Enhancing Radiological Biomedical Natural Language Processing Tasks with Radiology-Specific Word Encodings: A Comparative Analysis of Word Embeddings Sources

Kamlesh Kumar Yadav [1*], Dhablia Dharmesh Kirit [1]

Abstract

Background: Machine Learning (ML)-based Biomedical Natural Language Processing (BNLP) techniques have garnered attention in radiology. However, these models typically depend on Word Encodings (WE) trained on generic datasets, as radiology-specific word libraries are limited. Objective: This study aimed to investigate the potential of radiography as a comprehensive database for generating Radiology-Specific Word Encodings (RSWE), enhancing the efficiency of BNLP tasks, especially in processing radiological texts. Methods: A systematic evaluation was conducted using WE derived from four databases: medical records, biomedical journals, Wikipedia, and news sources. Unstructured Electronic Medical Record (EMR) data from the Mayo Clinic and PubMed Central publications were used to train WE for medical-specific sources, while GloVe and Google News represented publicly available pre-trained WE for generic sources. Analytical evaluation employed medical keywords in three categories (illness, symptoms, drugs), and a 2-D graphical plot was created for 380 medical words. Numerical evaluation consisted of internal and external assessments. Results: Findings revealed that RSWE derived from EMR and PubMed Central outperformed generic WE, better capturing medical word meanings and identifying medically essential terms, aligning more closely with expert assessments. Conclusion: The study demonstrates the value of radiography as a radiology-specific resource for generating RSWE, with promising implications for improving BNLP in radiology.

Keywords: Radiology-specific word encodings (RSWE), Medical natural language processing (BNLP), Word embeddings (WE), Radiopaedia dataset, Electronic health records (EHR)

Significance | This study demonstrates Radiology-specific word encodings potential in improving the performance of radiology-focused BNLP tasks using diverse word embeddings.

*Correspondence. Kamlesh Kumar Yadav , Department of CS & IT, Kalinga University, Raipur, India
E-mail: ku.kamleshkumaryadav@kalingauniversity.ac.in

Editor Surendar Aravindhan, Ph.D., And accepted by the Editorial Board Sep 04, 2024 (received for review Jul 15, 2024)

1. Introduction

NLP approaches include various technologies used to extract valuable information from unstructured texts. It is crucial in radiology because it facilitates the conversion of information contained in medical images into a textual format, namely the radiology record (Pudasaini et al., 2022). Radiology records represent a large medical data collection that annotates the corresponding medical images. However, their unorganized and unrestricted language style sometimes poses challenges when transforming them into a computer-friendly structure. Language specifications such as RadLex and CT were developed to provide a more organized framework for textual information (Langlotz, 2006). The extensive analysis and organization of this large volume

Author Affiliation.
[1] Department of CS & IT, Kalinga University, Raipur, India

Please cite this article.
Kamlesh Kumar Yadav, Dhablia Dharmesh Kirit (2024). Enhancing Radiological Biomedical Natural Language Processing Tasks with Radiology-Specific Word Encodings: A Comparative Analysis of Word Embeddings Sources, Journal of Angiotherapy, 8(9), 1-6, 9873

of unstructured information have created significant opportunities for BNLP applications in radiology.

Historically, NLP systems in radiography were developed using syntax-based frameworks to organize unstructured narrative material (Nobel et al., 2020). However, rule-based techniques require substantial time and resources to develop, as they

they necessitate the creation of domain-specific tasks (Richardson, 2007). These approaches are also inflexible in accommodating variations in human and institutional behaviors (Liu et al., 2018). In recent years, ML techniques have gained popularity due to their ability to operate without labor-intensive manual rule creation (Sorin et al., 2020). CNN and RNN methodologies have been extensively applied to text-categorization tasks, such as identifying lung embolism or bronchitis in radiology reports (Banerjee et al., 2019).

