AI-Augmented Biomedical Entity Linking in Social Media Data

Anonymous ACL submission

Abstract

Social media has become a widely used and convenient venue for mass population to share their health experiences. Health concept expressions on the platforms such as Twitter frequently contain misspellings, abbreviations, paraphrases, and metaphorical language that differ substantially from standardized biomedical terminology, limiting the effectiveness of traditional lexical matching approaches, and making biomedical entity linking (BEL) from social media text challenging. This study proposes a hybrid BEL framework that combines LLMs for symptom span extraction and linguistic normalization with ontology-aware semantic retrieval using SapBERT embeddings with FAISS-based nearest-neighbor search over the UMLS terms and identifiers. Leveraging the ability of LLMs, our framework extracts symptom spans from tweets and generates preferred terms (PTs) from the identified spans, and then retrieves UMLS Concept Unique Identifiers (CUIs) through embedding-based retrieval. The proposed framework was evaluated on a corpus of 600 COVID-19 Twitter posts containing 2,125 manually annotated symptom mentions. Under SQuAD-style lexical evaluation, the framework achieved an F1-score of 0.910, while the evaluation of semantic match using LLMs achieved an F1-score of 0.916 for symptom span extraction. The extracted symptom concepts showed substantial alignment with clinically recognized COVID-19 symptoms published by WHO and CDC, demonstrating the framework's potential for applications such as social media-based public health surveillance.

1 Introduction

The rapid growth of social media platforms has transformed how individuals communicate about personal health experiences. During large-scale public health events such as the COVID-19 pandemic, platforms like Twitter became important spaces for users to share symptoms of the infection, describe illness progression, and seek health-related information in real time (Chen et al., 2020). Mining such data for clinically relevant symptom information has great potential for public health surveillance, outbreak monitoring, and epidemiological research. However, extracting and standardizing symptom mentions from social media text remains a challenging natural language processing (NLP) problem due to the noisy, informal, and highly variable nature of consumer-generated language.

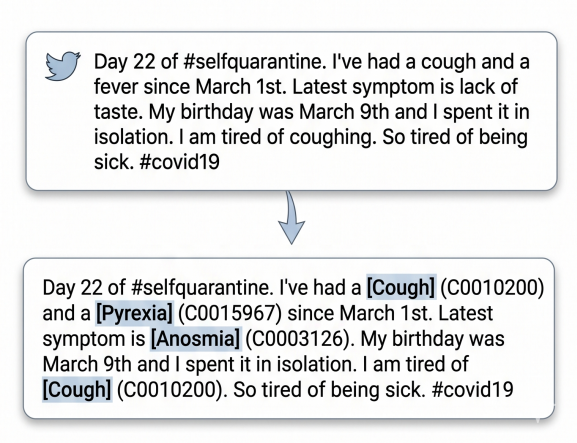

Figure 1: Example of biomedical entity linking (BEL) for symptom entities in Twitter data. The top panel shows the original tweet, while the bottom panel shows the tweet augmented with normalized UMLS preferred terms and their corresponding Concept Unique Identifiers (CUIs).

Unlike clinical notes or biomedical literature, social media posts frequently contain misspellings, abbreviations, colloquial expressions, paraphrases, and metaphorical language (Scepanovic et al., 2020; Tamine and Goeuriot, 2021). A single biomedical concept such as fever may be expressed in many different forms, including "burning up," "running a temp," or "felt like I was on fire." This linguistic variability creates major challenges for

biomedical entity linking (BEL), particularly for approaches that rely heavily on lexical overlap with standardized biomedical terminology.

BEL refers to the task of identifying biomedical entities in raw text and linking them to standardized concepts within biomedical knowledge bases such as the Unified Medical Language System (UMLS) (French and McInnes, 2023; Chen et al., 2026). In social media settings, BEL is especially difficult because health concept expressions are often embedded within informal narratives rather than appearing as easily identifiable biomedical phrases, as illustrated in Figure 1. Effective BEL therefore requires both accurate health concept expression span extraction and robust concept normalization capable of handling highly variable consumer language.

