<u>Original Paper</u>

Disambiguating Clinical Abbreviations by One-to-All Classification: Algorithm Development and Validation Study

Sheng-Feng Sung[1,2], MD, PhD; Ya-Han Hu[3], PhD; Chong-Yan Chen[3], MBA

[1]Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan
[2]Department of Nursing, Fooyin University, Kaohsiung, Taiwan
[3]Department of Information Management, National Central University, Taoyuan City, Taiwan

Corresponding Author:

Ya-Han Hu, PhD
Department of Information Management
National Central University
300 Zhongda Rd, Zhongli District
Taoyuan City, 32001
Taiwan
Phone: 886 34227151 ext 66560
Email: yhhu@mgt.ncu.edu.tw

Abstract

Background: Electronic medical records store extensive patient data and serve as a comprehensive repository, including textual medical records like surgical and imaging reports. Their utility in clinical decision support systems is substantial, but the widespread use of ambiguous and unstandardized abbreviations in clinical documents poses challenges for natural language processing in clinical decision support systems. Efficient abbreviation disambiguation methods are needed for effective information extraction.

Objective: This study aims to enhance the one-to-all (OTA) framework for clinical abbreviation expansion, which uses a single model to predict multiple abbreviation meanings. The objective is to improve OTA by developing context-candidate pairs and optimizing word embeddings in Bidirectional Encoder Representations From Transformers (BERT), evaluating the model's efficacy in expanding clinical abbreviations using real data.

Methods: Three datasets were used: Medical Subject Headings Word Sense Disambiguation, University of Minnesota, and Chia-Yi Christian Hospital from Ditmanson Medical Foundation Chia-Yi Christian Hospital. Texts containing polysemous abbreviations were preprocessed and formatted for BERT. The study involved fine-tuning pretrained models, ClinicalBERT and BlueBERT, generating dataset pairs for training and testing based on Huang et al's method.

Results: BlueBERT achieved macro- and microaccuracies of 95.41% and 95.16%, respectively, on the Medical Subject Headings Word Sense Disambiguation dataset. It improved macroaccuracy by 0.54%-1.53% compared to two baselines, long short-term memory and deepBioWSD with random embedding. On the University of Minnesota dataset, BlueBERT recorded macro- and microaccuracies of 98.40% and 98.22%, respectively. Against the baselines of Word2Vec + support vector machine and BioWordVec + support vector machine, BlueBERT demonstrated a macroaccuracy improvement of 2.61%-4.13%.

Conclusions: This research preliminarily validated the effectiveness of the OTA method for abbreviation disambiguation in medical texts, demonstrating the potential to enhance both clinical staff efficiency and research effectiveness.

Introduction

The advent of electronic medical records (EMRs) has revolutionized data management in medical institutions by enabling the storage and collection of extensive patient data. EMRs integrate records and reports from various hospital departments, documenting diverse patient conditions and providing a comprehensive repository of information,

including previous laboratory and examination reports, hospitalization and surgical procedure records, and medication histories [1-3]. EMRs contain two types of data: structured, such as physiological measurements, laboratory results, diagnostic and drug codes, and assessment scales, and unstructured, primarily consisting of textual medical records like surgical and imaging reports, pathology reports, and discharge summaries [4-10].

Recent studies have leveraged natural language processing (NLP) tools, including MetaMap, MedLEE, and Clinical Text Analysis and Knowledge Extraction System (cTAKES), to extract valuable patient information from EMRs' clinical text [11-14]. These applications range from identifying specific medical concepts to complex analyses, such as discerning relationships between medical conditions or predicting patient outcomes and disease progression [10,15-19]. However, the prevalent use of abbreviations in clinical documents poses significant challenges for NLP in clinical decision support systems, as abbreviations often have multiple meanings depending on their context, and unstandardized or local abbreviations further complicate text interpretation [20,21]. This ambiguity impedes the extraction of meaningful information, affecting clinical decision support system performance and highlighting the need for effective methods for abbreviation disambiguation in clinical NLP applications.

Abbreviation disambiguation in NLP involves identifying the correct expansion of an abbreviation based on its context [22,23]. In this process, one-to-one (OTO) and one-to-all (OTA) approaches are two distinct strategies for resolving the meaning of abbreviations [24]. The OTO approach involves training a separate machine learning model for each specific abbreviation, learning its unique patterns and contextual cues to disambiguate its meaning. In contrast, the OTA approach uses a single machine learning model trained to disambiguate all abbreviations across various contexts.

