Article

Automated Clinical Impression Generation for Medical Signal Data Searches

Woonghee Lee [1] , **Jaewoo Yang** [2] , **Doyeong Park** [2] **and Younghoon Kim** [2,*]

[1] BK21 Education and Research Center for Artificial Intelligence in Healthcare, Department of Applied Artificial Intelligence, Hanyang University, Ansan 15588, Republic of Korea; woongheelee@hanyang.ac.kr
[2] Department of Applied Artificial Intelligence (Major in Bio Artificial Intelligence), Hanyang University, Ansan 15588, Republic of Korea; onnoo@hanyang.ac.kr (J.Y.); enter1994@hanyang.ac.kr (D.P.)
[*] Correspondence: nongaussian@hanyang.ac.kr

Abstract: Medical retrieval systems have become significantly important in clinical settings. However, commercial retrieval systems that heavily rely on term-based indexing face challenges when handling continuous medical data, such as electroencephalography data, primarily due to the high cost associated with utilizing neurologist analyses. With the increasing affordability of data recording systems, it becomes increasingly crucial to address these challenges. Traditional procedures for annotating, classifying, and interpreting medical data are costly, time consuming, and demand specialized knowledge. While cross-modal retrieval systems have been proposed to address these challenges, most concentrate on images and text, sidelining time-series medical data like electroencephalography data. As the interpretation of electroencephalography signals, which document brain activity, requires a neurologist's expertise, this process is often the most expensive component. Therefore, a retrieval system capable of using text to identify relevant signals, eliminating the need for expert analysis, is desirable. Our research proposes a solution to facilitate the creation of indexing systems employing electroencephalography signals for report generation in situations where reports are pending a neurologist review. We introduce a method incorporating a convolutional-neural-network-based encoder from DeepSleepNet, which extracts features from electroencephalography signals, coupled with a transformer which learns the signal's auto-correlation and the relationship between the signal and the corresponding report. Experimental evaluation using real-world data revealed our approach surpasses baseline methods. These findings suggest potential advancements in medical data retrieval and a decrease in reliance on expert knowledge for electroencephalography signal analysis. As such, our research represents a significant stride towards making electroencephalography data more comprehensible and utilizable in clinical environments.

Keywords: electroencephalography; indexing method; text generator; transformer; convolutional neural network

1. Introduction

Medical retrieval systems have garnered interest among clinicians [1]. However, commercial retrieval systems like Apache Lucene [2] and Elasticsearch [3], which hinge on term-based indexing, lack efficacy when dealing with continuous medical data, such as electroencephalography (EEG) data. Moreover, the processes of annotating, classifying, and interpreting medical data are resource intensive and demand expert domain knowledge. Conversely, the cost of data recording systems has become increasingly affordable [4].

In response to these challenges, researchers have suggested cross-modal retrieval systems. However, these proposals predominantly focus on images and texts [5–9], leaving time-series medical data like EEG data, which documents brain activity, largely overlooked [10]. Analyzing and diagnosing EEG signals necessitates the expertise of a neurologist, which makes it the costliest step. Thus, a proficient retrieval system that can use text to locate corresponding signals without requiring expert analysis is in high demand.

The aim of this work is to aid the creation of indexing systems that utilize EEG signals for report generation. The proposed system in this work is designed for the following scenario: A medical center collects EEG signals from patients and stores them in a dedicated database. However, the accompanying reports are yet to be produced by a neurologist. If a neurologist or researcher requests specific signals using a query, such as ''`abnormal EEG due to: Asymmetric EEG background with left arrhythmic delta activity`'', as depicted in Figure 1, the signals need to be characterized with their respective impressions for storage in a term-based indexing system.
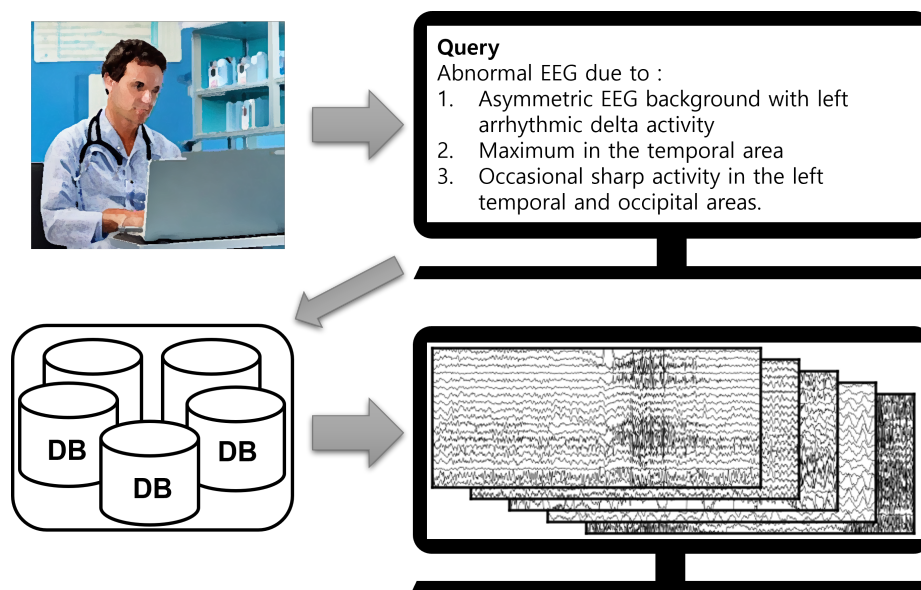
Figure 1. Flowchart for retrieving EEG signals based on impression queries.