Prior work on biomedical named entity recognition (BioNER) and BEL has primarily focused on formal biomedical text such as clinical notes and scientific abstracts. Traditional systems such as MetaMap (Aronson, 2001), QuickUMLS (Soldaini and Goharian, 2016), and dictionary-based retrieval methods perform well when input text closely resembles standardized biomedical terminology, but their performance deteriorates substantially on noisy social media language (Limsopatham and Collier, 2016).

Several studies have specifically explored medical concept normalization in consumer-generated text. For example, Tutubalina et al. (2018) proposed a recurrent neural network framework for medical concept normalization from social media posts, demonstrating the importance of semantic modeling for handling informal health expressions. Shared tasks such as SMM4H (Weissenbacher et al., 2019) further highlighted the persistent challenges associated with extracting and normalizing symptom mentions from social media text.

More recent embedding-based approaches such as BioSyn (Sung et al., 2020) and SapBERT (Liu et al., 2021) improved biomedical concept retrieval through dense semantic representations. In particular, SapBERT introduced synonym-aware biomedical embeddings trained directly on UMLS synonym pairs, enabling semantic retrieval beyond exact lexical matching.

At the same time, large language models (LLMs) have shown a state-of-the-art capability in interpreting informal language and generating medically grounded representations of consumer health expressions (Havlík, 2024; Dobbins, 2025). However, relatively fewer studies have explored integrating LLM-based extraction and normalization with ontology-aware semantic retrieval in a complete end-to-end BEL pipeline for social media text.

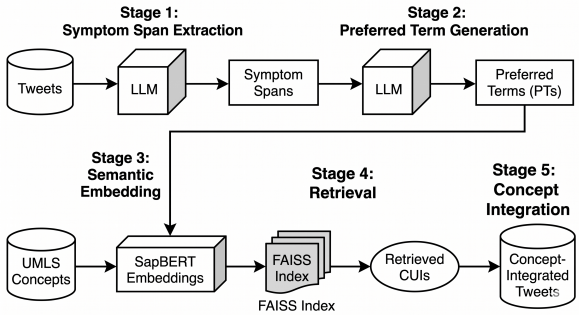

Figure 2: Overview of the proposed biomedical entity linking pipeline.

To address these challenges, this study proposes a hybrid BEL framework that combines LLM-based symptom extraction and linguistic normalization with ontology-aware semantic retrieval using SapBERT embeddings and FAISS-based nearest-neighbor search over the UMLS Metathesaurus. Leveraging the powerful ability of LLMs, the framework first identifies symptom spans from raw tweets and generates medically grounded preferred terms (PTs) from the identified spans, and then retrieves standardized UMLS concepts and their corresponding Concept Unique Identifiers (CUIs) through embedding-based semantic retrieval using cosine similarity.

2 Method

2.1 Overview

In this study, we focus on symptom entity linking and biomedical embeddings. The proposed framework is evaluated on a dataset of 600 COVID-19–related Twitter posts containing symptom mentions and personal health experiences. Performance is assessed using both SQuAD-style lexical evaluation (Rajpurkar et al., 2016) and LLM-based semantic evaluation.

Illustrated in Figure 2 is our proposed hybrid biomedical entity linking (BEL) framework for identifying symptom mentions in social media posts and mapping them to standardized biomedical concepts in the Unified Medical Language System (UMLS).

The pipeline consists of five stages: (1) symptom span extraction from raw tweets using an LLM, (2)

System: You are a biomedical NLP assistant specializing in symptom extraction.
Task: Your task is to extract and return the shortest, most specific positive symptom spans directly from the passage without paraphrasing or normalizing them.
Respond ONLY with a valid JSON list of strings with no explanations or extra text.

Example:
Input: "Day 1 I have extreme fever and sore throat. Day 2 I have slight fever."
Output: ["fever", "sore throat", "fever"]

Passage: *<tweet>*

Figure 3: Prompt for extracting symptom spans from tweets.

generation of medically grounded preferred terms (PTs) using an LLM, (3) creation of SapBERT embeddings for semantic representation, (4) retrieval of standardized UMLS concepts and CUIs using FAISS-based nearest-neighbor search, and (5) integration of the linked biomedical concepts back into the original tweet context.

2.2 Dataset

The study uses a corpus of 600 COVID-19–related Twitter posts describing personal health experiences. Tweets were posted in early 2020 and were manually annotated with symptom spans to facilitate evaluation of the symptom extraction component. In total, the dataset contains 2,125 annotated symptom mentions.

