International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN(Online): 2347-5552, Volume-12, Issue-3, May 2024
https://doi.org/10.55524/ijircst.2024.12.3.19
Article ID IJIRD-1298, Pages 120-125
www.ijircst.org

Innovations in Time Related Expression Recognition Using LSTM Networks

Qishi Zhan[1], Yuhan Ma[2], Erdi Gao[3], Dan Sun[4], and Haowei Yang[5]

[1]Marquette University, Milwaukee, USA
[2]Johns Hopkins University, Baltimore, USA
[3]New York University, New York, USA
[4]Washington University in St. Louis, St. Louis, USA
[5]University of Houston, Houston, USA

Correspondence should be addressed to Qishi Zhan; qishizhan7@gmail.com

Received: 01 May 2024 Revised: 15 May 2024 Accepted: 28 May 2024

ABSTRACT- The proposed architecture leverages the strengths of both Convolutional Neural Network (CNN) and Bidirectional Long Short-Term (BLSTM) to create a robust model for temporal expression recognition in clinical texts. The CNN component effectively captures morphological and orthographic features at the character level, which enriches the semantic understanding of complex medical terminologies that are often abbreviated or have unique suffixes and prefixes. The BLSTM component excels in capturing long-range dependencies in text, which is crucial for understanding the context in which temporal expressions occur. By integrating these models with a CRF layer, the system not only predicts discrete labels but also ensures that the sequence of predicted labels is coherent and contextually appropriate, addressing the limitations of models that predict labels independently. The integration of pre-trained biomedical word vectors provides significant contextual grounding tailored to the medical domain, enhancing the model's ability to discern and interpret the nuances of medical language. This is crucial in clinical contexts where accurate interpretation of temporal phrases can be critical for patient management and treatment timelines. Further, experiments conducted on the dataset validate the effectiveness of the proposed model, demonstrating a notable improvement over traditional methods that rely heavily on hand-crafted features and rule-based approaches. Future work could explore the adaptability of this model to other subdomains of the medical field and its efficacy in processing multilingual texts, potentially increasing its applicability in global healthcare settings, with further refinement of the neural architecture and optimization of training strategies potentially yielding even better performance and faster processing times essential for real-time clinical decision support systems.

KEYWORDS- Deep Learning, Target Detection, Bidirectional Long Short-Term, Convolutional Neural Network

I. INTRODUCTION

The extraction of temporal information from medical records has become a focal point in recent biomedical research. Such information aids healthcare professionals and researchers in understanding disease progression patterns and dynamic medical phenomena, forming a basis for clinical pathway studies and the development of intelligent decision support systems [1-3]. Due to the fast-paced and specialized nature of their work, temporal information in these records often lacks uniformity in format, displays irregular expressions, and is intertwined with medical events. These characteristics pose significant challenges for identifying temporal phrases and hinder the effective use of temporal information. Consequently, the automatic identification of temporal phrases in medical records has emerged as a crucial research topic, garnering increasing attention. In the context of English medical records, the series of Clinical TempEval shared tasks have been particularly influential. These tasks have been evaluated at the SemEval conference. The main focus of these tasks is the extraction of time-related information from hospital records, which is divided into three subtasks: medical temporal phrase extraction, medical event extraction, and the extraction of temporal and event relationships. Among these, the recognition of medical temporal phrases is the first and most critical step, as it forms the foundation for identifying the relationships between medical temporal information and events, serving as the core of the entire task.

II. RELATED WORK AND METHODOLOGY

Recent advancements in deep learning, particularly in the domains of natural language processing (NLP) and computer vision, have significantly influenced the approach to processing and analyzing clinical data. The intersection of convolutional neural networks (CNNs) and recurrent neural networks (RNNs), including their bidirectional variants (BLSTMs), has facilitated novel methodologies for handling complex data structures such as medical texts and images.

Dai et al. (2023)[4] explored an LSTM and attention-based model to mitigate unintended biases in toxicity detection, which demonstrates the utility of LSTM architectures in capturing the sequence-dependent nature of text for NLP tasks. Their approach is particularly relevant as it showcases how attention mechanisms can enhance model performance by focusing on specific parts of the input sequence, a concept that can be adapted to temporal expression recognition in clinical texts where identifying the relevant temporal phrases is crucial for accurate analysis. Similarly, Xiao et al. [5] leveraged CNNs for classifying cancer cytopathology images, focusing on breast cancer. Their work underscores the effectiveness of CNNs in extracting intricate patterns and features from medical images, which parallels the use of CNNs in our proposed architecture to extract morphological and orthographic features from clinical texts. Li et al. [6] further elaborated on the use of CNNs in image classification and semantic segmentation, which could be insightful for our research, particularly in enhancing the semantic understanding of complex medical terminologies through the CNN component of our architecture.