The OTA approach in abbreviation disambiguation offers several advantages over the OTO approach. OTA is easier to scale, requiring the maintenance and updating of only a single model, whereas OTO necessitates multiple models for each abbreviation, making it less scalable. OTA is more efficient in terms of computational resources and ensures a uniform disambiguation approach, reducing inconsistencies. Additionally, OTA simplifies model management, streamlining changes and improvements. By learning general patterns and contextual cues applicable to various abbreviations, OTA enhances overall context understanding, making it suitable for applications with diverse abbreviation needs. This flexibility makes OTA particularly useful in the biomedical domain, where abbreviations can have varied meanings in different contexts.

This study aims to enhance the application of the OTA abbreviation disambiguation framework for clinical abbreviation expansion. We propose constructing an OTA

disambiguation model by creating context-candidate pairs and refining word embeddings using Bidirectional Encoder Representations From Transformers (BERT) [25]. The model's effectiveness was assessed based on its predictive performance on real clinical data for the task of clinical abbreviation expansion.

Methods

Data

This study conducted experimental evaluations using 3 datasets: 2 publicly available datasets and 1 independently collected from a regional hospital in Taiwan. The first dataset, the Medical Subject Headings Word Sense Disambiguation (MSH WSD) dataset, was extracted from MEDLINE abstracts [26]. The MSH WSD dataset comprises 203 polysemous words and is divided into three sections: abbreviation set, term set, and term/abbreviation set. The abbreviation set, containing 106 ambiguous acronyms, was selected as one of our investigated datasets. The second dataset, originating from the University of Minnesota (UMN), comprises deidentified clinical text sourced from the university's hospitals [27]. The UMN dataset includes 440 frequently used abbreviations and acronyms, carefully selected from a pool of 352,267 dictated clinical notes. These two datasets are valuable resources for both NLP and medical informatics, particularly for disambiguation tasks within the health care domain [20,28,29].

Lastly, the Chia-Yi Christian Hospital (CYCH) dataset aggregates present illness data from patients at the Neurology Department of Ditmanson Medical Foundation Chia-Yi Christian Hospital. Abbreviation disambiguation results were validated by a neurologist. We narrowed the scope of abbreviations for evaluation and asked the doctor to mark the answers in advance. Specifically, we selected five frequently appearing abbreviations—ER, DM, CVA, PM, and PA—from both the UMN and CYCH datasets. We verified that the correct interpretations of abbreviations in the CYCH clinical documents matched the candidate sets for the UMN abbreviations. Due to manpower constraints, we limited our extraction to the first 1000 abbreviations for annotation by a physician. After removing one erroneous data entry and two initially overlooked abbreviations, we had a total of 998 sentences that included the five selected abbreviations. All 3 datasets were preprocessed and organized into the same format for subsequent model construction and evaluation. As shown in Table 1, the text "...He is status post a BK amputation on the right side and..." is partitioned into three parts: left, right, and target. Left denotes the text to the left of the target abbreviation, right represents the text to the right, and target indicates the target abbreviation. The remaining two fields include the correct expansion word for the target abbreviation (label) and the collection of all incorrect candidate expansion words (negs).

Table 1. Data schema after data preprocessing and an example sample text.

Field	Description	Example
Index	Document ID	1
Target	The target abbreviation	BK
Left	The text to the left of the target abbreviation	...He is status post a
Right	The text to the right of the target abbreviation	amputation on the right side and...
Label	The correct expansion word for the target abbreviation	below knee
Neg	The collection of all incorrect candidate expansion words. If there are multiple, separate them with commas.	BK(virus)

Ethical Considerations

The study protocol received formal approval from the Ditmanson Medical Foundation Chia-Yi Christian Hospital Institutional Review Board (2022074). Patient identifiers were replaced by a unique study identification number to ensure confidentiality. Informed consent was thus exempted.

The Proposed Framework

Figure 1 illustrates the proposed framework. We begin by retrieving text containing polysemous abbreviations from the 3 investigated datasets. The polysemous abbreviations are kept in their original form and marked accordingly.

