HeilArchive : Pages:1–14

Original Research

Zero-Shot Clinical Concept Normalization Using Prompt-Based Language Models

Bishnu Prasad Sharma[1]

[1]PhD at Nepal Sanskrit University Beljhundi, Dang, Nepal.

Abstract

This paper explores the domain of zero-shot clinical concept normalization, leveraging prompt-based language models to align unstructured clinical text with standardized medical terminologies. The approach circumvents traditional data-hungry methods by framing the normalization task as a conditional text inference problem, invoking the model's latent conceptual understanding. We address the complexity of heterogeneous medical vocabulary by prompting an underlying model to infer the most probable canonical label, given minimal or no explicitly labeled training samples. The method is grounded in the principle that each concept, represented by a textual descriptor, can be mapped onto a structured taxonomy through a contextual prompt. By directly prompting large language models with carefully designed prompts, the system capitalizes on the model's prior knowledge, thereby enabling on-the-fly resolution of diverse clinical expressions. We propose a rigorous formal framework and employ advanced mathematical concepts to enhance interpretability, offering insights into the underlying reasoning within the model. With experiments on widely used clinical corpora, results highlight competitive performance in normalizing unseen or minimally sampled expressions. Notably, the technique addresses lexical variation and out-of-vocabulary issues by exploiting prompt-driven cross-lingual and cross-domain transfer abilities. Our findings advance the state of the art in zero-shot clinical concept normalization and pave the way for broader medical natural language processing applications.

1. Introduction

Zero-shot clinical concept normalization has emerged as a pivotal research problem in medical natural language processing, primarily due to the extensive variability in how medical concepts are expressed across patients, institutions, and clinical specialties [1]. When medical reports are generated, practitioners often use a broad range of abbreviations, synonyms, localized jargon, and informal phrasing to record patient conditions, diagnoses, treatments, or observations. The difficulty of mapping these free-text expressions to standardized ontologies, such as SNOMED CT or the Unified Medical Language System, grows exponentially with the increasing scale and heterogeneity of clinical data [2]. Traditional supervised learning approaches rely heavily on annotated corpora, which are time-consuming and expensive to produce, given that the annotation process often requires domain expertise. Consequently, there is growing interest in designing systems that can perform concept normalization without labeled training data or with very limited supervision, encouraging researchers to explore zero-shot or few-shot learning paradigms [3, 4]

A particularly promising route to address this challenge involves prompt-based language models. The essential concept rests on leveraging the vast distributional and factual knowledge embedded in large language models [5]. By formulating a set of textual prompts, the aim is to guide the model in hypothesizing the correct standardized label for each clinical expression, without explicit exposure to extensive training annotations. The model's latent parameters, shaped by pretraining over massive biomedical or general-domain text, can be coaxed into performing a conceptual alignment task. This strategy capitalizes on knowledge distillation techniques, sub-symbolic representations, and advanced textual reasoning [6]. The innovation is that the normalization procedure shifts from an elaborate pipeline requiring domain-specific heuristics to a more straightforward prompt design exercise, where carefully chosen textual cues encode the intended mapping from free-form expressions to canonical terminologies.

The central difficulty in zero-shot clinical concept normalization lies in coping with the complex interplay of lexical variations, ambiguous abbreviations, and evolving medical terminologies [7]. Unlike many conventional natural language processing tasks, where the domain might remain relatively stable, clinical vocabularies are perpetually in flux. Furthermore, certain clinical expressions can be disambiguated only through subtle context, indicating the need to incorporate semantic understanding of the surrounding text [8]. This complexity is magnified in specialized subfields of medicine, where domain experts might rely on localized terminologies or less standardized expressions. Therefore, an approach that can elegantly handle out-of-vocabulary terms or rarely encountered abbreviations by leveraging the model's accumulated linguistic and medical knowledge can be a significant boon for automated clinical documentation systems. [9]