Machine learning approaches, particularly linear statistical models, are predominantly used for extracting medical temporal phrases. In SemEval Task 12, most teams utilized such methods. The UTHealth team [7] combined Hidden Markov Models (HMM) and Support Vector Machines (SVM) for sequence labeling, achieving the highest F-score. LIMSI [8] and GU IR [9] employed Conditional Random Fields (CRF) based sequence labeling algorithms, while ULISBOA [10] used SVM-based sequence labelers. These teams extensively utilized syntactic, lexical, and domain-specific features, with UTHealth also incorporating results from the rule-based SUTime system. Feature engineering is time-consuming and may introduce inaccuracies, impacting recognition performance.

Xu et al. [11] discussed the application of multimodal generative adversarial networks (GANs) within deep learning frameworks. Although their focus is on a different application, the integration of multiple data modalities in their work can inform future extensions of our model where multimodal data (text and images) from clinical records could be utilized for a comprehensive analysis. Zhao et al. [12] introduced a neural network-based approach for e-commerce webpage recommendation, utilizing semantic mining techniques. This approach to semantic analysis could be adapted to improve the extraction and interpretation of medical temporal phrases by enhancing the model's ability to understand and categorize the semantic contexts of these phrases. Zhang et al.[13] focused on optimizing deep learning algorithms for medical image processing, highlighting the importance of performance evaluation in models designed for the medical field. Their emphasis on optimization and evaluation is crucial as it aligns with our need to continuously refine our architecture to handle the nuances of medical language in clinical texts effectively.

These studies collectively illustrate a broad application of deep learning techniques across different domains and provide a strong foundation for developing sophisticated models capable of handling the specific challenges posed by the extraction of temporal information in medical records. The integration of these insights into our research could potentially lead to a more robust and efficient model, contributing to the broader field of biomedical informatics and clinical decision support systems.

Adding to the aforementioned studies, Yan et al. [14]have made significant strides in applying neural networks for survival prediction across diverse cancer types. Their work, which focuses on leveraging neural networks to predict patient outcomes based on clinical data, offers valuable insights into the potential of deep learning techniques in prognostic medical applications. This reference is particularly pertinent to our research as it exemplifies the capability of neural networks to handle varied and complex medical datasets, an aspect that could be integral for further refining our approach to temporal expression recognition in clinical texts.

III. EXPERIMENTAL DESIGN

A. Model and Algorithm Description

The task of extracting temporal phrases in medical records can be considered a sequence labeling task. Referring to previous successful cases, this involves identifying the type of temporal expression while recognizing it. The BIO scheme is used for result labeling. For instance, "B-Timex-Date" indicates the beginning of a date-type temporal expression, while "I-Timex-Time" denotes a non-beginning part of a time-type temporal expression. Figure 1 shows the neural network architecture based on Long Short-Term Memory (LSTM) units proposed in this paper. It is divided into three layers: the first layer is the word vector representation layer, the second layer is the bi-directional LSTM layer, and the third layer is the label output layer. First, the sentence is split into a sequence of words, which are converted into corresponding word vectors by querying a word vector table. These word vectors are then concatenated with character-level representation vectors of the words trained using a Convolutional Neural Network (CNN). The concatenated word vectors are used as input to the bi-directional LSTM layer. After computation, the layer outputs vectors with contextual semantics of the words. Finally, the Conditional Random Fields (CRF) in the label output layer optimizes the output labels holistically.