Subsequent steps involve common text preprocessing techniques, such as converting text to lowercase and removing certain special symbols. The preprocessed text is then adjusted to meet the input format required by BERT. Finally, the processed text is divided into training and testing sets. Three existing pretrained BERT-based models, including BERT-base-uncased [25], ClinicalBERT [30], and BlueBERT [31], are chosen and fine-tuned using these datasets, and prediction results are subsequently generated for evaluation. BERT-base-uncased is specifically used due to the common inconsistencies in capitalization within clinical texts, where lowercase letters are frequently used, even at the beginning of sentences or in abbreviations.

Figure 1. Research framework. BERT: Bidirectional Encoder Representations From Transformers; CYCH: Chia-Yi Christian Hospital; MSH WSD: Medical Subject Headings Word Sense Disambiguation; UMN: University of Minnesota.
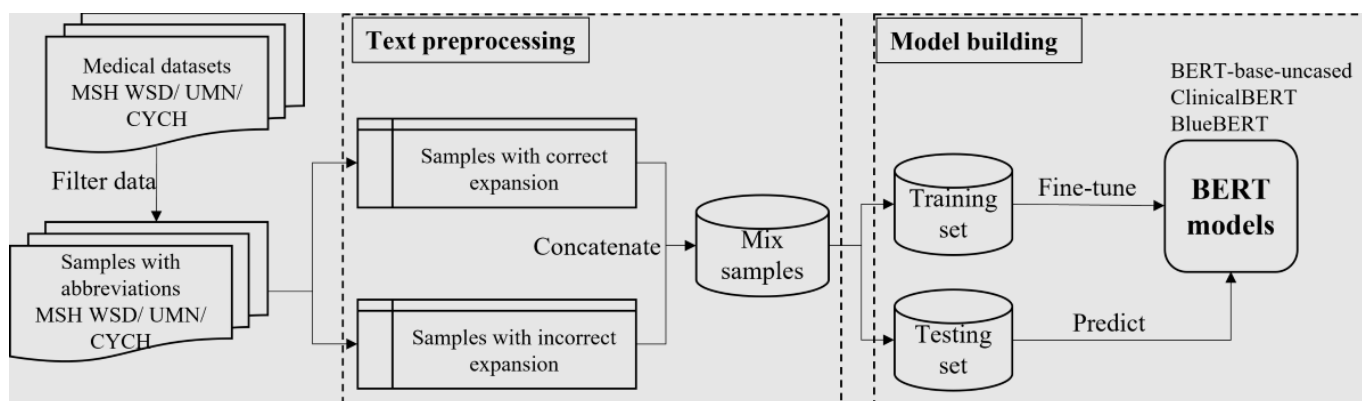
Text Preprocessing

This study converts the text into the context-candidate pair format and adjusts the word embedding values for BERT input. Specifically, we apply GlossBERT [32] to train our model using the samples consisting of abbreviations and all of their candidate expansions. If an abbreviation has n candidate expansions, with only one correct answer, we produce n samples. This includes one sample marked as the correct expansion (indicated as 1) and $n - 1$ samples marked as incorrect expansions (indicated as 0).

Before training, we use BERT's tokenizer to convert text into WordPieces, breaking words such as "amputation" into ['amp', '##utation']. Special tokens are then added: [CLS] at the start, [SEP] to separate sentences or differentiate sections, and [PAD] to equalize sequence lengths for batch

processing. For instance, when processing the sentence "He is status post a BK amputation..." with "BK" having expansions "BK(virus)" and "below knee," we generate two sequences: "[CLS] He is status post a BK amputation... [SEP] BK(virus) [SEP], 0" and "[CLS] He is status post a BK amputation... [SEP] below knee [SEP], 1."

Due to BERT's token limit of 512, sequences exceeding this are truncated. We manage sequence lengths by first converting text into WordPieces and adding necessary tokens. If the combined length of a sequence and its expansions exceeds BERT's limit, we employ a first in, first out (FIFO) strategy to ensure compliance with the token restriction.

BERT Tuning

Due to the limited dataset size, retraining a full BERT encoder was not feasible for this study. Instead, we

fine-tuned existing pretrained models to assess our abbreviation disambiguation method. We selected two health care–related models, ClinicalBERT and BlueBERT, along with a generic BERT-base-uncased model as a baseline.

We adapted these models by adding a fully connected output layer. This layer consists of two linear layers and a rectified linear unit activation function, simplified as:

$$y = f\left(\sum_{i=1}^{n} w_k x_i + b_k\right) \tag{1}$$

where w_k represents the weights applied to inputs x_i, and b_k is the bias term.

