ray imaging. The mean imaging for all phases in EEG is the overall embedding, represented as

$$E^i = \frac{\sum_n E_n^{(i)}}{T} \tag{3}$$

The M2 module is dedicated to Pattern Construction, where it creates a repository at the sentence level. This repository contains distinctive sentences, along with their representations, author details, and frequency statistics. This valuable resource is derived from an extensive medical report database. It serves as a dynamic search repository, offering initial sentence structures for the generation of sentences within new reports, enhancing the efficiency and consistency of report writing.

Pattern learning [17] and memory networks [18] offer distinct approaches for integrating data instances into neural networks.

Both methods share a fundamental concept, which involves the creation of a collection of patterns.

$$X_p = [X_{p1}, X_{p2}, \dots \dots X_{pn}] \quad (4)$$

It is represented as,

$$\{f(p)|p \in X_p \quad (5)$$

M3, the Query module, offers clinicians greater control in the interactive creation of personalized medical reports. This module accommodates clinician queries in two distinct forms: anchor words, representing the global context and serving as phenotype keywords relevant to the entire report, and optional prefix text, which includes partially composed sentences entered by users during the interactive editing process. This approach empowers clinicians to tailor reports to their specific needs, combining both global and local context cues for precise report customization.

The Query Component provides interactive report generation easier, enabling consumers to quickly and effectively create reports phrase by phrase. It offers two modes of interaction:

Anchor words (A): The aforementioned terms that gives the report's high-level context. Anchor words in EEG reports may contain terminology such as focal slowing, epi-leptiform, typical rest, and seizures. As shown in [19], anchor terms in X-ray findings might also include “Pneumonia, Cardiomegaly, Lung Lesion, Airspace Opacity, Edoema, Pleural Effusion, and Fracture”.

Prefix text ($P_n^{(m)}$): This specifies a partial sentence for sentence n in report m. Prefix text allows users to customize and exercise control. It's important to note that the use of prefix text is entirely optional in DLCRAC. Both anchor words and prefix text play a role in the Recover module when seeking relevant sentences from the pattern repository.

M4, the Recover and Edit module, actively engages users in report generation, drawing upon data representations, anchor words, and prefix text to guide the process. This module follows a sequential approach to report creation. Initially, the retrieval module extracts the most pertinent sentences from the pattern repository. Subsequently, the edit module employs a phrase-to-phrase framework [20] to refine the recovered sentences, incorporating input from the data representation, anchor words, and prefix text. This dynamic interaction ensures the production of tailored and context-aware medical reports.

During the retrieval phase, an information retrieval system is employed to locate the most pertinent sentences within the pattern repository. This process emulates a doctor referencing their prior reports to pinpoint sentences that require modification. With an anchor word, $A_l^{(i)}$, and the possibility of including prefix text, this module recovers a template sentence, S_l, from the pattern repository.

In the absence of anchor words, DLCRAC employs an initial prediction of anchor words. This is achieved by training a classifier using data embeddings $E^{(i)}$ to generate anchor words A^i. In contrast to alternative retrieval methods, this proposed approach is characterized by increased flexibility and scalability, harnessing the capabilities of retrieval systems.

During the editing phase, the obtained sentence undergoes modifications to yield the final sentence for the current report. A sequence-to-sequence model was employed, comprising an encoder and a decoder. The encoder maps the input, combining the sentence template S_l and data embedding $E^{(i)}$, into compressed representations. The decoder, in turn, reconstructs the output, resulting in the revised sentence as the output sequence.

4. Results and Discussion

DLCRAC's capability to enhance clinical report quality is evident in this research. A comparison was conducted between DLCRAC and state-of-the-art baseline techniques, encompassing report-level auto-completion with predefined anchor words, report-level auto-completion without predefined anchor words, and sentence-level auto-completion. The summary of report-level performance on X-ray and EEG datasets is presented in Table 1. DLCRAC outperforms the leading baseline models, achieving a remarkable 30% improvement in the CIDEr score. This outcome underscores the efficiency of the retrieval process from the pattern repository.