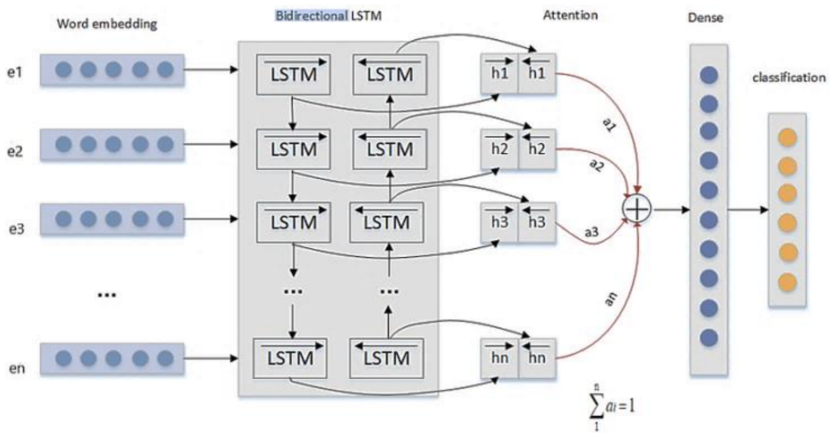

Figure 1: Architecture of BLSTM

B. Word Vectors

The input to the word vector representation layer is individual words, which are output as vector representations. In this research, the word vector representation is composed of two concatenated vectors: one obtained by querying a pre-trained word vector table, and the other computed as a character-level representation vector using a Convolutional Neural Network (CNN). The word vector table can be trained on large-scale, unlabeled datasets using word2vector [15] , or sourced from publicly available domain-specific vector sets, which are trained through unsupervised learning on large corpora and provide excellent generalization capabilities.

The CNN structure used in this study is depicted in Figure 2. For a word *w={a1,a2,...,an}* of length *n*, where a_i is the *i-th* character, *emb*(a_i) represents the character vector of a_i. Assuming the convolutional neural network's window size is *C*=3, the final vector representation of character a_i is the concatenation of the character vectors of itself and its adjacent characters, denoted as $r(a_i) = [emb(a_{i-1}), emb(a_i), emb(a_{i+1})]$ .The convolution operation employs a fixed-size convolution kernel to extract local features from the character vector matrix of the word, and a max-pooling operation is subsequently used to obtain the character-level feature vector representation of the entire word.

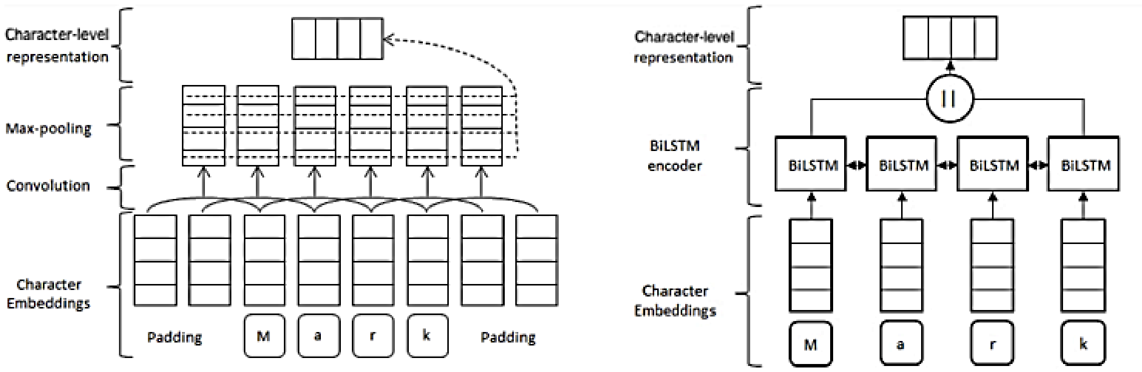

Figure 2 Architecture of CNN

C. Bidirectional Long Short-Term Memory (BLSTM) network

The Bidirectional Long Short-Term Memory (BLSTM) network represents a specialized Recurrent Neural Network (RNN) model designed to mitigate the gradient vanishing issue that arises in traditional RNN models when dealing with extended sequences [16-17]. This model has been effectively employed in a range of natural language processing tasks[18], such as named entity recognition[19], text classification, and medical imaging segmentation[20-21]. While the BLSTM network shares a structural resemblance with the bidirectional RNN, the critical distinction lies in its substitution of hidden units with LSTM units [22]. These LSTM units manage contextual information through the operation of input gates, memory cells, forget gates, and output gates. The structure of the BLSTM can be formally described as

$$f_t = \sigma\left(w_f \cdot [h_{t-1} \oplus x_t] + b_f\right) \tag{1}$$

The output from the bidirectional Long Short-Term Memory (LSTM) layer is derived by concatenating the forward and backward vectors, as described by the following formula:

$$h_t = \tanh\left(W_h \cdot \left(h_t^{(l)} \oplus h_t^{(r)}\right) + b_h\right) \tag{2}$$