The output layer's parameters are set as (50, 2), reflecting the size of the output from the previous layer and the number of classes (1 or 0). During prediction, the model calculates probabilities for each class. We focus primarily on the accuracy of the predictions for class 1, applying a softmax operation to enhance decision-making based on class 1's probability scores. This process optimizes our approach to evaluating the effectiveness of the trained models in context-sensitive disambiguation tasks.

Experimental Setup and Performance Measure

In our experimental evaluation, we aim to compare our proposed approach with several representative methods from prior studies on abbreviation disambiguation, focusing on model adaptability and performance across various datasets. The structure of our study is divided into two main parts.

Experiment 1 assesses the prediction performance of abbreviation expansion using both our proposed approach and baseline models (both OTO and OTA). We utilized two public datasets: MSH WSD and UMN. For MSH WSD, the OTO baselines included k-nearest neighbors [33], naive Bayes [26], and long short-term memory (LSTM) [34] models. For the UMN dataset, we referred to Wu et al [35] who used a combination of Word2Vec + support vector machine (SVM) as the OTO baseline. We further adapted this approach by substituting the original Word2Vec model with BioWordVec [36] (BioWordVec + SVM), which offers biomedical word embeddings via fastText, to better suit our study's focus on clinical data. Each clinical note was represented as a 200-dimensional vector. For the OTA baseline models, we implemented non–sense-based methods using BERT/XLNet, as described by Kim et al [37], which include deepBioWSD_random embeddings and deepBioWSD_pretrained sense embeddings. Additionally, we employed sense-based methods using bidirectional LSTM, outlined by Pesaranghader et al [38], specifically masked language modeling and permutation language modeling.

Experiment 2 evaluates the prediction performance of abbreviation disambiguation within the CYCH dataset, aiming to address abbreviation ambiguity in clinical contexts. This involved training models using the UMN dataset and testing them on the CYCH dataset. The experiment was designed to test how well the fine-tuning of pretrained models could adapt to a new hospital setting, using a combination of internal and external datasets to assess accuracy in a real-world clinical environment.

We conducted experiment 2 under two distinct scenarios to assess the adaptability and effectiveness of our model in handling abbreviation disambiguation. In the first scenario, we excluded CYCH text, utilizing only the UMN dataset for training. This approach tested the model's ability to generalize from an external dataset to a new environment, applying it subsequently to 998 entries from the CYCH dataset. In the second scenario, we incorporated a small subset of CYCH text into the training process. This was designed to explore incremental learning, where the model adapts to new data while retaining previously learned information, thereby enhancing its predictive performance with minimal data and brief training periods.

Moreover, to maintain consistency and validity in our training process, it was crucial to ensure that all context-candidate pairs appeared in the training set. Consequently, the dataset was carefully screened before splitting, opting for a simple 9:1 ratio between the training and test sets instead of using cross-validation. To evaluate the model's performance, we employed metrics such as accuracy, microaccuracy, and macroaccuracy. These metrics were derived from the confusion matrix for each abbreviation, providing detailed insights into the model's efficacy across different contexts.

Results

Experiment 1

The results for experiment 1, using the MSH WSD dataset, are summarized in Table 2. For the OTO baselines, k-nearest neighbors achieved a macroaccuracy of 94.34%, with microaccuracy data unavailable. The naive Bayes method recorded a macroaccuracy of 93.86%, but microaccuracy was not reported. The LSTM method displayed both macro- and microaccuracy scores, which were 94.87% and 94.78%, respectively. For the OTA baselines, the sense-based deepBioWSD_random embeddings [38] achieved macro- and microaccuracy scores of 93.88% and 93.71%, respectively. The deepBioWSD_pretrained sense embeddings [38] improved prediction performance, with macro- and microaccuracy scores of 96.82% and 96.24%, respectively. Meanwhile, the non–sense-based methods, masked language modeling and permutation language modeling [37], recorded macroaccuracies of 95.89% and 96.83%, respectively, with microaccuracy not reported.