A crucial stage in several ML-based NLP models involves converting words into vector-based representations, known as WE (Shetty & Mahale, 2023). GloVe and word2vec are widely used techniques for generating WE and have been applied in many BNLP models (Johnson et al., 2024). Studies have shown that initializing WE with domain-specific data leads to improved outcomes (Neelima et al., 2024). However, radiology NLP models often rely on WE trained on broad corpora or their training sets (Yuan et al., 2019). It is hypothesized that WE trained on a radiology-specific dataset may effectively capture the underlying medical ontology, improving model performance in BNLP tasks.

An increasing body of research explores using text extraction and WE algorithms for analyzing EMR data. For instance, Miotto et al. (2016) aimed to leverage patient data in clinical notes using DL techniques (Sindhusaranya et al., 2023). However, the study showed that plain text could not be effectively utilized as meaningful attributes in these algorithms. Instead, the authors used medical ontologies to extract patient phenotype terms and employed WE methods like word2vec to transform text into a DL-compatible format. Other researchers have developed similar approaches, encoding diagnoses, procedures, and prescriptions into vectors to enhance prediction accuracy (Pham et al., 2016).

Furthermore, Liu et al. (2018) trained WE for clinical abbreviation expansion, achieving accuracy comparable to skilled humans. Another study developed an automated system for removing personal medical data using WE and RNN models (Yang et al., 2019). Despite the progress made in analyzing clinical descriptions in EMRs using WE, little research has focused on applying WE to structured medical codes. One study used a skip-gram model to generate vector representations of medical codes and introduced a prediction technique, Patient-Diagnostic Projection Similarity (PDPS), which outperformed logistic regression in prediction tasks (Ma et al., 2020).

Lastly, Li et al. (2019) developed an enhanced version of Word2Vec and PDPS, which allows data integration from multiple sources without sharing individual-level information. This study also explored the potential of radiography as a radiology database for developing RSWE, which could improve BNLP task performance in radiological texts.

2. Materials and Methods

The primary text from Radiopaedia entries in 2021 was extracted, together with their corresponding "System" label(s), using Python (Richardson, 2007). No articles were omitted. The text from Radiopaedia instances was excluded. The text underwent preprocessing using word tokenization. Commonly used English stop-words such as "the" punctuation marks, capitalizing, and special characters have been eliminated. Tokens were created by splitting words that included punctuation or special characters at the location where a character was deleted. The WE used in this study were acquired from Stanford NLP, a corpus containing 5 billion tokens. Figure 1 provides a summary of the features of the article database.

2.1 Analytical and Numerical assessment

We systematically chose medical terms from three distinct medical WE categories: disease, symptom, and medication. WE derived from four distinct datasets were used to calculate the five most comparable terms to each chosen medical term based on cosine likeness. We thoroughly examined the semantic relationship between the desired term and the most associated words. Let $w1\ and\ w2$ be two distinct words. The relationship between $w1\ and\ w2$ can be described as:

$$Relationship(w1, w2) = \frac{\partial 1.\partial 2}{\|\partial 1\|\|\partial 2\|} \quad (1)$$

Where $\partial 1\ and\ \partial 2$ are the vector depictions for $w1\ and\ w2$, respectively. If the desired word is a medical phrase p1 composed of many words, denoted as $p1 = w1, w2, w3, \dots \dots wn$, the relationship function is modified accordingly.

$$Relationship(p1, w2) = \frac{\varphi 1.\varphi 2}{\|\varphi 1\|\|\varphi 2\|} \quad (2)$$

where $\varphi 1 = (\frac{1}{n}) \sum_i^n \partial i \quad (3)$

Equation (3) is the depiction for $p1$ in the WE space. We ranked the terms in the vocabulary by assessing their resemblance to the target word. From this ranking, we selected the five words that were rated highest.