The tweets contain highly informal consumer-generated language, including misspellings, abbreviations, paraphrases, and metaphorical expressions. Examples include expressions such as "burning up" for fever and "can't catch my breath" for dyspnea. These linguistic variations make biomedical entity linking much more challenging for traditional lexical matching approaches.

2.3 Symptom Span Extraction

Symptom span extraction was formulated as a biomedical named entity recognition (BioNER) task. Given a raw tweet, the LLM was prompted to identify the shortest spans corresponding to symptom mentions while ignoring unrelated text. GPT-4o was used for symptom span extraction in this study. Figure 3 shows the prompt used.

The prompting strategy emphasized: (1) extraction of only symptom-related spans, (2) preservation of the original wording from the tweet, and (3) exclusion of inferred or hallucinated symptoms.

For example, given the tweet:

> "I've had a cough and fever since last night and now I can't taste anything.",

the model extracts:

```
["cough", "fever", "can't taste anything"].
```

The extracted spans were evaluated against human annotations using both lexical and semantic matching criteria.

2.4 LLM-Based Linguistic Normalization

Social media symptom expressions frequently differ from standardized biomedical terminology. To reduce linguistic variability before concept retrieval, the extracted symptom spans were normalized into medically grounded preferred terms (PTs) using an LLM.

Rather than directly predicting UMLS concept identifiers, the model generates clinically interpretable PTs showing standard biomedical terminology.

For example, given the symptom spans:

```
["cough", "fever", "can't taste anything"],
```

the LLM predicts:

```
["Cough", "Pyrexia", "Ageusia"].
```

This intermediate normalization step reduces lexical variability and improves downstream semantic retrieval performance.

2.5 UMLS Terminology Preparation

The Unified Medical Language System (UMLS) Metathesaurus (2025AA release) was used as the biomedical knowledge base for concept retrieval (U.S. National Library of Medicine, 2025). UMLS integrates terminology from multiple biomedical vocabularies and organizes synonymous expressions under shared Concept Unique Identifiers (CUIs).

Within UMLS, a single biomedical concept may contain multiple synonymous surface forms. Among these terms, one representation is designated as the *Preferred Term* (PT), which is computed from a list of ranked source vocabularies and serves as the standardized name for the concept.

For example, given the preferred terms:

```
["Cough", "Pyrexia", "Ageusia"],
```

the corresponding UMLS CUIs are:

```
[C0010200, C0015967, C0003126].
```

Since each concept is associated with a preferred term within each source vocabulary, it may have multiple PTs across different vocabularies. For example, the concept of fever may appear as "fever" in one vocabulary and "pyrexia" in another, while both are linked to the same CUI.

Mapping informal symptom expressions to medically grounded PTs improves alignment between input text and standardized biomedical terminology, particularly when handling abbreviations, misspellings, or nonstandard language commonly found in social media data.

In this study, only English-language concepts were extracted from the UMLS Metathesaurus, resulting in a collection of approximately 3.59 million UMLS terms.

2.6 Biomedical Embeddings with SapBERT

Biomedical concept embeddings were generated using SapBERT (Liu et al., 2021), a transformer-based model initialized from PubMedBERT and further trained using UMLS synonym pairs. SapBERT applies metric learning to align synonymous biomedical expressions within the embedding space while separating unrelated concepts.

Given an anchor term x_i and a synonymous term x_j sharing the same UMLS CUI, the training objective is defined as:

$$L = -\log \frac{\exp(\mathrm{sim}(x_i, x_j)/\tau)}{\sum_k \exp(\mathrm{sim}(x_i, x_k)/\tau)}$$

where x_k denotes other terms in the batch, $\mathrm{sim}(\cdot)$ is a similarity function, and τ is a temperature scaling parameter. This objective encourages semantically equivalent biomedical terms to occupy nearby regions in the embedding space.

Each biomedical term was encoded into a 768-dimensional dense vector representation using SapBERT. Following prior embedding-based retrieval approaches, all vectors were L2-normalized:

$$\hat{x} = \frac{x}{\|x\|}$$

This normalization ensures that similarity comparisons depend on semantic orientation rather than vector magnitude.