The BERT-base-uncased achieved macro- and microaccuracy scores of 93.64% and 93.38%, respectively. ClinicalBERT recorded a macroaccuracy of 94.77% and a microaccuracy of 94.59%. BlueBERT displayed a competitive performance with macro- and microaccuracies of 95.41% and 95.16%, respectively, on par with other evaluated methods. BlueBERT's macroaccuracy was only slightly lower than the highest performing models, deepBioWSD$_{pretrained\ sense\ embeddings}$ (96.82%) and permutation language modeling (96.83%), but higher than deepBioWSD$_{random\ embeddings}$ (93.88%). This demonstrates BlueBERT's robustness and effectiveness in sequence classification within this study.

The abbreviation disambiguation results for the UMN dataset, presented in Table 3, highlight the performance of various models. BlueBERT excelled, achieving macro- and microaccuracies of 98.4% and 98.22%, respectively, indicating its strong potential for disambiguation tasks. BERT-base-uncased and ClinicalBERT also showed strong performance, though slightly less than BlueBERT. In contrast, OTO-based models like Word2Vec + SVM and BioWordVec + SVM had lower accuracy scores, underscoring the advanced capabilities of the BERT models.

Table 2. Abbreviation disambiguation results (Medical Subject Headings Word Sense Disambiguation).

Method	Macroaccuracy (%)	Microaccuracy (%)
One-to-one		
k-nearest neighbors [33]	94.34	—[a]
Naive Bayes [26]	93.86	—
Long short-term memory [34]	94.87	94.78
One-to-all		
deepBioWSD$_{random\ embeddings}$ [38]	93.88	93.71
deepBioWSD$_{pretrained\ sense\ embeddings}$ [38]	96.82	96.24
Masked language modeling [37]	95.89	—
Permutation language modeling [37]	96.83	—
BERT-base-uncased	93.64	93.38
ClinicalBERT	94.77	94.59
BlueBERT	95.41	95.16

[a]Not applicable.

Table 3. Abbreviation disambiguation results (University of Minnesota).

Method (work)	Macroaccuracy (%)	Microaccuracy (%)
One-to-one		
Word2Vec + SVM[a] [35]	95.79	—[b]
BioWordVec + SVM	94.27	—
One-to-all		
Masked language modeling [37]	98.39	—
Permutation language modeling [37]	98.28	—
BERT-base-uncased	97.59	97.27
ClinicalBERT	98.27	98.01
BlueBERT	98.40	98.22

[a]SVM: support vector machine.
[b]Not applicable.

Overall, the proposed OTA method, especially when implemented using the pretrained BlueBERT model, outperformed the OTO-based approaches. The OTA method's reliance on a single model, as opposed to the multiple models required by OTO methods, improves maintainability and scalability.

Experiment 2

Table 4 displays the abbreviation disambiguation results for the CYCH dataset using BlueBERT. The table compares accuracy percentages for each abbreviation when trained exclusively on external data versus including incremental amounts of CYCH data (5 and 10 samples, respectively). For example, the model recorded a 62.07% accuracy for the abbreviation DM when trained without CYCH data. With the inclusion of CYCH data, the accuracy slightly improved to 70.27% with 5 samples but then slightly decreased to 100% with 10 samples. This trend of initial improvement followed by a marginal decline was observed for other abbreviations as well. Notably, the abbreviation PA showed a substantial increase in performance; it had 0% accuracy when trained

without CYCH data but reached 100% accuracy when trained with either 5 or 10 CYCH samples.

Table 4. Abbreviation disambiguation results of the Chia-Yi Christian Hospital (CYCH) dataset.

Abbreviation	Training without CYCH data, accuracy (%)	Training with CYCH data	
		Accuracy, includes 5 documents (%)	Accuracy, includes 10 documents (%)
ER	97.91	98.03	97.75
DM	62.07	70.27	100
CVA	96.20	98.65	100
PM	77.27	100	100
PA	0.00	100	100

Discussion

Automatic abbreviation disambiguation is crucial in clinical settings as it enhances the clarity and readability of medical records. By accurately interpreting abbreviations, it ensures that health care professionals have a precise understanding of patient information, facilitating accurate diagnoses and effective treatment plans. This automation also speeds up data processing, supports decision-making, and reduces errors, thereby improving overall health care delivery and patient safety.

Traditional OTO methods for abbreviation expansion involve constructing independent models for each abbreviation. Although this method offers high accuracy, it presents challenges in terms of maintenance and generalizability, complicating clinical applications due to the high number of models and associated maintenance costs. In contrast, this study proposes an approach that reduces the number of required models and offers better performance in clinical abbreviation restoration, thereby lowering both the operational and maintenance costs.