To demonstrate various facets of medical notions represented by WE learned from distinct databases, we obtained 380 medical phrases from the collection. We subsequently displayed the WE for these medical phrases in a 2-D plot using t-distributed WE. Figure 2 displays groups of medical phrases in the WE. Figure 2a illustrates

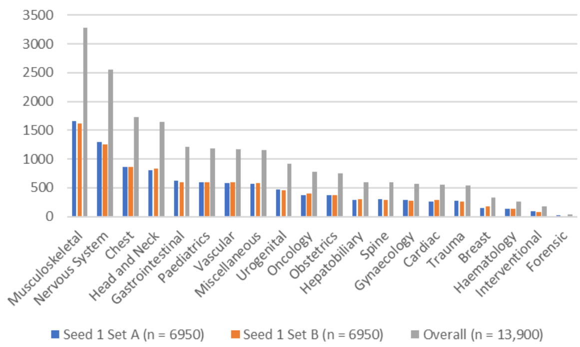

Figure 1 Summary of the features of the article database

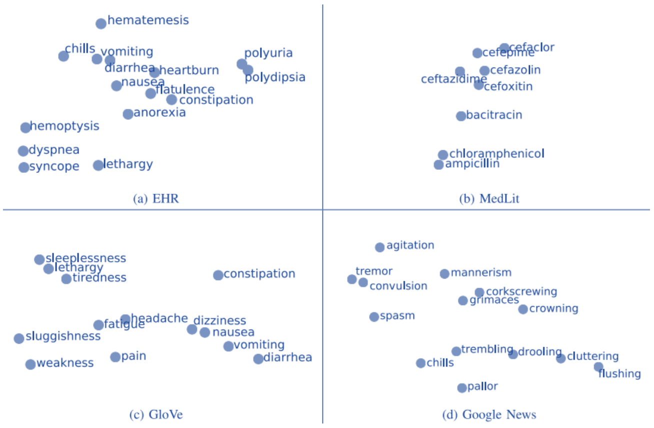

Figure 2 Groups of medical phrases in the WE

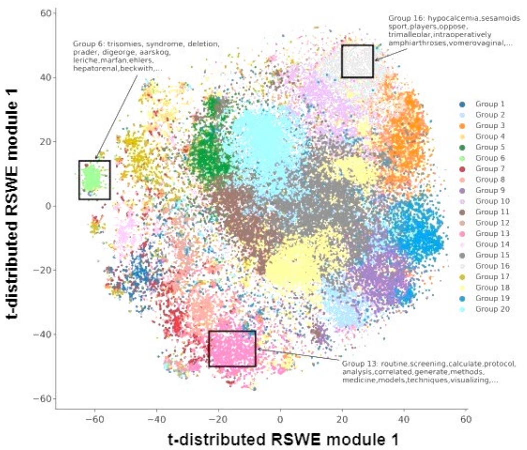

Figure 3 A 2D t-distributed RSWE map, representing the 50-three-dimensional Radiopaedia WE.