2.7 FAISS-Based Semantic Retrieval

To enable scalable retrieval across millions of UMLS concepts, all SapBERT embeddings were indexed using Facebook AI Similarity Search (FAISS) (Douze et al., 2025). FAISS supports efficient nearest-neighbor retrieval over high-dimensional biomedical embeddings.

Given a normalized query embedding $\hat{q}$ and a database embedding $\hat{x}_i$, semantic similarity is computed using cosine similarity:

$$\mathrm{sim}(\hat{q}, \hat{x}_i) = \frac{\hat{q} \cdot \hat{x}_i}{\|\hat{q}\|\|\hat{x}_i\|}$$

Because all embeddings are L2-normalized, cosine similarity reduces to the inner product between vectors. During inference, the predicted biomedical concept is obtained through nearest-neighbor retrieval:

$$x^* = \arg\max_i (\hat{q} \cdot \hat{x}_i),$$

where x^* denotes the retrieved UMLS concept embedding.

An exact nearest-neighbor search strategy was implemented using the FAISS `IndexFlatIP` structure. This retrieval framework enables efficient semantic search across approximately 3.59 million English-language UMLS terms.

2.8 Performance Evaluation

The framework was evaluated for symptom extraction at the tweet level, where each tweet contains a set of annotated symptom spans. Performance was measured using precision, recall, and F1-score under two evaluation settings:

- **Lexical evaluation**: SQuAD-style lexical matching based on exact and partial matches between predicted and annotated symptom spans, implemented as string comparisons.
- **Semantic evaluation**: LLM-assisted one-to-one semantic matching between conceptually equivalent symptom expressions, even when lexical forms differ.

For lexical evaluation, text was first normalized by converting to lowercase and removing punctuation and articles. Predicted and annotated symptom spans were then tokenized, and metrics were computed based on token overlap following the standard SQuAD evaluation framework.

System: You are a linguist capable of finding semantically similar terms.
Task: You are comparing two lists of symptom expressions.

Rules:
- Treat each list item as a separate occurrence (duplicates allowed).
- Match semantically similar terms.
- No external medical knowledge.
- One-to-one matching only.
- Each index may appear at most once.

Return ONLY valid JSON:
```
{{
  "matches": [{{"list1_idx": i, "list2_idx": j}}],
  "unmatched_list_1": [],
  "unmatched_list_2": []
}}
```

List 1:
<symptom list 1>

List 2:
<symptom list 2>

Figure 4: Prompt for finding semantically similar symptom spans.

Table 1: Lexical and semantic evaluation results for symptom extraction from tweets.

Evaluation Type	Precision	Recall	F1
Lexical (SQuAD-style)	0.948	0.902	0.910
Semantic (LLM-assisted)	0.935	0.915	0.916

Semantic evaluation was introduced to account for conceptually equivalent symptom expressions that may differ lexically. For example, phrases such as "high temperature" and "fever" represent the same biomedical concept despite low lexical overlap. GPT-4o was used to perform semantic matching between predicted and annotated symptom spans. Figure 4 shows the prompt used for semantic evaluation.

3 Results

Table 1 shows the tweet-level lexical and semantic evaluation results for LLM-based symptom span extraction across the 600 COVID-19 Twitter posts. Under SQuAD-style lexical evaluation, the framework achieved an F1-score of 0.910, while semantic evaluation achieved an F1-score of 0.916.

Table 2 further summarizes the lexical matching outcomes. Among the 600 evaluated tweets, 391 produced exact matches, 200 produced partial matches, and 9 produced complete mismatches.

Representative examples of symptom mentions, generated preferred terms (PTs), and mapped UMLS Concept Unique Identifiers (CUIs) are shown in Table 3. Across the dataset, 2,125 annotated symptom mentions were normalized into 199 unique PTs and mapped to 174 unique UMLS CUIs, demonstrating substantial reduction of linguistic variability through normalization and semantic retrieval. Informal expressions such as "hard to breathe" and "no sense of smell" were successfully normalized to standardized biomedical terminology prior to concept retrieval.

Table 2: Tweet-level lexical matching outcomes under SQuAD-style evaluation.