Compared to OTO, the OTA approach provides greater scalability, efficiency, and consistency, with a unified model that is easier to maintain and update. However, OTA approaches can be costly in terms of model retraining. Kim et al [37] highlighted that retraining the encoder necessitated high-end GPUs and substantial memory, requiring up to 14 days. Our study adopted a tuning approach using existing pretrained models, substantially cutting down training time to approximately an hour and a half by utilizing free online resources like the K80 GPU through Kaggle Notebook. This method effectively reduces both hardware and time costs, especially beneficial in clinical settings where frequent model updates may be necessary.

This study further demonstrates the practicality of this method in various hospital scenarios, particularly addressing cross-hospital and interdepartmental issues. Our incremental learning approach has been shown to significantly improve prediction results, thereby saving considerable retraining costs.

This study has the following limitations. First, although it preliminarily validates the exceptional effectiveness of the OTA method for abbreviation disambiguation in medical texts, the evaluation is limited by the size of the datasets used. More extensive and comprehensive clinical data are required before application to further validate this method. Second, our study is constrained by the maximum sequence length restriction of the BERT model. Longer clinical notes exceeding the 512-token limit must be truncated, risking the loss of information. Analysis shows that about 16.85% of the MSH WSD dataset and only 0.03% of the UMN dataset exceed this limit. The experimental results indicate superior accuracy for the UMN dataset; however, the performance for the MSH WSD dataset is lower, likely due to significant truncation of longer texts.

Additionally, in generating context-candidate pairs, we retain all candidates and use a FIFO approach for trimming the context. If an abbreviation appears at both the beginning and end of a context and exceeds the token limit, the FIFO method may remove the initial occurrence. Conversely, a last in, first out method could remove an abbreviation appearing at the end. If the same abbreviation carries different meanings in different parts of the text, identical context-candidate pairs may be created post trimming, potentially distorting model training and leading to incorrect predictions.

This study presents an innovative approach to the disambiguation and expansion of abbreviations in clinical medical texts by utilizing context-candidate pairs and the BERT model. This method enhances the readability of medical texts, improving the efficiency of clinical staff who review EMRs and saving time for cross-disciplinary researchers analyzing clinical data, thereby increasing the effectiveness of their studies. Given that clinical medical texts are replete with abbreviations, accurate disambiguation is essential for improving text clarity and usability. Automating this process greatly assists both medical professionals and researchers. The successful application of this model on the investigated datasets underscores its effectiveness and establishes it as a valuable reference for future research in clinical abbreviation expansion.

Data Availability

The codes used for training and evaluating the models on the Medical Subject Headings Word Sense Disambiguation and University of Minnesota datasets are available at [39] and [40] on Kaggle.

Conflicts of Interest

None declared.

References

1. Komeda Y, Handa H, Watanabe T, et al. Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience. Oncology. 2017;93 Suppl 1:30-34. [doi: 10.1159/000481227] [Medline: 29258081]

2. Park HJ, Kim SM, La Yun B, et al. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of breast masses on ultrasound. Medicine (Balt). Jan 2019;98(3):e14146. [doi: 10.1097/MD.0000000000014146] [Medline: 30653149]

3. Sato Y, Takegami Y, Asamoto T, et al. A computer-aided diagnosis system using artificial intelligence for hip fractures-multi-institutional joint development research. arXiv. Preprint posted online on Mar 11, 2020. [doi: 10.48550/arXiv.2003.12443]

4. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data. May 24, 2016;3(1):160035. [doi: 10.1038/sdata.2016.35] [Medline: 27219127]

5. Abhyankar S, Demner-Fushman D, Callaghan FM, McDonald CJ. Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. J Am Med Inform Assoc. 2014;21(5):801-807. [doi: 10.1136/amiajnl-2013-001915] [Medline: 24384230]

6. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. BMC Med Inform Decis Mak. Oct 29, 2020;20(1):280. [doi: 10.1186/s12911-020-01297-6] [Medline: 33121479]

7. Hatef E, Rouhizadeh M, Nau C, et al. Development and assessment of a natural language processing model to identify residential instability in electronic health records' unstructured data: a comparison of 3 integrated healthcare delivery systems. JAMIA Open. Apr 2022;5(1):ooac006. [doi: 10.1093/jamiaopen/ooac006] [Medline: 35224458]