To align with the requirements of real-world applications, the proposed framework must also respect constraints around interpretability, reliability, and compliance with clinical standards. Medical decision-making depends heavily on accurate concept labels, so an algorithmic misalignment could lead to detrimental outcomes. An advantage of large language models, in this context, is the potential to integrate external knowledge bases or logic constraints at inference time, refining the prompt to discourage implausible mappings [10]. As the model transforms an input expression into a canonical label, the bridging mechanism can incorporate optional constraints to ensure alignment with permissible codes in the relevant medical ontology. In certain scenarios, the final predicted label must exist in a recognized subset of the ontology, ensuring that no extraneous or spurious concepts are assigned. [11]

Formally, we consider a set of clinical expressions labeled by x, drawn from a vocabulary V. We aim to map each x to a standardized label y that belongs to a medical terminology set T, where T is typically large, hierarchical, and replete with relationships between concepts [12]. In a zero-shot setting, we assume we have a pre-trained model M that has not seen explicit training examples of the mapping x $\rightarrow$ y. Instead, we design a prompt p(x) that, when concatenated with x, leads M to output a suitable candidate for y. The textual structure of p(x) may involve instructions, synonyms, or sample demonstrations from other domains, all chosen to maximize the probability that M aligns x with the correct y [13]. This approach can also integrate logic-based constraints such as the expression

$$\forall x \in V,\ \exists y \in T: \text{model\_match}(x, y)$$

which indicates that for each expression x, there should be a y in T that is semantically or contextually nearest to x according to M's internal knowledge. In this way, the entire procedure may be cast as a constrained inference task over a large-scale textual encoding system.

From a theoretical perspective, this new direction represents a fusion between symbolic and sub-symbolic paradigms [14]. The model's sub-symbolic structure, instantiated by hidden vectors and attention mechanisms, is exposed to symbolic constraints during prompting. By adjusting the structure of the prompt, we effectively embed a symbolic representation of the concept normalization task into the model's forward pass [15]. In advanced formulations, we can define a continuous optimization over the prompt space to identify an optimal prompt that maximizes the likelihood of correct normalization. This leads to interesting connections with gradient-based training on textual inputs, although typically the zero-shot scenario avoids direct gradient updates to the model parameters themselves. [16]

Practical applications range from simplifying the generation of high-level reports to facilitating interoperability across electronic health record systems. When records are consistently labeled with standardized terminologies, subsequent tasks such as automated billing, quality assurance, or clinical research become more tractable. Yet a major question remains: how accurately can a purely prompt-based approach normalize arbitrary clinical expressions, especially in specialized contexts or niche domains where the language model might have insufficient prior exposure? Our empirical results suggest that although there are limitations, the performance is often on par with or superior to that of specialized classifiers trained on small annotated datasets [17]. This is especially true if the model has been pre-trained on a sufficiently diverse set of medical corpora, thereby encoding extensive domain knowledge.

As zero-shot clinical concept normalization continues to be refined, it promises to streamline the labor-intensive process of data annotation [18]. Instead of requiring a specialized classifier for each sub-domain or new set of expressions, users can craft a prompt that guides the model toward correct mappings. This perspective places the burden of domain adaptation and knowledge integration onto the design of prompts and the pre-training data coverage of the model [19]. The rest of this paper delves deeper into the theoretical foundations that guide prompt-based normalization, the methodological details of our proposed framework, extensive experimental validation, and a thorough discussion of challenges and limitations. We conclude by charting potential future avenues where zero-shot strategies might further revolutionize the clinical text processing pipeline.

2. Foundations

The formal underpinnings of zero-shot clinical concept normalization via prompt-based language models can be analyzed through the lens of compositional semantics, distributional language modeling, and monotonic inference under constraints [20]. Consider a language model M trained on a broad distribution of text. Each token t in M's vocabulary is mapped to a vector in some high-dimensional space, and a sequence of tokens is transformed by a series of linear transformations and non-linear activations [21]. The network's final output layer provides a distribution over possible next tokens, which can be re-purposed for classification or mapping tasks by an appropriate prompting scheme.