The proposed method in this work incorporates an encoder that employs convolutional neural networks (CNNs) from DeepSleepNet [11] to extract EEG signal features, along with a transformer [12] that learns the signal's auto-correlation and the relationship between the signal and the target report. Experimental evaluation with real-world data showed the proposed approach in this work surpasses baseline methods.

The contributions of this work can be summarized as follows.

- Development of a method supporting indexing systems for EEG signals in the medical domain, specifically tailored for report generation purposes.
- Introduction of a cost-effective system that minimizes the requirement for extensive neurologist involvement in the EEG indexing process.
- Demonstration of superior performance and effectiveness through rigorous experimentation using real-world data, surpassing baseline approaches.

2. Methods

In this context, the recorded EEG from a patient in a session is represented as S, while the corresponding impression provided by a neurologist is denoted as T. The EEG signal S consists of a sequence of data points e_i for a specific time i within the duration from 0 to t. This can be expressed as $S = \{e_0, \ldots, e_{t-1}\}$. For instance, if an EEG signal S is recorded at a frequency of 200 Hz for a span of one hour, then the total number of data points in S would be 720,000. Hence, the signal can be represented as $S = \{e_0, \ldots, e_{720,000-1}\}$. It is important to note that EEG signals in practice are recorded by multiple channels. However, for the sake of simplicity, the multiple channels are abstracted as e_i for a specific time i. Therefore, for each time i, the data point e_i can represent a vector which includes three scalar values if there are three channels, such as FP1, FP2, and F3.

2.1. Text Generation for EEG Signal Using DeepSleepNet Encoder and a Transformer

In an information retrieval system, indexing a sentence can be easily implemented using an inverted index [13]. However, because the EEG signal is a continuous value, it is difficult to directly index it into an existing database and use search engines, such as Lucene [2] or Elasticsearch [3]. Therefore, this work proposes a text generator for the signal. Once the signal is described by the generated text, it can be stored using a word-level inverted index [14]. For example, as depicted in Figure 2, the signals named *Doc 1* already have an impression reported by neurologists, stating "Abnormal EEG. Excess drowsiness. Low voltage pattern". However, the signals named *Doc 2* do not have any report yet. Thus, the goal of this research is to generate an associated report that precisely describes the signal using our text generator and to apply an inverted index using a tokenizer, such as a one-gram or bi-gram tokenizer. To achieve this goal, this work introduces a text generator for word-level inverted indexes that combines convolutional neural networks (CNNs) borrowed from DeepSleepNet [11] and a transformer [12].
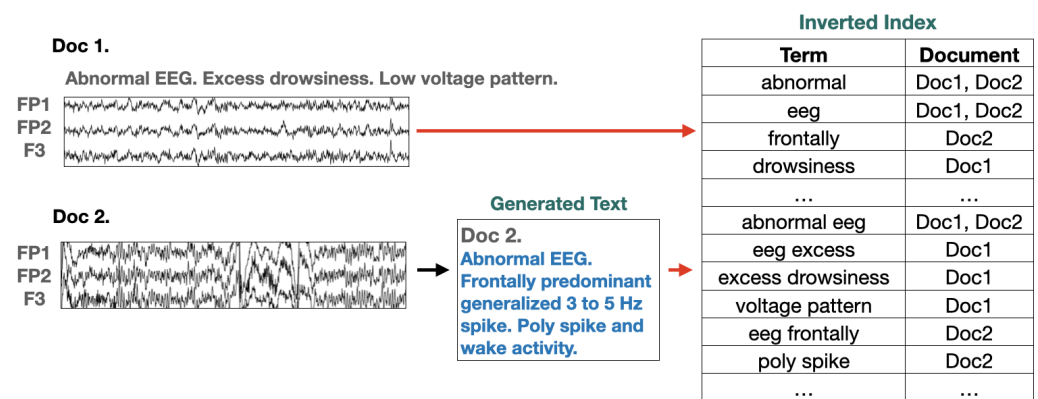

Figure 2. Flowcharts for inverted indexing of signals with impressions, both provided by a neurologist (*Doc 1*) and generated by the proposed method (*Doc 2*).