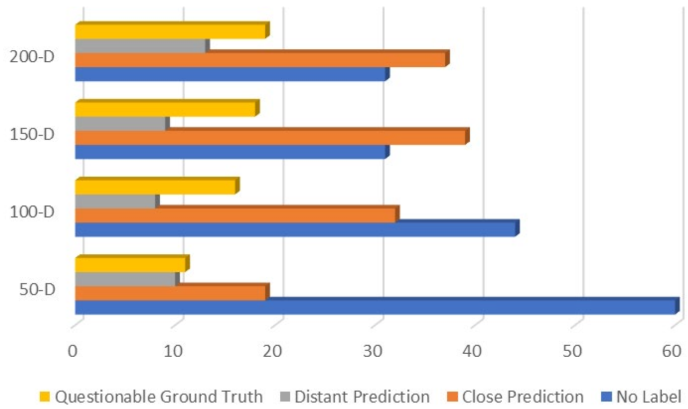

Figure 4 Categorization of errors in article labeling for Radiopaedia

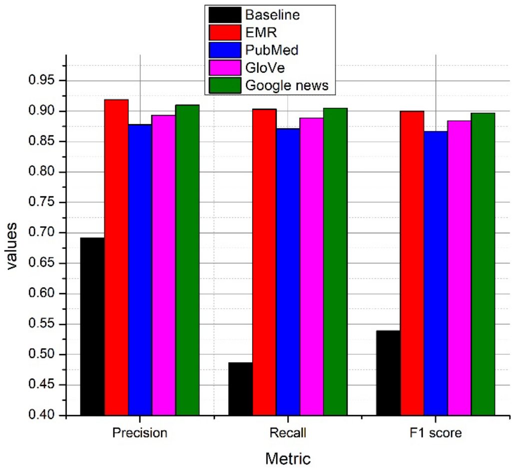

Figure 5 Systematic evaluation of RSWE derived from four distinct databases

a grouping of signs and symptoms, including burning sensations, nausea, and vomiting, derived from the WE learned on EMR. Figure 2b displays a grouping of antibiotic drugs, including bacitracin and chloramphenicol, as determined by PubMed WE. Figure 2c and 2d show groups of symptoms derived from the GloVe and Google News WE, respectively. As we failed to employ any grouping technique, these groupings were detected based on intuition from the 2-D graphic. A comprehensive depiction of the whole collection of 380 medical phrases is shown, using WE derived from four distinct databases.

An analysis of the k-means clustering of the WE demonstrated that GloVe successfully grouped words with similar meanings together. Figure 3 displays the RSWE map with a t-distribution and provides a selection of sample words from certain locations. Specific clusters, like the Group 17 cluster, seem to have less clearly defined borders for their points. This might be because a single t-distributed RSWE map is ineffective in visualizing non-metric connections. The t-distributed RSWE graphic demonstrates that the GloVe model successfully captured the semantic relationships between words by properly clustering related keyword groupings in close proximity.

Figure 4 depicts the categorization of errors in article labeling for Radiopaedia. Figure 4 illustrates the range of error types across the 101 randomly chosen erroneously tagged articles for every model. A chi-square test was conducted to determine the degree of homogeneity for every set of models with identical parameters. Our research has shown that articles from Radiopaedia.org may serve as a radiology database for training WE (Radiopaedia). These WE possess a robust internal comprehension of medical terminology and enhance the efficiency of a model on a radiology-focused BNLP task. The significant improvement in similarity completion effectiveness, seen when comparing the top accuracies of Radiopaedia-trained RSWE, suggests that the outputs of these WE are more closely aligned with the actual truth. The use of Radiopaedia RSWE in models resulted in a considerable improvement in matching accuracy. Specifically, there was a 6-12% gain at 50-D and 100-D and around 3.5% gain at 150-D and 200-D, which were numerically significant.

Figure 5 illustrates the systematic evaluation of RSWE derived from four distinct databases. The EHR-RSWE is better than all the others because it gets the best scores in every category, with an F1 score of 0.900, a Precision score of 0.919, and a Recall score of 0.903. Based on these results, the WE made from EHRs works best for radiology-specific tasks compared to the other options that were looked at. The Google News-RSWE also does well, especially in Recall (0.905), which shows that it can find related words correctly. The performance of MedLit and GloVe WE is about the same as that of EHR and Google News WE, but it is a little worse. This suggests that while they work, they might be unable to capture the complexities of radiology like EHR encodings thoroughly do.

3. Conclusion

This research demonstrates the application of radiography as a comprehensive radiology library to develop radiology-specific word encodings (RSWE). Through a comparative analysis of four Word embedding (WE) sources—Electronic Health Records (EHR), PubMed Central, GloVe, and Google News—the study identified that RSWE derived from EHR data significantly outperformed others in radiology-specific BNLP tasks. The EHR-RSWE demonstrated the highest scores across F1, Precision, and Recall metrics, suggesting its superior capability in capturing the intricate nuances of medical terminology. Additionally, the analysis showed that embeddings trained on domain-specific corpora like EHR and PubMed Central offer higher prediction accuracy and better clustering of medical terms compared to general-purpose models like GloVe and Google News. The results indicate that utilizing RSWE from radiology-specific datasets such as Radiopaedia can substantially enhance the performance of machine learning models in medical NLP tasks, thus improving the accuracy and efficiency of clinical decision-making in radiology.