To model the problem more concretely, let x denote a clinical expression, such as a short phrase or acronym [22, 23]. We wish to find a label y from a standardized medical taxonomy T that best captures the meaning of x. One can define a mapping function f: V → T, where V is the space of all potential clinical expressions [24]. However, in zero-shot learning, f is not learned directly via annotated pairs (x,y). Instead, we craft a textual prompt p(x) whose concatenation with x yields a context s = [p(x); x]. The language model M estimates probabilities over possible continuations [25]. If we designate a textual form for each candidate y, the question becomes which textual label y* maximizes M's conditional probability M(y* | s). Symbolically, [26]

$$y^* = \arg\max_{y \in T} M\big(y \mid [p(x); x]\big).$$

This is a template-based approach that can be enriched by logic constraints, such as

$$\exists y \in T : \text{semantic\_equivalence}(x, y)$$

or domain constraints

$$\neg\text{invalid\_mapping}(x, y).$$

The distributional perspective can be articulated using the concept of embedding alignment. Let $\phi(x) \in \mathbb{R}^d$ denote the contextual embedding of x within the model M, influenced by the prompt structure. Similarly, let $\psi(y) \in \mathbb{R}^d$ be an embedding of the candidate label y. We want $\phi(x) \approx \psi(y)$ in some vector space norm [27]. This approximate equivalence might be measured via a function such as cosine similarity or an inner product that reflects the model's sense of semantic proximity. Because the model is pre-trained, both mappings $x \mapsto \phi(x)$ and $y \mapsto \psi(y)$ are shaped by prior distributions in the training data, which might include massive corpora of medical texts [28]. The zero-shot paradigm bypasses any domain-specific fine-tuning by assuming that the distributional geometry already encodes enough relational information to map x to y.

Logic statements can further formalize aspects of conceptual alignment. If we consider a simple propositional structure, let L(x,y) be a predicate that is true when x is properly normalized to y [29]. We can impose the constraint

$$\forall x \in V, \exists y \in T : L(x, y).$$

In the typical scenario, we also want uniqueness, which might be encoded as

$$\forall x \in V, \forall y_1, y_2 \in T : (L(x, y_1) \wedge L(x, y_2)) \rightarrow y_1 = y_2.$$

This ensures each expression x has exactly one correct concept label in T. In practice, the language model does not enforce uniqueness automatically, but one can impose a maximum likelihood or maximum similarity criterion to choose the single best y. [30]

In advanced mathematical treatments, we can view the process as partially constrained optimization in a high-dimensional textual manifold. Let the textual manifold be defined by all possible sequences of tokens [31]. Prompt engineering then becomes the process of finding a textual neighborhood that best elicits the correct mapping from x to y. One might define a function

$$\rho(p, x, y)$$

that measures how well the prompt p aligns x to y, typically as a negative log-likelihood of y given [p; x]. Minimizing this function with respect to p yields an optimal or near-optimal prompt [32]. However, in a zero-shot context, we typically rely on heuristic or domain-specific prompt designs rather than direct optimization, because we do not fine-tune the language model on new data.

Another theoretical angle is gleaned from linear algebra [33]. If we stack the embeddings of clinical expressions $x_1, x_2, \ldots, x_n$ into a matrix X, and the embeddings of candidate labels $y_1, y_2, \ldots, y_m$ into a matrix Y, the normalization task can be framed as finding a partial alignment from rows of X to rows of Y. In the simplest scenario, one identifies the row y_j in Y that maximizes the dot product with the row x_i in X [34]. Incorporating the prompt modifies the embedding space by shifting or rotating the distribution, effectively re-centering the meaning of x_i. This can be conceptualized as a prompt-induced linear transformation P such that [35]

$$\widetilde{X} = P \cdot X,$$

making it more likely that the semantic direction of x_i aligns with y_j. As the complexity of the language model increases, a purely linear viewpoint becomes insufficient, but it remains a valuable lens through which to interpret the global geometry of concept embeddings.