DeepSleepNet [11] is a model proposed for detecting sleep stages in time series EEG data. It is composed of both convolutional neural networks (CNNs) and long short-term memory networks (LSTMs), which are used to capture patterns of EEG activity in both local and global areas, respectively. In this work, the CNN portion of the model is only considered since the proposed method utilizes the transformer to learn patterns in the global area. The convolutional layers of DeepSleepNet are designed to detect low- and high-frequency patterns along the time axis using two branches of large and small filters, making it a popular choice for research aiming to capture features from EEG signals [15–17]. The two-branch CNN structure proposed in [11] is adopted in the suggested method in this work for mapping the signal into a feature space. However, the network is modified to be multi-channel CNNs to accommodate EEG signals recorded from diverse channels. The modified version is named in this work DeepSleepNet Encoder (DE), which is depicted in Figure 3. Given an EEG signal $S = \{e_0, \ldots, e_{t-1}\}$, the signal is divided into fixed-length segments, such as every 30 s, yielding segments $\{e_0, \ldots, e_{6000-1}\}, \ldots, \{e_{t-6000} \ldots, e_{t-1}\}$. Then, each segment is fed into the DeepSleepNet Encoder to extract local features, resulting in a set of features $F = \{f_0, \ldots, f_m\}$.

Transformer [12] is an encoder–decoder model which is based on a self-attention mechanism for sequence transduction, especially machine translation. The self-attention mechanism was proposed to learn a sentence representation by using query Q, key K, and value V. In the encoder, the self-attention mechanism discovers relationships among the inputs using the scaled dot-product attention proposed in [12], which is defined as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

where $Q = FW^Q$, $K = FW^K$, and $V = FW^V$ in our problem for the initial layer and W is the matrix of the trainable parameters. The encoder of the transformer is utilized to discover relationships among the divided signal segments. Additionally, the decoder of the transformer is exploited to capture the correlation among words in the target sentence, as well as to learn the relationship between the words and the input features given by the encoder using the cross-attention mechanism. During training, the decoder captures the relationship between the sentence and the signal using the features of the sentence as query Q, and the features from the encoder as K and V, as shown in Figure 3.
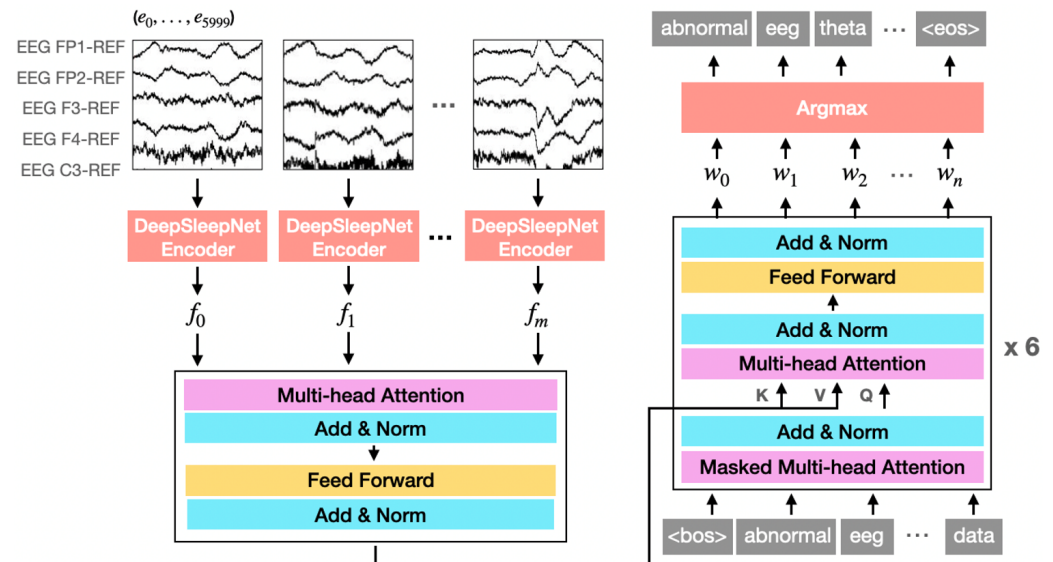
Figure 3. Diagram of the encoder–decoder architecture of the proposed method, using DeepSleepNet and a transformer, presented in this work.

To summarize, the objective of this work is to minimize the negative log-likelihood (NLL), given pairs (S, T) of signals S and texts $T = \{w_0, \ldots, w_n\}$ in the training set as shown below.

$$\min NLL = -\sum_{i=0}^{n} \log P(w_i | w_{i-1}, \ldots, w_0, S) \tag{2}$$

3. Experiments

In this section, the experimental setup is described and the results are discussed. The experiments were conducted on a workstation with an Intel(R) Core(TM) i9-7900X CPU @ 3.30GHz, 128GB of RAM, and an NVIDIA GeForce GTX 1080Ti GPU running on Ubuntu 20. Additionally, this work aims to demonstrate the practicality of the proposed method for use in existing retrieval systems. Therefore, Elasticsearch [3] is utilized to perform the retrieval and calculate the relevance score.