Table 1. Report Level Performance

Dataset	Technique	CIDEr	BLEU1	BLEU2	BLEU3
IU X-Ray Image	CNN-RNN	0.254	0.291	0.192	0.100
	Adaptive Attention	0.257	0.220	0.158	0.090
	DLCRAC (Predicted anchor words)	0.378	0.332	0.345	0.195
	DLCRAC (Defined anchor words)	0.397	0.376	0.365	0.210
(EEG) TUH	MP	0.278	0.623	0.462	0.543
	TAM	0.296	0.713	0.567	0.661
	DLCRAC (Predicted anchor words)	0.412	0.786	0.653	0.689
	DLCRAC (Defined anchor words)	0.479	0.798	0.660	0.712

For a more in-depth analysis of individual module contributions, DLCRAC was assessed without the edit module, focusing solely on sentence retrieval from existing reports. Despite still surpassing baseline models in CIDEr performance, this version of DLCRAC falls short compared to the fully-featured DLCRAC with predicted anchor words, which utilizes both retrieval and edit modules.

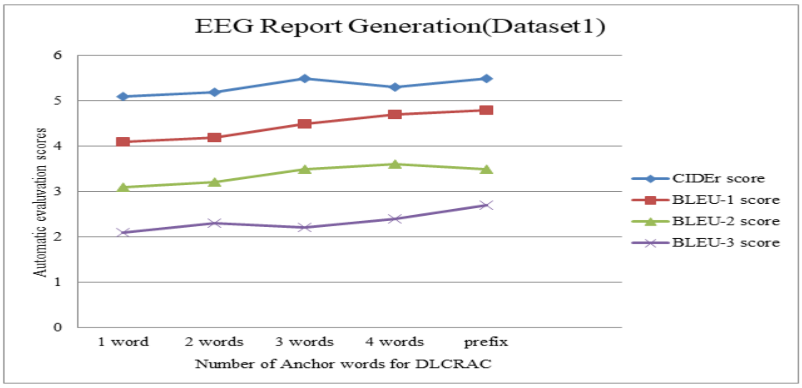

Fig. 3. Score for EEG Report Generation (Dataset 1)

The plot presented figure 3, figure 4 and figure 5 illustrates the progressive rise in CIDEr and BLEU scores concerning the generation of EEG and X-ray reports as the number of anchor words increases. This observed trend of growing scores with an expanding set of anchor words serves as compelling evidence that anchor words play a crucial role in enhancing the quality of reports generated by DLCRAC. The increase in anchor words effectively guides the DLCRAC system in selecting superior candidate sentences and refining them, reflecting the valuable contribution of clinicians as they furnish more anchor words to the system.

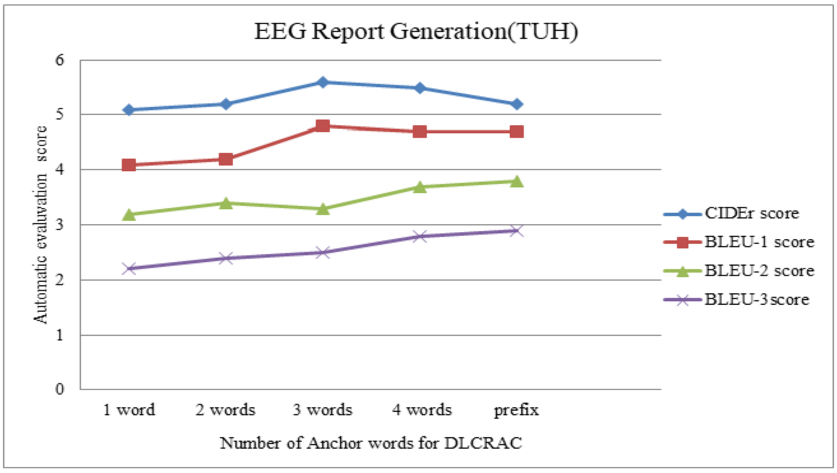

Fig. 4. Score for EEG Report Generation (TUH)

To optimize the performance of DLCRAC in sentence-by-sentence interactive report auto-completion using anchor words and prefix text, we conducted evaluations with varying numbers of anchor words and prefix sentences. These evaluations encompassed the utilization of anchor words until 5 with DLCRAC and variable-length prefix sentences, allowing us to explore the effects of increasing anchor words on the system's performance.