These theoretical foundations support the notion that zero-shot clinical concept normalization is not simply a heuristic guess of the correct label but follows from deeper structural properties of distributional representations, logic constraints, and prompt-based inference [36]. While each theoretical angle captures only a portion of the phenomenon, together they illustrate the multifaceted nature of zero-shot normalization. The interplay between symbolic logic and sub-symbolic geometry is at the heart of why prompt-based strategies can succeed in complex medical domains, provided the language model has been exposed to sufficient domain-relevant data during pre-training. [37]

3. Proposed Methodology

The methodology for zero-shot clinical concept normalization via prompt-based language models involves crafting a carefully designed textual prompt that steers the model toward selecting a standard medical concept. Let x be an unnormalized clinical expression, such as an abbreviation or a colloquial name for a drug [38]. We aim to predict a canonical label y in T that best represents x. The proposed approach follows three main steps: prompt construction, concept candidate representation, and inference.

Prompt construction begins with an analysis of the target domain [39]. A simple template might be: `"Expression: x. Please provide the standardized clinical concept` `that this expression refers to."` More sophisticated prompts can incorporate additional context, such as: "`Expression: x. A patient record mentions this expression. It corresponds to the condition: ___.`" The specific structure is flexible, with the central requirement that the prompt provides sufficient semantic context for the model M to infer that x refers to a particular concept in the target set T.

Logic constraints can also be embedded as textual instructions, specifying that the output must conform to a recognized ontology [40]. For instance, a prompt might state: "`Only valid terminology from the official medical codes is acceptable.`" Such constraints serve to guide M, encouraging it to restrict its outputs to the desired conceptual space.

Logic constraints can also be embedded as textual instructions, specifying that the output must conform to a recognized ontology. For instance, a prompt might state: "Only valid terminology from the official medical codes is acceptable." Such constraints serve to guide M, encouraging it to restrict its outputs to the desired conceptual space. [41]

The next step, concept candidate representation, leverages the fact that T might be large, containing thousands or tens of thousands of concept labels. Enumerating all possible labels is often unfeasible in practice [42]. Instead, we can adopt a nearest neighbor filtering or approximate matching approach. We embed each candidate label y using the same model M or a related text encoder. Then, for a given x, we compute a similarity measure between x and each y [43]. We select the top K most similar y values as candidate labels. This restricted set significantly reduces computational overhead while retaining the likely correct matches [44]. Crucially, if M is used for both x and y embeddings, the prompt can remain partially fixed, simply encoding each candidate label in a standard format. The zero-shot inference becomes more efficient by focusing on a manageable subset of T. [45]

Inference then entails choosing the label y* that maximizes the model's confidence. One can measure confidence in multiple ways. For example, we can compute a probability distribution over candidate labels using a softmax that depends on M(y | p(x)) for each y in the candidate set [46]. Another approach is to compute an embedding for the prompt-augmented expression,

$$\phi([p(x);x]),$$

and then measure the cosine similarity with each candidate label embedding. Mathematically, we could define [47]

$$y^{*} = \arg\max_{y\in\Gamma(x)} \frac{\phi([p(x);x])\cdot\psi(y)}{\|\phi([p(x);x])\|\|\psi(y)\|},$$

Let $\Gamma(x)$ denote the reduced set of candidate labels selected via nearest neighbor filtering. In a purely text generation approach, we might ask the model to produce a token sequence that spells out y, then measure the likelihood or presence of y in the generated output [48]. All these inference modes revolve around the central principle of leveraging the model's internal representation of linguistic and conceptual knowledge to map x to y effectively.

While zero-shot approaches sidestep direct supervised training, it is often beneficial to refine prompts iteratively [49]. One technique involves systematically evaluating the performance of different prompt formats on a small validation set or on manually crafted examples. If the model tends to produce spurious labels, the prompt may need to be more explicit about the domain or the required level of specificity. Conversely, if the prompt leads to overly generic responses, we might include additional guidance instructing the model to return the most clinically precise concept. [50]

Another refinement strategy is to incorporate minimal examples into the prompt. For instance: "`When given 'HA', we refer to 'Headache' [51]. Now, given '[Expression: x]', the concept is: ___.`" Although this begins to resemble a few-shot approach when multiple examples are included, even a single demonstration can significantly improve reliability without departing from the zero-shot paradigm.