3.1. Data Description and Preprocessing

The EEG data corpus collected by the Temple University Hospital (TUEG) [10] is utilized, which is publicly available (https://isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml, accessed on 28 October 2022). The dataset comprises more than 30k clinical EEG recordings collected since 2002, organized by patient and corresponding session. For each session, there is a pair of an EEG recording and a physician report, which includes the patient description, clinical history, medications, and clinical impression. Since only the impression in the report is considered, sessions without impressions were removed.

The average length of EEG recordings is around 30 min. Sessions in which the signal was longer than 30 min were removed. The raw EEG signals were recorded in the range of 250–1000 Hz and across various channels. Therefore, the signals were resampled at 200 Hz

and utilized 19 channels: FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, CZ, A1, and A2. Additionally, bipolar montage was applied to the channels, as physicians consider it a clearer and more symmetric way to visualize and assess the signals [18]. The power line noise is considered as an artifact that exists in the frequency band around 50 Hz [19]. Therefore, the noise was removed using 4th-order Butterworth filters, cutting off frequencies between 47.5 and 52.5 Hz and 57.5 and 62.5 Hz, as recommended in [20].

In addition, as mentioned in Section 2.1, the signal is split into 30 s segments. Therefore, each signal can be divided into at most 60 segments. We divided the data into a training set (70%), validation set (15%), and test set (15%) based on patient. As a result, the obtained sets include 7917, 1685, and 1700 sessions from a total of 11,302 sessions for the training, validation, and test sets, respectively.

3.2. Implemented Models and Details

The proposed model in this work and the baseline models were implemented in PyTorch version 1.12.1 [21] using the ADAM optimizer [22] with a learning rate of 0.0001 and a batch size of 32, except for the CNN-LSTM model, which had memory issues. Additionally, each model was trained for 40 epochs and the best epoch was selected based on the BLEU [23] score measured on the validation set after training. The implemented models and their details are described as follows.

- **Ours:** Two branches of convolutional neural networks (CNNs) are utilized to process EEG signals divided into 30 s segments. The proposed architecture is adapted from DeepSleepNet, but modified to accommodate multi-channel inputs. Detailed specifications of the CNN encoder, including layer sequences and hyper-parameters, are presented in Table 1. Alongside this, a transformer is incorporated into the system. Crucial hyper-parameters for the transformer, including the dimension of the embedding layer, the number of heads, and the quantity of encoder and decoder layers, are delineated in Table 2. We adhere to the hyper-parameter values specified in the original DeepSleepNet [11] and transformer model [12]. Implementations using the DeepSleepNet Encoder also follow these same hyper-parameter values, except for the number of layers, which are indicated in Table 2. The number of layers is selected using a grid search from 1 to 6.
- **CNN-LSTM:** In [24,25], text generators are introduced that exploit domain knowledge given by doctors or retrieved sentences from a database. Their models are based on CNNs and LSTMs for embedding input signals and sequence-to-sequence modeling, respectively. Since a text generator is considered given an EEG signal without additional domain knowledge written in sentences, a sequence-to-sequence model is implemented similar to [25]. The batch size for this implementation is 4, as it does not compress input signals enough and it causes memory issues. Note that the best epoch is selected based on the BLEU score using a validation set to ensure a fair comparison.
- **DE-LSTM:** Given that the CNN-LSTM model previously mentioned does not compress input signals into the feature space adequately, leading to memory issues, the DeepSleepNet Encoder (DE) is employed as the signal encoder in the model proposed in by this research, effectively supplanting the CNN component.
- **DE-LSTM-ATTENTION:** The attention mechanism, as proposed by Bahdanau et al. [26], is a powerful tool for capturing the relationship between input and output in a neural network. In this research, Bahdanau attention is employed to understand the correlation between segmented EEG signals and the tokens generated from them. This attention mechanism is integrated into the DE-LSTM model, which is named DE-LSTM-Attention in this section.

Table 1. Detailed structure and hyper-parameters of DeepSleepNet Encoder.