Match Type	Tweets	Percentage
Exact Match	391	65.2%
Partial Match	200	33.3%
Complete Mismatch	9	1.5%

Table 3: Examples of symptom mentions, preferred terms, and their mapped UMLS CUIs.

Symptom Mention	Preferred Term	UMLS CUI
dry tickles at back of throat	Throat irritation	C0700184
eyes on fire	Burning eyes	C0015395
feeling nauseous	Nausea	C0027497
hard to breathe	Dyspnea	C0013404
lost my sense of taste	Ageusia	C2364111
muscles ache	Myalgia	C0231528
no sense of smell	Anosmia	C0003126
runny nose	Rhinorrhea	C1260880
stuffy nose	Nasal congestion	C0027424
temp keeps going up and down	Fluctuating temperature	C1504363

Figure 5 illustrates an end-to-end example of the proposed biomedical entity linking pipeline, including symptom extraction, linguistic normalization, and final CUI mapping.

The top 10 most frequently identified symptom concepts in the tweet dataset are presented in Figure 6. Common concepts such as cough, fatigue, pyrexia, and dyspnea appear with the highest frequency, with cough being the most frequent ($n = 298$).

Figure 7 shows the number of unique mention variants associated with each symptom concept. Symptoms such as dyspnea, fatigue, fever, and cough exhibit greater linguistic diversity in social media expressions, with dyspnea showing the highest variability at 63 distinct variants. Figure 8 further illustrates the distribution of mention variants for "fever," where multiple expressions (e.g., "temperature," "feverish," and "elevated temp") refer to the same underlying concept alongside a long tail of less frequent variants grouped under "Other."

Figure 9 presents the alignment between normalized symptom concepts identified in the tweet dataset and standardized COVID-19 symptom

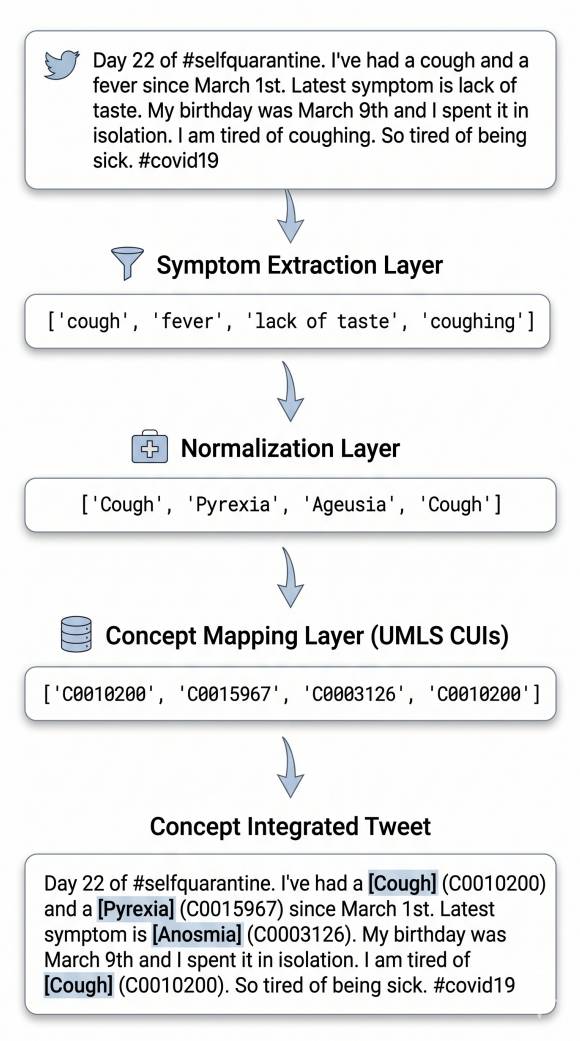

Figure 5: Example of end-to-end concept normalization and CUI mapping.

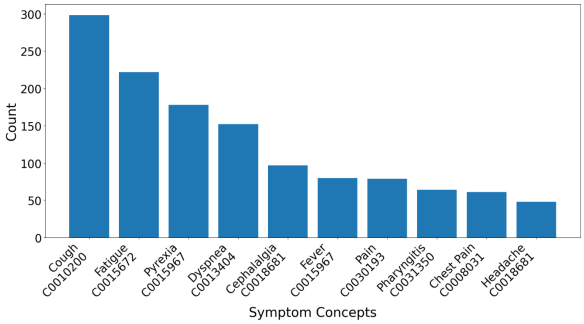

Figure 6: Top 10 mentioned symptom concepts by frequency.