D. Label Output Layer

Conditional Random Fields (CRF), introduced by Lafferty et al. [23] in 2001, is an undirected probabilistic graphical model. CRF achieves a globally optimal labeling sequence by taking into account the dependencies between adjacent labels. In recent years, it has demonstrated excellent performance across various sequence labeling tasks. This study employs the CRF algorithm to enhance the output from the BLSTM layer, thereby obtaining the globally optimal label output. For a given sentence:

$$x = \{x_1, x_2, \cdots, x_n\} \tag{3}$$

Define *P* as the scoring result output from the BLSTM layer. *P* is an $n \times k$ matrix, where k is the number of output label categories. Define $P_{i,j}$ as the probability of the *i-th* word in the sentence outputting the *j-th* label. For a predicted sequence $y=\{y_1, y_2, ..., y_n\}$,its score can be defined as Training Parameters.

$$S(x, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i} \tag{4}$$

The training process employs the Adagrad stochastic gradient descent model [24], with a learning rate of 0.03 and a regularization parameter set to 10^{-8}. To reduce overfitting, a dropout with a rate of 0.5 is applied to the input/output sections of the BLSTM layer. The parameters are fine-tuned based on the development set outcomes. This study utilizes the publicly available 200-dimensional PubMed [25] vector set as the initial word vector lookup table, which is derived from a vast corpus of biomedical literature and abstracts. The rationale for selecting this vector set will be detailed in Section 3.2.1. All other vectors are also defined as 200-dimensional. The CNN window size is set to 1, and the output vector length is specified as 20.

IV. EXPERIMENTAL DESIGN

A. Dataset and Evaluation

In this study, we conducted various experiments using the cancer patient pathology reports and clinical record corpora (THYME corpus) annotated by the Mayo Clinic, provided by SemEval-task12 [26]. The dataset is divided into three parts: training set, development set, and test set. The training set contains 283 clinical record documents, the development set and the test set contain 146 and 153 documents respectively, with 3833, 2078, and 1932 time expressions respectively. Incorporating the linked data methodology, as explored by Li et al. [27] in their investigation of creating accessibility linked data based on publicly available datasets, we enhance the structure of our dataset. By adopting RDF (Resource Description Framework) standards and linking our data elements to established ontologies, we ensure that the data from different sources within the THYME corpus are semantically interoperable and accessible. This approach not only facilitates a better understanding of the relationships between data elements but also supports the accessibility of data across different platforms and applications, thereby extending the utility of the THYME corpus beyond its current scope. The application of linked data principles aids in the dynamic integration and retrieval of data, making the dataset more adaptable to various research needs in the field of medical informatics. The evaluation method consistent with the task is used in the experiment. In Table 1, P represents precision, R represents recall, and F1 represents the F-measure.

B. Experimental Results

The representation of word vectors has a significant impact on the results of sequence labeling tasks. This paper conducts comparative experiments using the following three types of word vectors: a) word vectors with random initialization lengths of 100 dimensions, 200 dimensions, and 300 dimensions; b) word vectors trained using the word2vec tool[28] on the THYME corpus provided by the task, with dimensions of 100, 200, and 300; c) 200-dimensional PubMed word vector set trained with large-scale biomedical corpora. The experimental results, as shown in Table 1, indicate that the third type of vector achieves the best generalization results.

Table. 1: Results with different choices of word embeddings

Word Vector	P	R	F1
Random Initialization 100D	81.74	78.3	79.98
Random Initialization 200D	82.34	78.67	80.46
Random Initialization 300D	82.35	78.42	80.34
THYME Corpus Training 100D	81.92	78.83	80.35
THYME Corpus Training 200D	82.21	79.83	81
THYME Corpus Training 300D	82.4	79.74	81.05
Pubmed 200D Vector	83.71	80.02	81.82

C. Layered Testing of Neural Network Architectures

We performed a series of layered tests on the neural network architecture, comparing the outcomes to evaluate the contribution of each module within the model. The results of these experiments are summarized in Table 2. The term "-CRF" denotes the exclusion of the CRF layer from the model, with softmax used for label output instead. The term "-pre-train" signifies that the model does not utilize the pre-trained Pubmed vector set as the initial vectors, opting instead for randomly generated initial word vectors. The term "-CNN" implies the removal of the CNN layer, using vectors directly obtained from the Pubmed vector set as word vectors. The "CNN-BLSTM-CRF" model is the final neural network model chosen for this study.