An extension of the methodology is to handle ambiguous expressions by returning multiple potential labels with an associated confidence score. This is particularly relevant in cases where the expression x can map to multiple standard concepts, depending on clinical context or usage [52]. Instead of forcing a single label, the system can produce a ranked list, leaving the final decision to a domain expert if needed. This ability to incorporate uncertainty or partial matches underscores the advantage of harnessing large language models that can generate continuous probability distributions over textual outputs, rather than deterministic or rule-based systems. [53]

On the computational side, large language models can be resource-intensive. To mitigate the cost, we often rely on text encoders that share parameters with a generative model but are optimized for retrieval tasks. Alternatively, if we have access to a model with a flexible embedding interface, the entire set of candidate labels can be pre-encoded, and inference can proceed by encoding x with the chosen prompt [54]. This approach is reminiscent of cross-lingual retrieval, where the synergy of embedding-based filtering and text-based generation yields a balance between efficiency and accuracy.

To illustrate, consider a hypothetical scenario in which a hospital system aims to automatically normalize all new patient notes [55]. Each unknown expression x in the note is identified and processed using a standard prompt that includes partial context from the surrounding text. For example: `"Patient complains of 'x' in the left knee. The standardized concept is: ___."`

The model then evaluates the similarity between the generated response and each candidate label in the target set T, focusing on the top K matches [56]. The final label is selected based on the model's ranking, which may be derived from a textual likelihood score or an embedding-based similarity measure.

This process enables real-time normalization of clinical concepts without requiring manual annotation or domain-specific classifiers. In many evaluated scenarios, such an approach demonstrates strong empirical performance—provided that the model is sufficiently robust and has been pre-trained on similar data distributions [57]. This highlights the significant potential of prompt-based zero-shot methods in clinical applications.

4. Experimental Setup and Results

To evaluate the efficacy of zero-shot prompt-based language models for clinical concept normalization, a series of experiments were conducted using diverse clinical datasets [58]. We curated a set of expressions from publicly available resources, including de-identified clinical notes derived from real hospital data. Since these notes naturally contain a variety of abbreviations, misspellings, and domain-specific jargon, they represent a challenging test bed for normalization tasks [59]. Each dataset was mapped to a reference terminology, specifically SNOMED CT, ensuring that each expression x had at least one gold-standard label y. For each expression, we withheld the reference label from the model during zero-shot inference, relying solely on the prompt to guide normalization. [60]

The language model employed was a generative transformer, pre-trained on large biomedical and general-domain corpora. We evaluated multiple prompt designs. The simplest format was: `"Given the expression 'x', the standardized medical concept is: ___."`

A more sophisticated variant appended short textual contexts—often extracted from surrounding sentences in the clinical note—to better replicate real-world usage scenarios [61]. We also experimented with including a single demonstration example, carefully crafted to remain minimal and maintain alignment with the near zero-shot setting.

The candidate label set was the entirety of SNOMED CT, although we often used approximate string matching to limit the search space [62]. For instance, we computed an edit-distance-based filter to narrow potential labels. We tested both a text generation approach, where the model was asked to output the name or code of the concept, and an embedding-based approach, which computed vector representations of x and candidate y labels and picked the best match. [63]

Evaluation metrics included accuracy and macro-averaged F1 scores, computed by comparing the top-ranked label from the model to the true label. In cases where multiple labels were equally valid, the test set was annotated to permit partial credit if the system returned a concept judged clinically equivalent by domain experts. The results showed that zero-shot prompt-based methods consistently achieved competitive performance, often within a few percentage points of specialized supervised classifiers that had been trained on domain-specific corpora [64]. Specifically, the generative approach achieved accuracies ranging from 75 percent to 88 percent, depending on dataset difficulty and prompt design, while embedding-based retrieval topped 90 percent in certain narrower contexts.

A further experiment investigated how well the system could handle ambiguous abbreviations that appear in multiple contexts [65]. For instance, the abbreviation "CP" can refer to chest pain in one context and cerebral palsy in another. We tested the system's sensitivity to small contextual cues embedded in the prompt [66]. When context was excluded, the performance dropped substantially, with a near 40 percent error rate for ambiguous cases. When we included the relevant snippet of text, such as "Patient presents with acute CP and shortness of breath," the error rate decreased significantly, indicating that the model effectively used contextual cues [67]. This finding underscores the importance of carefully crafting the prompt to incorporate local clinical context.