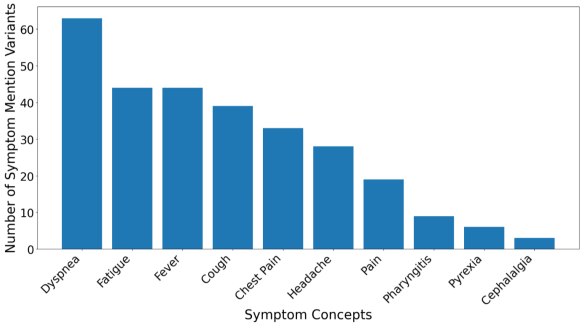

Figure 7: Number of unique symptom mention variants per top 10 symptom concepts.

lists published by the World Health Organization (WHO) (World Health Organization, 2025) and United States Centers for Disease Control and Prevention (CDC) (Centers for Disease Control and Prevention, 2025).

4 Discussions

The above results demonstrate that the proposed hybrid framework effectively supports biomedical entity linking from noisy social media text. The strong performance under both lexical and semantic evaluation settings suggests that the LLM is able to identify symptom spans with high agreement to human annotations, even when symptom expressions vary lexically.

The difference between lexical and semantic evaluation highlights the limitations of strict token-level matching for social media health text. Many symptom expressions conveyed equivalent biomedical meaning despite differences in wording, abbreviations, or paraphrasing. The semantic evaluation therefore provides a more realistic assessment of concept-level symptom extraction performance. Our LLM-based approach for recognizing symptom spans narrows the performance gap between lexical and semantic evaluations.

The normalization examples further demonstrate the importance of the intermediate preferred term generation step. Informal consumer expressions were successfully transformed into medically grounded terminology before semantic retrieval within the UMLS concept space.

The linguistic variability analysis shown in Figures 7 and 8 illustrates the large number of distinct surface forms associated with the same symptom concept in social media data. This variability presents a major challenge for traditional lexical matching systems and motivates the use of LLM-assisted normalization combined with embedding-based retrieval.

The extracted symptom concepts showed substantial overlap with clinically recognized COVID-19 symptoms published by WHO and CDC, as shown in Figure 9. These findings demonstrate that the proposed framework can effectively cap-

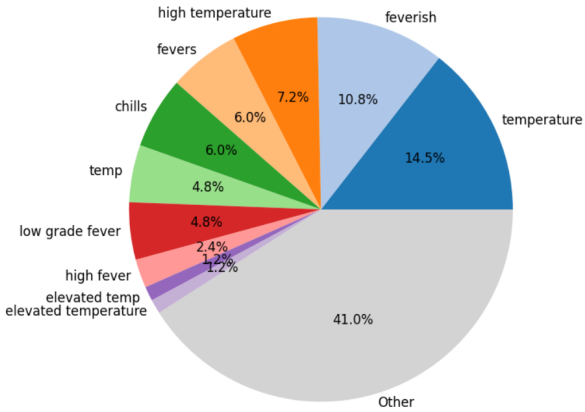

Figure 8: Distribution of top symptom mention variants for "fever".

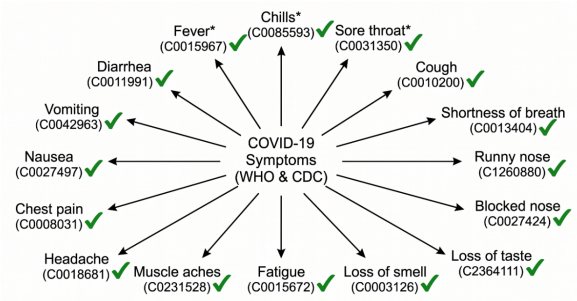

Figure 9: Alignment of normalized symptom concepts with standardized COVID-19 symptoms defined by WHO and CDC. Note: * indicates the most common symptoms identified by WHO.