Input: a segment of the signal	
Conv1d (C, 64, sfreq//2, sfreq//16) BatchNorm1d (64) ReLU	Conv1d (C, 64, sfreq*4, sfreq//2) BatchNorm1d (64) ReLU
MaxPool1d (8, 8) Dropout	MaxPool1d (4, 4) Dropout
Conv1d (64, 128, 8, 1) BatchNorm1d (128) ReLU	Conv1d (64, 128, 6, 1) BatchNorm1d (128) ReLU
Conv1d (128, 128, 8, 1) BatchNorm1d (128) ReLU	Conv1d (128, 128, 6, 1) BatchNorm1d (128) ReLU
Conv1d (128, 128, 8, 1) BatchNorm1d (128) ReLU	Conv1d (128, 128, 6, 1) BatchNorm1d (128) ReLU
MaxPool1d (4, 4)	MaxPool1d (2, 2)
Output: concatenation along the last axis	

Table 2. Hyper-parameters of the transformer.

Parameter Name	Value
Embedding dimension	512
The number of heads	8
The number of encoder layers	1
The number of decoder layers	6

3.3. Evaluation Metric

To evaluate the proposed method, normalized discounted cumulative gain (NDCG) was utilized, which measures the performance of an information retrieval system [27]. NDCG is a commonly used evaluation metric in ranking-based recommendation systems. Specifically, it is exploited to measure the quality of an information retrieval system. NDCG is based on cumulative gain (CG). CG represents the summation of the relevance scores of retrieved items. In a retrieval system, the CG of items is discounted (DCG) in order to penalize items which are less relevant using logarithmic reduction factors. Finally, NDCG is obtained by the DCG of predicted items out of the DCG of true items as follows:

$$\text{NDCG@k} = \frac{1}{Z} \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2(i+1)} \tag{3}$$

where Z is the ideal DCG of the items, which is computed by sorting the items by true relevance scores, and k is the number of retrieved items from the retrieval system.

3.4. Retrieval Performance

The retrieval performance of the methods described in Section 3.2 was compared. The impression given by every signal using each implemented method was generated. After that, they were indexed and stored into Elasticsearch using a standard tokenizer [28] and n-gram from 1 to 4. A total of 1700 queries were searched in the test set. The relevance score was calculated by measuring the similarity between query and the documents stored in Elasticsearch using default settings such as BM25 [13]. Although commercial search engines typically display 10 documents per page (e.g., Google and Bing), the comparisons in this study provide NDCG scores for 5 to 30 documents for more detailed comparisons,

in increments of five, as depicted in Figure 4. Since NDCG is calculated based on the number of search results, a higher score is achieved with more search results.
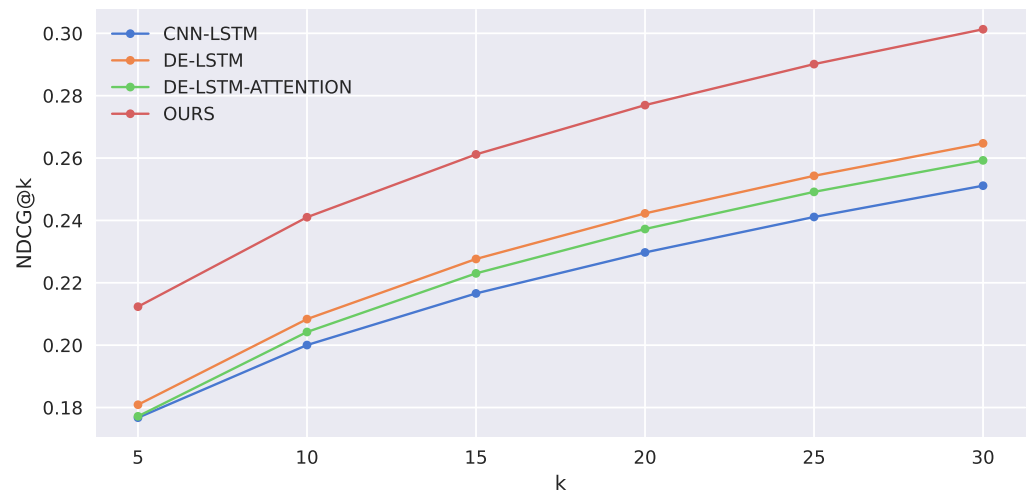
Figure 4. Comparison of average NDCG scores between the proposed model and baselines for each set of five relevant documents. k denotes the number of relevant documents.

The figure shows that the proposed method in this work achieves the best performance compared to the others. It is speculated that the DeepSleepNet Encoders (DE) capture the signal features effectively. As shown in the graph, the proposed method and the DE-LSTM and DE-LSTM-ATTENTION methods, which include the DE, outperform the CNN-LSTM method without DE, while the attention mechanism in DE-LSTM-ATTENTION performs better than DE-LSTM. The self-attention and cross-attention in the transformer used in the proposed method have a greater impact on indexing the signal.

Additionally, in order to ensure the validity of the retrieval performance, the proposed method in this work was compared with each baseline using a t-test. The p-values obtained from the t-test for CNN-LSTM, DE-LSTM, and DE-LSTM-ATTENTION are 0.0174, 0.0377, and 0.0116, respectively. Therefore, the proposed method in this paper significantly outperforms the retrieval performance of the baselines.