Table. 2: Test Performance Evaluation

model	P	R	F1
-CNN	80.53	79.81	80.17
-pretrain	82.34	78.67	80.46
-CRF	79.85	78.33	79.08
CNN-BLSTM-CRF	83.71	80.02	81.82

After the aforementioned three sets of experiments, the paper selects the Pubmed vectors with a dimension of 200 and compares the experimental results with those of other outstanding scholars. The results are shown in Table 3.

Table. 3: Results with different teams

Team	P	R	F1
CDE-IIITH	0.604	0.55	0.576
Brundlefly	68.59	41.49	51.69
UFPRSheffield	0.301	79.49	44.69
UTHealth	83.59	75.69	79.49
LIMSI-1	83.99	50.99	63.49
CNN-BLSTM-CRF	83.71	80.02	81.82

This comparative analysis indicates that the neural network model presented in this paper has achieved the best performance to date without employing any manual features.

V. CONCLUSION

This paper has demonstrated a robust approach for extracting temporal phrases from medical records using a deep learning framework, specifically leveraging a deep neural network architecture. By employing a convolutional neural network (CNN), the model excels at capturing intricate morphological features of words. Simultaneously, it harnesses the capabilities of a bidirectional long short-term memory (BLSTM) network to interpret contextual semantic nuances effectively. Furthermore, the Conditional Random Field (CRF) algorithm enhances the precision of the output labels, ensuring the model's robustness in real-world applications. A key advantage of our approach is its independence from manually crafted features and specialized medical knowledge, which simplifies the model's deployment and scalability. Comparative analysis with existing systems demonstrates that our model not only meets but exceeds the performance of the best available systems in this domain. This achievement underscores the efficacy of deep learning methodologies in handling complex pattern recognition tasks within medical texts. The model's applicability extends beyond temporal phrase extraction to other sequence labeling tasks, such as the identification of medical events, which suggests its potential as a versatile tool in medical informatics. Looking forward, the model's adaptability could be further enhanced by integrating multi-task joint learning. For instance, developing a joint neural network that concurrently learns to identify both temporal phrases and medical events could significantly improve the system's accuracy and generalization capabilities. This prospective advancement may lead to a more holistic understanding of patient records, ultimately contributing to more informed and effective healthcare solutions. This paper not only advances the field of medical text analysis but also sets a foundation for future innovations in automated medical record management.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Moharasar, G., & Tu, B. H. (2016). A semi-supervised approach for temporal information extraction from clinical text. In Proceedings of IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (pp. 7-12). Piscataway, NJ: IEEE Press.

[2] Wang, S., Liu, Z., & Peng, B. (2023, December). A Self-training Framework for Automated Medical Report Generation. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 16443-16449).

[3] Gong, Y., Qiu, H., Liu, X., Yang, Y., & Zhu, M. (2024). Research and Application of Deep Learning in Medical Image Reconstruction and Enhancement. Frontiers in Computing and Intelligent Systems, 7(3), 72-76.

[4] Dai, W., Tao, J., Yan, X., Feng, Z., & Chen, J. (2023, November). Addressing Unintended Bias in Toxicity Detection: An LSTM and Attention-Based Approach. In 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 375-379). IEEE.

[5] Xiao, M., Li, Y., Yan, X., Gao, M., & Wang, W. (2024). Convolutional neural network classification of cancer cytopathology images: taking breast cancer as an example. arXiv preprint arXiv:2404.08279.

[6] Li, M., Zhu, Z., Xu, R., Feng, Y., & Xiao, L. (2024). Research on Image Classification And Semantic Segmentation Model Based on Convolutional Neural Network. Journal of Computing and Electronic Information Management, 12(3), 94-100.

[7] Lee, H. J., Xu, H., Wang, J., et al. (2016). UTHealth at SemEval-2016 task 12: An end-to-end system for temporal information extraction from clinical notes. In Proceedings of the 10th International Workshop on Semantic Evaluation (pp. 1292-1297).