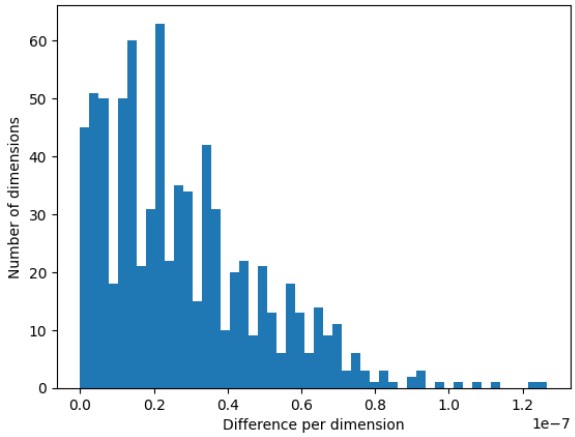

Figure 10: Per-dimension embedding differences between two retrieved CUIs for concept "pain". Each embedding is of 768 dimensions, and this chart shows the statistics of the numeric difference between each pair of the corresponding dimension.

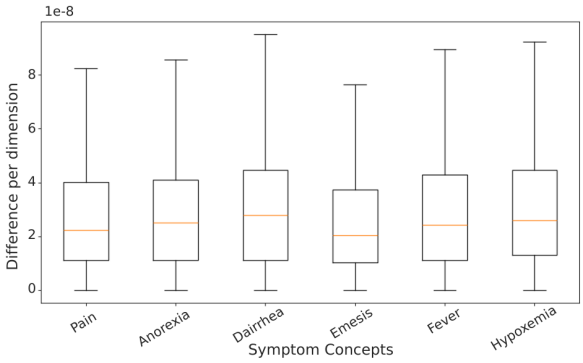

Figure 11: Distribution of per-dimension embedding differences for CUI pairs with identical similarity scores across six representative cases.

ture clinically relevant symptom information from noisy social media text and map it to standardized biomedical concepts.

An intriguing observation in this study involves the behavior of embedding-based retrieval when semantically similar biomedical concepts occupy nearly identical regions of the embedding space. In certain cases, some UMLS concepts received near-identical similarity scores due to extremely close embedding representations.

For example, when querying the term "pain," FAISS retrieved two different CUIs associated with different semantic types: C1549543 (*Functional Concept*) and C0030193 (*Sign or Symptom*). Both concepts produced a cosine similarity score of 1.0 with the query embedding, indicating nearly identical alignment in the vector space. The observed embedding differences were extremely small, on the order of 10^{-7}, suggesting only minor numerical precision differences rather than meaningful semantic dissimilarity.

This behavior may be attributed to the design of SapBERT, which intentionally places synonymous or highly related biomedical terms close together in the embedding space. While this property improves semantic retrieval performance, it can also introduce ambiguity when multiple concepts share nearly equivalent vector representations.

A representative example for "pain" is shown in Figure 10, where most of the 768 dimensions exhibit near-zero differences. This observation is further supported by Figure 11, which shows the distribution of embedding differences across all the six identified cases in the dataset and demonstrates consistently minimal variation across concepts.

5 Conclusion

This study presented a hybrid biomedical entity linking framework that combines LLM-based symptom extraction and linguistic normalization with ontology-aware semantic retrieval using Sap-

BERT embeddings and FAISS-based search over the UMLS. The framework effectively extracted symptom spans from raw tweet text, normalized these diverse informal symptom expressions referring to the same biomedical concept into medically grounded preferred terms and mapped them to standardized UMLS CUIs. The extracted concepts showed substantial alignment with clinically recognized COVID-19 symptoms published by WHO and CDC. These findings demonstrate the potential of integrating LLMs with biomedical embedding-based retrieval systems for robust concept normalization and biomedical entity linking in noisy social media environments. However, challenges remain in distinguishing semantically overlapping biomedical concepts with highly similar embedding representations. Future work may explore improved disambiguation strategies, evaluation across broader clinical domains, and adaptation to multilingual and real-time public health surveillance settings.