3.5. Retrieval Results and Generated Sentences

While demonstrating the performance of proposed approach in this work across a range of examples is crucial, the limitations of the BLEU score as a measure of text generation quality must be acknowledged. This score may not fully capture the superiority of our method in EEG signal retrieval over baseline approaches. Therefore, examples chosen at random may not present an accurate portrayal of the comparative performance of our method. As a remedy to this, specific queries were selected representing both normal and abnormal EEGs, shown in Figures 5 and 6, respectively. The results are organized based on relevance score. Since the signals span about 30 min and are too lengthy for visualization in a single figure, a 5 s subset of each signal is presented. The associated original impression is displayed below the signal. It is important to note that every signal is indexed using the impression generated by each method, rather than the original impression.

For the query ''`normal eeg in wakefulness and brief stage 2 sleep`'' showcased in Figure 5, the proposed method in this work accurately indexes normal signals exhibiting both wakefulness and stage 2 sleep. These results are presented in each column. Contrarily, the baseline methods misinterpret some abnormal signals as normal. In the case of an abnormal EEG, as represented in Figure 6, the query ''`abnormal eeg due to: low voltage suppressed background.  absence of reactivity or variability`'' is utilized. The proposed method accurately retrieves the precise signal displayed in the

rightmost column, whereas the baseline methods return signals which only partially match the query sentence.
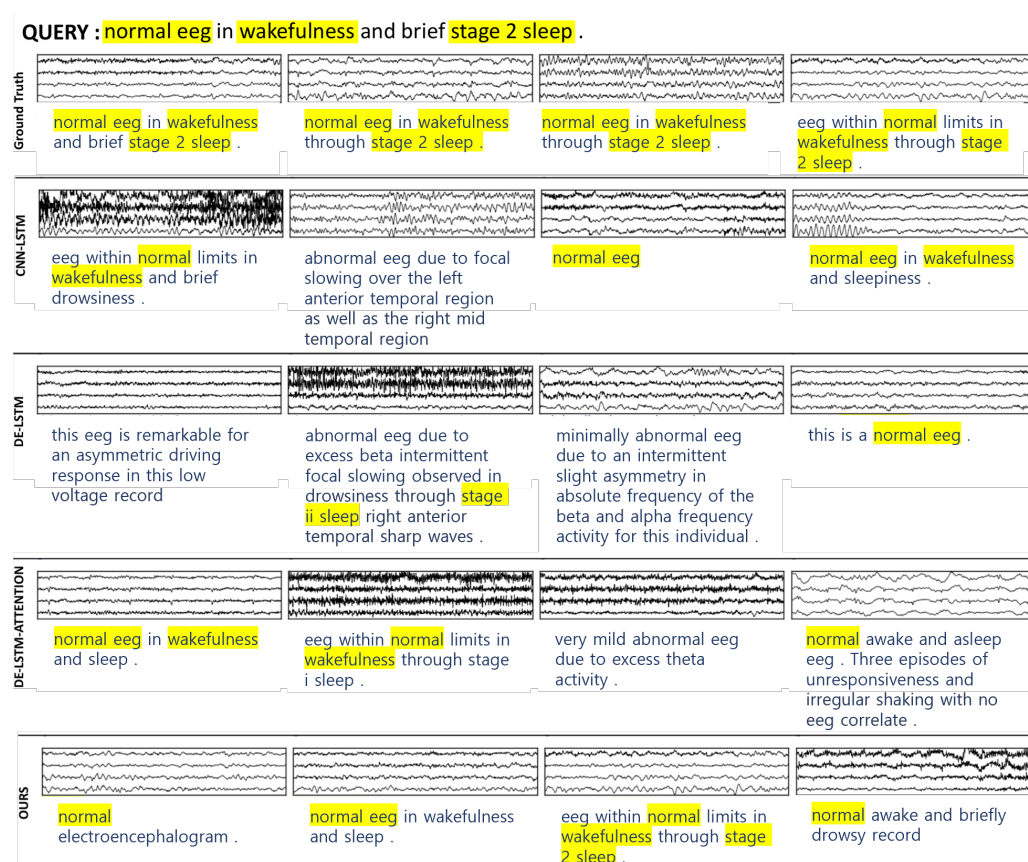
Figure 5. Retrieval results for the query 'normal EEG in wakefulness and brief stage 2 sleep'. Yellow highlights indicate matching descriptions between the query and the impressions.

3.6. The BLEU Scores for Each Methods

In addition, BLEU scores [23] were measured using five-fold Monte Carlo cross-validation [29] to compare the performance of text generation. Table 3 shows the BLEU scores for 1 to 4 g. Note that BLEU-n is the average BLEU score for 1 to n. The proposed method from this paper outperforms the BLEU scores of other methods, as indicated in the table.