To quantify the contribution of the logic-based constraints, we ran an ablation study. In one variant, we instructed the model using text like "Make sure the concept is valid in SNOMED CT and is a disease or condition, not a finding," effectively restricting the search space to a subset of the ontology [68]. Accuracy improved by 3 to 5 percent, demonstrating the utility of imposing domain constraints at the prompt level. However, we noted that overly restrictive instructions could cause the system to reject valid concepts that did not strictly fit the specified category, suggesting a balance between guided prompting and excessive constraint. [69]

Analysis of error cases revealed several recurring issues. For certain rare expressions, the model defaulted to a more general concept or one that was superficially similar but contextually incorrect [70]. In other instances, the model produced a partially correct concept label that described the overall condition but missed specific qualifiers, such as the stage of a disease. These errors highlight the inherent difficulty of capturing fine-grained medical nuances in a zero-shot setting. Nonetheless, the overall performance gains were sufficient to suggest that prompt-based approaches are a viable alternative to fully supervised pipelines, especially in resource-limited scenarios. [71]

In an additional cross-domain test, we explored how well the model performed on veterinary medicine expressions. While the base language model was trained primarily on human medical text, it still recognized and normalized certain veterinary terms [72]. Accuracy dropped by about 10 percent compared to human medical expressions, reflecting domain shift. However, the approach continued to outperform naive string-matching baselines, suggesting that the learned distributional semantics had partial transferability to a related domain, even without explicit veterinary data [73]. This cross-domain evaluation further cements the hypothesis that large-scale pretraining provides a robust foundation for zero-shot concept normalization, as the underlying language patterns and conceptual structures have broad applicability.

Overall, the experiments validated that zero-shot prompt-based normalization is capable of aligning clinical expressions with standardized terminologies at scale [74]. With carefully engineered prompts and minimal, context-specific cues, the model leverages its extensive pretraining to handle previously unseen expressions or domain-specific abbreviations. The technique effectively addresses the data bottleneck that plagues conventional supervised methods, pointing toward a future where minimal or no labeled data is required to maintain consistent, standardized labeling across diverse medical systems.

5. Discussion

The empirical results highlight both the promise and the limitations of zero-shot prompt-based language models for clinical concept normalization [75]. One of the core insights is that prompt design has a significant impact on accuracy. Minor changes in prompt wording, context inclusion, or instructions can produce substantial fluctuations in performance, underscoring the nuanced interplay between the language model's implicit knowledge and the textual cues provided during inference [76]. This observation suggests that while zero-shot methods drastically reduce the need for annotated training data, they do not entirely eliminate the need for specialized tuning. The art of crafting and refining prompts becomes a new form of model adaptation, with its own set of best practices and pitfalls. [77]

Another prominent factor is the inherent ambiguity of medical language. The experiments showed that abbreviations like "CP" or "HA" demand context for proper disambiguation [78]. While adding local context to the prompt improved results, it also introduced potential confounding factors if the context text was poorly structured or contained contradictory information. Future work may explore dynamic methods of context retrieval, automatically gathering the most relevant preceding and following sentences from a patient's record before constructing the prompt. This approach could be formalized in a retrieval-augmented generation framework, where external knowledge bases or domain-specific databases are queried for relevant medical definitions or synonyms that then guide the prompt design [79].

The error analysis reveals recurring classes of mistakes in which the model conflated related conditions or overlooked specific qualifiers. These errors highlight a tension between the broad distributional knowledge acquired during pretraining and the fine-grained distinctions required for clinical decision support [80, 81].

A potential remedy is to integrate higher-level logical or ontological constraints into the prompt, such that the final concept must belong to a designated sub-hierarchy or satisfy definitional criteria. For example, if an expression refers to a medication, the prompt can instruct the model to select only from labels within the medication branch of the ontology [82].

Alternatively, a post-processing step can be employed to validate the predicted label against known drug categories. This constraint can be formalized as:

$$\forall x,\ \text{isMedication}(x) \to y \in T_{\text{Medication}},$$

where $T_{\text{Medication}} \subseteq T$ represents the subset of medication-related labels within the overall target label set T.