8. Levis M, Levy J, Dufort V, Gobbel GT, Watts BV, Shiner B. Leveraging unstructured electronic medical record notes to derive population-specific suicide risk models. Psychiatry Res. Sep 2022;315:114703. [doi: 10.1016/j.psychres.2022.114703] [Medline: 35841702]

9. Wang M, Wei Z, Jia M, Chen L, Ji H. Deep learning model for multi-classification of infectious diseases from unstructured electronic medical records. BMC Med Inform Decis Mak. Feb 16, 2022;22(1):41. [doi: 10.1186/s12911-022-01776-y] [Medline: 35168624]

10. Sung SF, Lin CY, Hu YH. EMR-based phenotyping of ischemic stroke using supervised machine learning and text mining techniques. IEEE J Biomed Health Inform. Oct 2020;24(10):2922-2931. [doi: 10.1109/JBHI.2020.2976931] [Medline: 32142458]

11. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp. 2001:17-21. [Medline: 11825149]

12. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc. 2010;17(3):229-236. [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]

13. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. Nat Lang Eng. Mar 1995;1(1):83-108. [doi: 10.1017/S1351324900000061]

14. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc. 2010;17(5):507-513. [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]

15. Garla V, Lo Re V, Dorey-Stein Z, et al. The Yale cTAKES extensions for document classification: architecture and application. J Am Med Inform Assoc. 2011;18(5):614-620. [doi: 10.1136/amiajnl-2011-000093] [Medline: 21622934]

16. Sung SF, Chen CH, Pan RC, Hu YH, Jeng JS. Natural language processing enhances prediction of functional outcome after acute ischemic stroke. J Am Heart Assoc. Dec 21, 2021;10(24):e023486. [doi: 10.1161/JAHA.121.023486] [Medline: 34796719]

17. Gao J, He S, Hu J, Chen G. A hybrid system to understand the relations between assessments and plans in progress notes. J Biomed Inform. May 2023;141:104363. [doi: 10.1016/j.jbi.2023.104363] [Medline: 37054961]

18. Ye J, Yao L, Shen J, Janarthanam R, Luo Y. Predicting mortality in critically ill patients with diabetes using machine learning and clinical notes. BMC Med Inform Decis Mak. Dec 30, 2020;20(Suppl 11):295. [doi: 10.1186/s12911-020-01318-4] [Medline: 33380338]

19. Sung SF, Chen K, Wu DP, Hung LC, Su YH, Hu YH. Applying natural language processing techniques to develop a task-specific EMR interface for timely stroke thrombolysis: a feasibility study. Int J Med Inform. Apr 2018;112:149-157. [doi: 10.1016/j.ijmedinf.2018.02.005] [Medline: 29500013]

20. Moon S, Pakhomov S, Melton GB. Automated disambiguation of acronyms and abbreviations in clinical texts: window and training size considerations. AMIA Annu Symp Proc. 2012;2012:1310-1319. [Medline: 23304410]

21. Wu Y, Tang B, Jiang M, Moon S, Denny JC, Xu H. Clinical acronym/abbreviation normalization using a hybrid approach. In: Forner P, Navigli R, Tufis D, Ferro N, editors. Working Notes for (CLEF) 2013 Conference, Valencia, Spain, September 23-26, 2013. CEUR-WS.org; 2013. URL: https://ceur-ws.org/Vol-1179/CLEF2013wn-CLEFeHealth-WuEt2013.pdf [Accessed 2024-09-18]

22. Moon S, Berster BT, Xu H, Cohen T. Word sense disambiguation of clinical abbreviations with hyperdimensional computing. AMIA Annu Symp Proc. Nov 16, 2013;2013:1007-1016. [Medline: 24551390]

23. Wu Y, Denny JC, Trent Rosenbloom S, et al. A long journey to short abbreviations: developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD). J Am Med Inform Assoc. Apr 1, 2017;24(e1):e79-e86. [doi: 10.1093/jamia/ocw109] [Medline: 27539197]

24. Li Y, Wang H, Li X, Deng S, Su T, Zhang W. Disambiguation of medical abbreviations for knowledge organization. Inf Processing Manage. Sep 2023;60(5):103441. [doi: 10.1016/j.ipm.2023.103441]

25. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online on Oct 11, 2018. [doi: 10.48550/arXiv.1810.04805]