Table 3. The average BLEU scores for each method.

Methods	BLEU-1	BLEU-2	BLEU-3	BLEU-4
CNN-LSTM	0.2802	0.2079	0.1684	0.1397
DE-LSTM	0.3615	0.2699	0.2176	0.1796
DE-ATTENTION-LSTM	0.3098	0.2358	0.1936	0.1623
OURS	0.3929	0.2890	0.2304	0.1884

To ensure the validity of the performance, every baseline and the proposed method in this work were compared using a *t*-test. Therefore, the *t*-test was conducted three times for the three baselines, namely CNN-LSTM, DE-LSTM, and DE-LSTM-ATTENTION, against OURS. The hypothesis for each test is as follows:

- Null hypothesis: The BLEU-4 score of OURS is the same as that of the baseline.
- Alternative hypothesis: The BLEU-4 of OURS significantly outperforms the baseline.

For each *t*-test, the significance level (alpha) is set to 10%. Subsequently, the *p*-values of CNN-LSTM, DE-LSTM, and DE-LSTM-ATTENTION are 1.0031^{-06}, 0.0967, and 0.0037, respectively. Therefore, the null hypothesis is rejected for the baselines.

Figure 6. Results for the query 'abnormal EEG due to: low voltage suppressed background'. Yellow highlights denote matching descriptions between the query and the impressions.

4. Discussion

In this section, the experimental results are analyzed and discussed. Additionally, the potential applications of the proposed method in this work, as well as its limitations and future directions, are provided.

4.1. Analysis of Results

In the experiments, retrieval results were compared using NDCG, a commonly used metric for evaluating search engines, as described in Section 3.4. The proposed method in this work outperforms the baselines in terms of indexing EEG signals. Moreover, the *t*-test shows that the proposed method in this work significantly exceeds the retrieval performance of the baselines. Furthermore, retrieval samples for queries searching for `normal eeg` and `abnormal eeg` are provided in Section 3.5. As shown in this section, the proposed method generates more plausible sentences compared to the baselines. Finally,

the ability of text generation is compared in Section 3.6. The BLEU scores demonstrate that the proposed method significantly outperforms the baselines in terms of text generation.

4.2. Available Applications of Proposed Method

The proposed method in this research, which aids the creation of indexing systems utilizing EEG signals for report generation, has various potential applications in the medical domain. Some of the available applications are as follows.

1. Indexing and organization: The proposed method allows for the efficient indexing and organization of EEG signals in a dedicated database. Signals can be characterized with their respective impressions and stored in a term-based indexing system, making it easier to manage and access the data.
2. Efficient signal retrieval: Neurologists and researchers can use the system to retrieve specific EEG signals based on queries. For instance, using a query like ``eeg within normal limits in wakefulness and brief sleep for an adult of this age'', the system can locate relevant signals and their corresponding impressions, streamlining the retrieval process.
3. Report generation support: The system can be implemented in medical centers where EEG signals are collected from patients, but accompanying reports are pending neurologist review. By using the proposed method, the system can automatically generate impressions for the signals, facilitating the report generation process.

Overall, the proposed method offers a promising solution to enhance EEG data utilization in clinical settings, making it more accessible and valuable for neurologists and researchers alike.

4.3. Limitations and Future Directions

This work addresses the novel problem of indexing commercial term-based databases using EEG signals and proposes a promising method based on a CNN and a transformer-based deep learning architecture. However, a major challenge arises from the limited availability of large-scale EEG datasets due to the sensitivity and privacy concerns associated with medical signals. To overcome this limitation, the authors intend to explore the development of a continually learning system that can adapt and improve over time without relying heavily on sharing private and sensitive medical data.

5. Conclusions

Given a multi-channel signal such as EEG signals, we propose using a report generator to index the signal into a retrieval system. This work adopts DeepSleepNet to extract features from the signal. Additionally, this research utilizes a transformer to incorporate self-attention for learning correlations between features extracted from the signal and cross-attention for grasping the connection between the signal and the target description. In experiments with real-world data, the retrieval results using the proposed method outperformed the baselines, which were based on CNNs, LSTMs, and attention networks. The proposed system is expected to facilitate the implementation of various applications, including indexing and organization of EEG signals in a dedicated database, efficient signal retrieval for neurologists and researchers to locate specific EEG signals using natural language queries, and report generation support. Additionally, the future scope of the research includes expanding the proposed method to incorporate multi-modal medical data, such as EEG and functional magnetic resonance imaging (fMRI) data, to enhance the understanding of human brain functions.