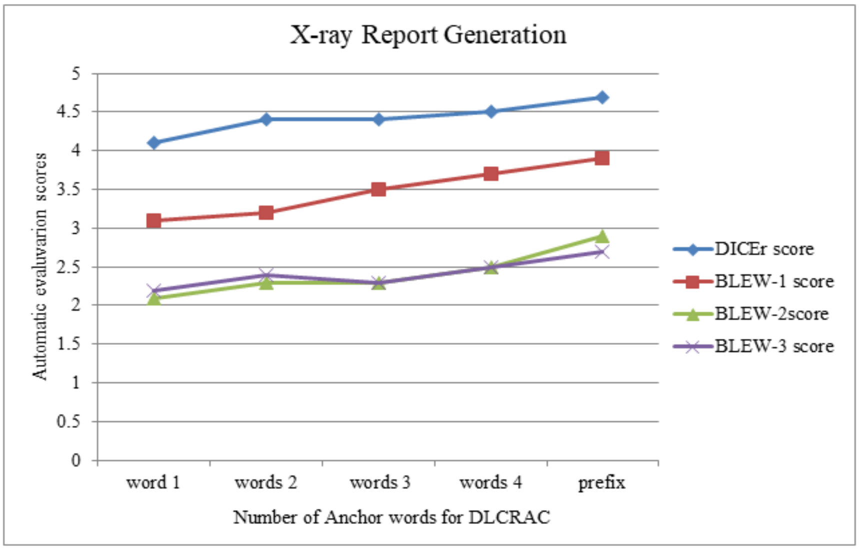

Fig. 5. Score for X-Ray Report Generation

Creating clinical reports from raw medical recordings like X-rays and electroencephalogram (EEG) is a crucial and routine duty for medical professionals. Nevertheless, crafting precise and comprehensive reports can be a time-intensive endeavor. According to the result graphs, DLCRAC achieved higher score and accuracy as shown in above figures. Using DLCRAC in conjunction with various anchor words guarantees that important findings are included in the analysis. Anchor words require the analysis generating unit to be operationally correct because they are founded on facts that are medically important. The suggested approach for overcoming obstacles using multimodal deep learning techniques and using these to apply to significant healthcare issues.

5. Conclusion

The process of crafting medical reports is vital but typically labor-intensive for human healthcare professionals. Existing methodologies for generating medical reports mainly concentrate on generating entire reports without close human interaction, which can lead to errors and doesn't align well with clinical workflows. In this study, DLCRAC is introduced as a computational solution designed to assist in the clinical report auto-completion task. DLCRAC enables doctors to compose clinical reports interactively, sentence by sentence. This approach amalgamates an information retrieval engine and neural networks, allowing the retrieval of pertinent sentences through information retrieval systems and their subsequent modification using neural networks. Empirical results from our experiments illustrate that DLCRAC excels in generating high-quality, clinically precise reports. It consistently outperforms various baseline methods, exhibiting remarkable improvements of up to 35% in CIDEr and BLEU-3 when compared to the best-performing baseline techniques.

References

[1] Nguyen, P. T., Huynh, V. D. B., Vo, K. D., Phan, P. T., Elhoseny, M., & Le, D. N. (2021). Deep Learning Based Optimal Multimodal Fusion Framework for Intrusion Detection Systems for Healthcare Data. *Computers, Materials & Continua*, *66*(3).

[2] Haleem, M. S., Ekuban, A., Antonini, A., Pagliara, S., Pecchia, L., & Allocca, C. (2023). Deep-Learning-Driven Techniques for Real-Time Multimodal Health and Physical Data Synthesis. *Electronics*, *12*(9), 1989.