26. Jimeno-Yepes AJ, McInnes BT, Aronson AR. Exploiting MeSH indexing in MEDLINE to generate a data set for word sense disambiguation. BMC Bioinformatics. Jun 2, 2011;12(1):1-14. [doi: 10.1186/1471-2105-12-223] [Medline: 21635749]

27. Moon S, Pakhomov S, Melton G. Clinical abbreviation sense inventory. University of Minnesota: University Digital Conservancy. Oct 31, 2012. URL: https://conservancy.umn.edu/items/6651323b-444a-479e-a41a-abca58c2e721 [Accessed 2024-09-18]

28. Finley GP, Pakhomov SVS, McEwan R, Melton GB. Towards comprehensive clinical abbreviation disambiguation using machine-labeled training data. AMIA Annu Symp Proc. Feb 10, 2016;2016:560-569. [Medline: 28269852]

29. Grossman Liu L, Grossman RH, Mitchell EG, et al. A deep database of medical abbreviations and acronyms for natural language processing. Sci Data. Jun 2, 2021;8(1):149. [doi: 10.1038/s41597-021-00929-4] [Medline: 34078918]

30. Alsentzer E, Murphy JR, Boag W, et al. Publicly available clinical BERT embeddings. arXiv. Preprint posted online on Apr 6, 2019. [doi: 10.18653/v1/W19-1909]

31. Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets. arXiv. Preprint posted online on Jun 13, 2019. [doi: 10.48550/arXiv.1906.05474]

32. Huang L, Sun C, Qiu X, Huang X. GlossBERT: BERT for word sense disambiguation with gloss knowledge. In: Inui K, Jiang J, Ng V, Wan W, editors. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics; 2019:3509-3514. [doi: 10.18653/v1/D19-1355]

33. Sabbir A, Jimeno-Yepes A, Kavuluru R. Knowledge-based biomedical word sense disambiguation with neural concept embeddings. Proc IEEE Int Symp Bioinformatics Bioeng. Oct 2017;2017:163-170. [doi: 10.1109/BIBE.2017.00-61] [Medline: 29399672]

34. Jimeno Yepes A. Word embeddings and recurrent neural networks based on long-short term memory nodes in supervised biomedical word sense disambiguation. J Biomed Inform. Sep 2017;73:137-147. [doi: 10.1016/j.jbi.2017.08.001] [Medline: 28797709]

35. Wu Y, Denny JC, Rosenbloom ST, et al. A preliminary study of clinical abbreviation disambiguation in real time. Appl Clin Inform. Jun 3, 2015;6(2):364-374. [doi: 10.4338/ACI-2014-10-RA-0088] [Medline: 26171081]

36. Zhang Y, Chen Q, Yang Z, Lin H, Lu Z. BioWordVec, improving biomedical word embeddings with subword information and MeSH. Sci Data. May 10, 2019;6(1):52. [doi: 10.1038/s41597-019-0055-0] [Medline: 31076572]

37. Kim J, Gong L, Khim J, Weiss JC, Ravikumar P. Improved clinical abbreviation expansion via non-sense-based approaches. Proc Mach Learn Res. May 10, 2020;136:161-178. [doi: 10.1038/s41597-019-0055-0]

38.   Pesaranghader A, Matwin S, Sokolova M, Pesaranghader A. deepBioWSD: effective deep neural word sense disambiguation of biomedical text data. J Am Med Inform Assoc. May 1, 2019;26(5):438-446. [doi: 10.1093/jamia/ocy189] [Medline: 30811548]
39.   Chen S. MSH_paper_bert. Kaggle. Feb 2022. URL: https://www.kaggle.com/code/dsaddicter/msh-paper-bert [Accessed 2024-09-12]
40.   Chen S. UMN_paper_bert. Kaggle. Feb 2022. URL: https://www.kaggle.com/code/dsaddicter/umn-paper-bert [Accessed 2024-09-12]

Abbreviations

BERT: Bidirectional Encoder Representations From Transformers
cTAKES: Clinical Text Analysis and Knowledge Extraction System
CYCH: Chia-Yi Christian Hospital
EMR: electronic medical record
FIFO: first in, first out
LSTM: long short-term memory
MSH WSD: Medical Subject Headings Word Sense Disambiguation
NLP: natural language processing
OTA: one-to-all
OTO: one-to-one
SVM: support vector machine
UMN: University of Minnesota