Limitations

The study focused primarily on symptom-related biomedical entities without examining broader biomedical entity categories within the UMLS semantic type hierarchy, such as diseases, medications, or procedures. In addition, the proposed framework was evaluated on a relatively small dataset of 600 COVID-19–related English-language Twitter posts, which may limit generalizability to other domains, languages, and social media platforms. The work relies on LLM-based extraction and normalization, where outputs may vary depending on prompting strategies and model behavior, potentially affecting reproducibility and consistency. The semantic evaluation also involved LLM-assisted judgment of conceptual similarity, which may introduce subjectivity compared to fully deterministic evaluation methods. Finally, the embedding-based retrieval approach using SapBERT and FAISS can produce ambiguity when semantically overlapping biomedical concepts occupy nearly identical regions in the embedding space, resulting in multiple CUIs receiving almost identical similarity scores.

Ethical Considerations

This study uses publicly available COVID-19–related Twitter posts for biomedical entity linking research. No attempt was made to identify individual users, and all use of the UMLS Metathesaurus complies with the U.S. National Library of Medicine license agreement. The proposed framework is intended for biomedical NLP research and public health surveillance rather than clinical diagnosis. Because the framework relies on LLMs and noisy social media data, outputs may contain biases or inconsistencies that affect downstream interpretation. Any practical deployment of this framework would require appropriate validation and domain expert oversight.

References

Alan R Aronson. 2001. Effective mapping of biomedical text to the umls metathesaurus: the metamap program. In *Proceedings of the AMIA Symposium*, page 17.

Centers for Disease Control and Prevention. 2025. Signs and symptoms of COVID-19. https://www.cdc.gov/covid/signs-symptoms/index.html.

Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Covid-19: The first public coronavirus twitter dataset.

Haihua Chen, Yuhan Zhou, Ruochi Li, Aryan Murthy Illa, Ana Cleveland, and Junhua Ding. 2026. A comprehensive survey on medical concept normalization: Datasets, techniques, applications, and future directions. *Journal of Biomedical Informatics*, page 105005.

Nicholas J Dobbins. 2025. Generalizable and scalable multistage biomedical concept normalization leveraging large language models. *Research Synthesis Methods*, 16(3):479–490.

Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2025. The faiss library. *IEEE Transactions on Big Data*.

Evan French and Bridget T McInnes. 2023. An overview of biomedical entity linking throughout the years. *Journal of biomedical informatics*, 137:104252.

Vladimír Havlík. 2024. Meaning and understanding in large language models. *Synthese*, 205(1):9.

Nut Limsopatham and Nigel Collier. 2016. Normalising medical concepts in social media texts by learning semantic representation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (volume 1: long papers)*, pages 1014–1023.

Fangyu Liu, Ehsan Shareghi, Zaiqiao Meng, Marco Basaldella, and Nigel Collier. 2021. Self-alignment pretraining for biomedical entity representations. In

Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies, pages 4228–4238.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 2383–2392.

Sanja Scepanovic, Enrique Martin-Lopez, Daniele Quercia, and Khan Baykaner. 2020. Extracting medical entities from social media. In *Proceedings of the ACM conference on health, inference, and learning*, pages 170–181.

Luca Soldaini and Nazli Goharian. 2016. Quickumls: a fast, unsupervised approach for medical concept extraction. In *MedIR workshop, sigir*, pages 1–4.

Mujeen Sung, Hwisang Jeon, Jinhyuk Lee, and Jaewoo Kang. 2020. Biomedical entity representations with synonym marginalization. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3641–3650.

Lynda Tamine and Lorraine Goeuriot. 2021. Semantic information retrieval on medical texts: Research challenges, survey, and open issues. *ACM Computing Surveys (CSUR)*, 54(7):1–38.

Elena Tutubalina, Zulfat Miftahutdinov, Sergey Nikolenko, and Valentin Malykh. 2018. Medical concept normalization in social media posts with recurrent neural networks. *Journal of biomedical informatics*, 84:93–102.

U.S. National Library of Medicine. 2025. Umls knowledge sources. https://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html. Release 2025AA, National Library of Medicine (US), Bethesda, MD.

Davy Weissenbacher, Abeed Sarker, Arjun Magge, Ashlynn Daughton, Karen O'Connor, Michael Paul, and Graciela Gonzalez. 2019. Overview of the fourth social media mining for health (smm4h) shared tasks at acl 2019. In *Proceedings of the fourth social media mining for health applications (# SMM4H) workshop & shared task*, pages 21–30.

World Health Organization. 2025. Coronavirus disease (COVID-19) key facts. https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-(covid-19).