[3] Chen, Q. Q., Li, J. P., Agbley, B. L. Y., Hussain, A., Khan, I., Khan, R. U., ... & Ali, I. (2023). A Multimodal Network Security Framework for Healthcare Based on Deep Learning. *Computational Intelligence and Neuroscience*, *2023*.\

[4] Hu, Y., Tang, J., Zhao, S., & Li, Y. (2022). Deep learning-based multimodal 3 T MRI for the diagnosis of knee osteoarthritis. *Computational and Mathematical Methods in Medicine*, *2022*.

[5] Tripathi, A., Waqas, A., Venkatesan, K., Yilmaz, Y., & Rasool, G. (2023). Building Flexible, Scalable, and Machine Learning-ready Multimodal Oncology Datasets. *arXiv preprint arXiv:2310.01438*.

[6] VJ, S. (2021). Deep learning algorithm for COVID-19 classification using chest X-ray images. *Computational and Mathematical Methods in Medicine*, *2021*.

[7] Hemdan, E. E. D., Shouman, M. A., & Karar, M. E. (2020). Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055*.

[8] Apostolopoulos, I. D., Aznaouridis, S. I., & Tzani, M. A. (2020). Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases. *Journal of Medical and Biological Engineering*, *40*, 462-469.

[9] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88.

[10] Zhang, Y. D., Dong, Z., Wang, S. H., Yu, X., Yao, X., Zhou, Q., ... & Gorriz, J. M. (2020). Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. *Information Fusion*, *64*, 149-187.

[11] Bellemo, V., Burlina, P., Yong, L., Wong, T. Y., & Ting, D. S. W. (2019). Generative adversarial networks (GANs) for retinal fundus image synthesis. In *Computer Vision–ACCV 2018 Workshops: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers 14* (pp. 289-302). Springer International Publishing.

[12] Wu, S., Ren, Y., Yang, A., Chen, X., Pu, X., He, J., ... & Yu, P. S. (2023). Deep Learning and Medical Imaging for COVID-19 Diagnosis: A Comprehensive Survey. *arXiv preprint arXiv:2302.06611*.

[13] Kashala Kabe, G., Song, Y., & Liu, Z. (2021, September). Novel distant domain transfer learning method for COVID-19 classification from X-rays images. In *Proceedings of the 5th International Conference on Algorithms, Computing and Systems* (pp. 127-134).

[14] Heiliger, L., Sekuboyina, A., Menze, B., Egger, J., & Kleesiek, J. (2022). Beyond medical imaging-A review of multimodal deep learning in radiology. *TechRxiv*, (19103432).

[15] Jing, B., Xie, P., & Xing, E. (2017). On the automatic generation of medical imaging reports. *arXiv preprint arXiv:1711.08195*.

[16] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[17] C. Y. Li, X. Liang, Z. Hu, and E. P. Xing, “Knowledge-driven encode, retrieve,paraphrase for medical image report generation,” arXiv preprint arXiv:1903.10122,2019.

[18] Uysal, A. S., Li, X., & Mulvey, J. M. (2023). End-to-end risk budgeting portfolio optimization with neural networks. *Annals of Operations Research*, 1-30.

[19] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., ... & Ng, A. Y. (2019, July). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 590-597).

[20] Skeggs, R. (2023). *Using domain specific language and sequence to* sequence *models as a hybrid framework for a natural language interface to a database solution* (Doctoral dissertation, Brunel University London).

[21] Pandey, J.K., Ahamad, S., Veeraiah, V., Adil, N., Dhabliya, D., Koujalagi, A., Gupta, A. Impact of call drop ratio over 5G network (2023) Innovative Smart Materials Used in Wireless Communication Technology, pp. 201-224.

[22] Mondal , D. (2021). Green Channel Roi Estimation in The Ovarian Diseases Classification with The Machine Learning Model . Machine Learning Applications in Engineering Education and Management, 1(1), 07–12.

[23] Ólafur, J., Virtanen, M., Vries, J. de, Müller, T., & Müller, D. Data-Driven Decision Making in Engineering Management: A Machine Learning Framework. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/108