[8] Grouin, C., & Moriceau, V. (2016). LIMSI at SemEval-2016 task 12: Machine learning and temporal information to identify clinical events and time expressions. In Proceedings of the 10th International Workshop on Semantic Evaluation (pp. 1225-1230).

[9] Cohan, A., Meurer, K., & Goharian, N. (2016). GUIR at SemEval-2016 task 12: Temporal information processing for clinical narratives. In Proceedings of the 10th International Workshop on Semantic Evaluation (pp. 1248-1255).

[10] Barros, M., Lamurias, A., Figueiro, G., et al. (2016). ULISBOA at SemEval-2016 task 12: Extraction of temporal expressions, clinical events and relations using IBEnt. In Proceedings of the 10th International Workshop on Semantic Evaluation.

[11] Xu, R., Yang, Y., Qiu, H., Liu, X., & Zhang, J. (2024). Research on Multimodal Generative Adversarial Networks in the Framework of Deep Learning. Journal of Computing and Electronic Information Management, 12(3), 84-88.

[12] Zhao, W., Liu, X., Xu, R., Xiao, L., & Li, M. (2024). E-commerce Webpage Recommendation Scheme Base on Semantic Mining and Neural Networks. Journal of Theory and Practice of Engineering Science, 4(03), 207–215. https://doi.org/10.53469/jtpes.2024.04(03).20

[13] Zhang, J., Xiao, L., Zhang, Y., Lai, J., & Yang, Y. (2024). Optimization and Performance Evaluation of Deep Learning Algorithm in Medical Image Processing. Frontiers in Computing and Intelligent Systems, 7(3), 67-71.

[14] Yan, X., Wang, W., Xiao, M., Li, Y., & Gao, M. (2024). Survival Prediction Across Diverse Cancer Types Using Neural Networks. arXiv preprint arXiv:2404.08713.

[15] Mikolov, T., Sutskever, I., Chen, K., et al. (2013, October 16). Distributed representations of words and phrases and their compositionality. arXiv. https://arxiv.org/abs/1310.4546

[16] Li, Z., Yu, H., Xu, J., Liu, J., & Mo, Y. (2023). Stock market analysis and prediction using LSTM: A case study on technology stocks. Innovations in Applied Engineering and Technology, 1-6.

[17] Yao, J., Wu, T., & Zhang, X. (2023). Improving depth gradient continuity in transformers: A comparative study on monocular depth estimation with cnn. arXiv preprint arXiv:2308.08333.

[18] Lu, S., Liu, Z., Liu, T., & Zhou, W. (2023). Scaling-up medical vision-and-language representation learning with federated learning. Engineering Applications of Artificial Intelligence, 126, 107037.

[19] Liu, Z., & Song, J. (2021, November). Comparison of Tree-based Feature Selection Algorithms on Biological Omics Dataset. In Proceedings of the 5th International Conference on Advances in Artificial Intelligence (pp. 165-169).

[20] Wang, Q., Schindler, S. E., Chen, G., Mckay, N. S., McCullough, A., Flores, S., ... & Benzinger, T. L. (2024). Investigating White Matter Neuroinflammation in Alzheimer Disease Using Diffusion-Based Neuroinflammation Imaging. Neurology, 102(4), e208013.

[21] Zhao, B., Cao, Z., & Wang, S. (2017). Lung vessel segmentation based on random forests. Electronics Letters, 53(4), 220-222.

[22] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5-6), 602-610.

[23] Lafferty, J. D., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning (pp. 282-289).

[24] Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12(7), 257-269.

[25] White, J. (2020). PubMed 2.0. Medical reference services quarterly, 39(4), 382-387.

[26] Bethard, S., Savova, G., Chen, W. T., Derczynski, L., Pustejovsky, J., & Verhagen, M. (2016, June). Semeval-2016 task 12: Clinical tempeval. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 1052-1062).

[27] Li, Y., Yan, X., Xiao, M., Wang, W., & Zhang, F. (2024). Investigation of Creating Accessibility Linked Data Based on Publicly Available Accessibility Datasets. In Proceedings of the 2023 13th International Conference on Communication and Network Security (pp. 77–81). Association for Computing Machinery.

[28] Church, K. W. (2017). Word2Vec. Natural Language Engineering, 23(1), 155-162.