References

1. Dureja, A.; Pahwa, P. Integrating CNN along with FAST descriptor for accurate retrieval of medical images with reduced error probability. *Multimed. Tools Appl.* **2022**, *82*, 17659–17686. [CrossRef]
2. Białecki, A.; Muir, R.; Ingersoll, G.; Imagination, L. Apache lucene 4. In Proceedings of the SIGIR 2012 Workshop on Open Source Information Retrieval, Portland, OR, USA, 12–16 August 2012; p. 17.
3. Elasticsearch, B. Elasticsearch. 2018. Available online: https://www.elastic.co/pt/ (accessed on 12 September 2019).
4. Guo, K.; Wang, Y.; Kang, J.; Zhang, J.; Cao, R. Core dataset extraction from unlabeled medical big data for lesion localization. *Big Data Res.* **2021**, *24*, 100185. [CrossRef]
5. Cao, Y.; Steffey, S.; He, J.; Xiao, D.; Tao, C.; Chen, P.; Müller, H. Medical image retrieval: A multimodal approach. *Cancer Inform.* **2014**, *13*, CIN-S14053.
6. Müller, H.; Unay, D. Retrieval From and Understanding of Large-Scale Multi-modal Medical Datasets: A Review. *IEEE Trans. Multimed.* **2017**, *19*, 2093–2104. [CrossRef]
7. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
8. Simpson, M.S.; Demner-Fushman, D.; Antani, S.K.; Thoma, G.R. Multimodal biomedical image indexing and retrieval using descriptive text and global feature mapping. *Inf. Retr.* **2013**, *17*, 229–264. [CrossRef]
9. Zhen, L.; Hu, P.; Wang, X.; Peng, D. Deep supervised cross-modal retrieval. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 10394–10403.
10. Obeid, I.; Picone, J. The temple university hospital EEG data corpus. *Front. Neurosci.* **2016**, *10*, 196. [CrossRef] [PubMed]
11. Supratak, A.; Dong, H.; Wu, C.; Guo, Y. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 1998–2008. [CrossRef] [PubMed]
12. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
13. Schütze, H.; Manning, C.D.; Raghavan, P. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008; Volume 39.
14. Hossain, M.Z.; Sohel, F.; Shiratuddin, M.F.; Laga, H. A comprehensive survey of deep learning for image captioning. *ACM Comput. Surv. (CsUR)* **2019**, *51*, 1–36. [CrossRef]
15. Cai, X.; Jia, Z.; Tang, M.; Zheng, G. BrainSleepNet: Learning Multivariate EEG Representation for Automatic Sleep Staging. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 976–979.
16. Mousavi, S.; Afghah, F.; Acharya, U.R. SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach. *PLoS ONE* **2019**, *14*, e0216456. [CrossRef] [PubMed]
17. Farahat, A.; Reichert, C.; Sweeney-Reed, C.M.; Hinrichs, H. Convolutional neural networks for decoding of covert attention focus and saliency maps for EEG feature visualization. *J. Neural Eng.* **2019**, *16*, 066010. [CrossRef]
18. Britton, J.W.; Frey, L.C.; Hopp, J.L.; Korb, P.; Koubeissi, M.Z.; Lievens, W.E.; Pestana-Knight, E.M.; St Louis, E. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infancts*; American Epilepsy Society: Chicago, IL, USA, 2016.
19. de Cheveigné, A. ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage* **2020**, *207*, 116356. [CrossRef]
20. Chatzichristos, C.; Dan, J.; Narayanan, A.M.; Seeuws, N.; Vandecasteele, K.; De Vos, M.; Bertrand, A.; Van Huffel, S. Epileptic seizure detection in EEG via fusion of multi-view attention-gated U-net deep neural networks. In Proceedings of the 2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 5 December 2020; pp. 1–7.

21. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. *Automatic Differentiation in Pytorch*; 2017. Available online: https://openreview.net/forum?id=BJJsrmfCZ (accessed on 28 October 2019)

22. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

23. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: A method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318.

24. Biswal, S.; Xiao, C.; Westover, M.B.; Sun, J. Eegtotext: Learning to write medical reports from eeg recordings. In Proceedings of the Machine Learning for Healthcare Conference, PMLR, Ann Arbor, MI, USA, 9–10 August 2019; pp. 513–531.

25. Biswal, S.; Xiao, C.; Glass, L.M.; Westover, B.; Sun, J. CLARA: Clinical report auto-completion. In Proceedings of the Web Conference 2020, Taipei, Taiwan, 20–24 April 2020; pp. 541–550.

26. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.

27. Järvelin, K.; Kekäläinen, J. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst. (TOIS)* **2002**, *20*, 422–446. [CrossRef]

28. Davis, M.; Iancu, L. Unicode text segmentation. *Unicode Stand. Annex* **2018**, *29*, 65.

29. Dubitzky, W.; Granzow, M.; Berrar, D.P. *Fundamentals of Data Mining in Genomics and Proteomics*; Springer Science & Business Media: New York, NY, USA, 2007.