Author contributions

K.K.Y. led the conceptualization, study design, and manuscript drafting. D.D.K. contributed to data analysis, interpretation of results, and manuscript review. Both authors have read and approved the final version of the manuscript.

Acknowledgment

Author was grateful to their department.

Competing financial interests

The authors have no conflict of interest.

References

Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., Moradzadeh, N., ... & Lungren, M. P. (2019). Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. Artificial Intelligence in Medicine, 97, 79-88.

Johnson, S. J., Murty, M. R., & Navakanth, I. (2024). A detailed review on word embedding techniques with emphasis on word2vec. Multimedia Tools and Applications, 83(13), 37979-38007.

Langlotz, C. P. (2006). RadLex: A new method for indexing online educational materials. Radiographics, 26(6), 1595-1597.

Li, Z., Roberts, K., Jiang, X., & Long, Q. (2019). Distributed learning from multiple EHR databases: contextual embedding models for medical events. Journal of Biomedical Informatics, 92, 103138.

Liu, Y., Ge, T., Mathews, K. S., Ji, H., & McGuinness, D. L. (2018). Exploiting task-oriented resources to learn word embeddings for clinical abbreviation expansion. arXiv preprint arXiv:1804.04225.

Ma, L., Zhang, C., Wang, Y., Ruan, W., Wang, J., Tang, W., ... & Gao, J. (2020, April). Concare: Personalized clinical feature embedding via capturing the healthcare context. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 01, pp. 833-840).

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Scientific Reports, 6(1), 1-10.

Neelima, S., Govindaraj, M., Subramani, K., ALkhayyat, A., & Mohan, C. (2024). Factors influencing data utilization and performance of health management information systems: A case study. Indian Journal of Information Sources and Services, 14(2), 146-152. https://doi.org/10.51983/ijiss-2024.14.2.21

Nobel, J. M., Puts, S., Bakers, F. C., Robben, S. G., & Dekker, A. L. (2020). Natural language processing in Dutch free text radiology reports: challenges in a small language area staging pulmonary oncology. Journal of Digital Imaging, 33, 1002-1008.

Pudasaini, S., Shakya, S., Lamichhane, S., Adhikari, S., Tamang, A., & Adhikari, S. (2022). Application of NLP for information extraction from unstructured documents. In Expert Clouds and Applications: Proceedings of ICOECA 2021 (pp. 695-704). Springer Singapore.

Radiopaedia.org, the wiki-based collaborative Radiology resource. Radiopaedia. https://radiopaedia.org/?lang=us. Accessed June 1, 2020.

Richardson, L. (2007). Beautiful soup documentation.

Shetty, S., & Mahale, A. (2023). Multimodal medical tensor fusion network-based DL framework for abnormality prediction from the radiology CXRs and clinical text reports. Multimedia Tools and Applications, 82(28), 44431-44478.

Sindhusaranya, B., Yamini, R., Manimekalai Dr, M. A. P., & Geetha Dr, K. (2023). Federated learning and blockchain-enabled privacy-preserving healthcare 5.0 system: A comprehensive approach to fraud prevention and security in IoMT. Journal of Internet Services and Information Security, 13(4), 199-209.

Sorin, V., Barash, Y., Konen, E., & Klang, E. (2020). Deep learning for natural language processing in radiology—fundamentals and a systematic review. Journal of the American College of Radiology, 17(5), 639-648.

Yang, X., Lyu, T., Li, Q., Lee, C. Y., Bian, J., Hogan, W. R., & Wu, Y. (2019). A study of deep learning methods for de-identification of clinical notes in cross-institute settings. BMC Medical Informatics and Decision Making, 19, 1-9.

Yuan, J., Zhu, H., & Tahmasebi, A. (2019). Classification of pulmonary nodular findings based on characterization of change using radiology reports. AMIA Summits on Translational Science Proceedings, 2019, 285.