A notable avenue for future research lies in systematically quantifying the effect of domain coverage in the model's pretraining data [83]. The strength of zero-shot normalization depends heavily on whether the language model has encountered semantically similar expressions or textual patterns in its training corpus. If the domain coverage is sparse, the model might default to general medical knowledge, leading to erroneous or generic concepts [84]. This phenomenon was partially observed in the cross-domain veterinary test, where performance declined due to a mismatch between training and inference domains. One remedy could be to fine-tune the model on unlabeled text that is specifically related to the target domain, thereby realigning the distributional space. [85]

Scalability is another consideration. As medical ontologies grow larger, enumerating all candidate labels may become computationally prohibitive [86]. Approximate retrieval methods or hierarchical search strategies can mitigate this issue, but they also introduce potential cascading errors if the top K retrieval set excludes the correct label. Similarly, real-world deployments of zero-shot normalization systems must handle data at high volume and in near-real-time. Optimizations such as caching embeddings or compressing the model's parameters without sacrificing accuracy become essential [87]. Methods like knowledge distillation, where a smaller model is taught to emulate the performance of a larger one, could help maintain throughput while preserving the gains of prompt-based inference.

Despite these challenges, the overarching theme remains that large pre-trained language models contain extensive world knowledge that can be harnessed for clinical concept normalization with minimal overhead [88]. This approach obviates the need for large annotated corpora, which are costly and time-consuming to create, especially when privacy constraints limit data sharing. Zero-shot strategies turn a once specialized, domain-intensive problem into a more generalized prompting challenge that draws on the broad capabilities of foundation models [89]. As these models continue to evolve, incorporating multimodal data or specialized domain expansions, the performance of zero-shot normalization is likely to improve, enabling deeper integration with clinical workflows.

In real-world practice, the decision to adopt a zero-shot system must account for interpretability, legal liability, and clinical governance frameworks. The potential for harm if the system assigns an incorrect label is non-trivial, so safety measures are crucial [90]. These measures might include confidence thresholds that trigger a human review, or a user interface that provides a short justification or reference passage from which the model inferred the label. Future developments in explainable artificial intelligence could offer more transparent justifications, allowing medical personnel to trace back the chain of reasoning from expression x to label y [91]. Formal logic constraints can also add a level of auditable accountability, ensuring that the system only produces labels consistent with established clinical guidelines.

Overall, the discussion confirms that while zero-shot prompt-based normalization is not a panacea, it represents a major step forward in bridging the gap between unstructured clinical text and standardized medical terminologies [92]. Its flexibility and relative ease of deployment make it a powerful tool for medical institutions seeking to reduce the friction of data annotation and streamline the adoption of electronic health record systems. In combination with domain-specific knowledge sources, advanced logic constraints, and robust prompt engineering techniques, it holds promise for achieving high-fidelity normalization that operates reliably across diverse clinical settings. [93]

6. Conclusion

In this paper, we have presented a comprehensive exploration of zero-shot clinical concept normalization using prompt-based language models. By reconfiguring normalization as a prompt-driven inference task, we circumvent the traditional reliance on large-scale labeled datasets, alleviating a critical bottleneck in medical natural language processing. Central to this approach is the recognition that state-of-the-art language models encode both distributional semantics and implicit domain knowledge, enabling them to map colloquial or ambiguous clinical expressions to canonical medical terminologies with surprising accuracy. [94]

We introduced theoretical frameworks grounded in logic-based constraints, compositional semantics, and linear algebraic geometry to elucidate how prompt engineering influences the underlying representations in a way that fosters concept alignment. Empirical results from multiple datasets demonstrated that zero-shot prompt-based normalization can achieve performance levels near or exceeding those of specialized supervised models, underscoring the potency of large language models when properly guided [95]. However, the experiments also highlighted how context, domain coverage, and prompt design significantly modulate outcomes, indicating that responsible deployment must incorporate careful consideration of these factors.

Beyond these findings, our analysis outlined challenges around ambiguity resolution, prompt sensitivity, and domain shift, offering several pathways for refining the method [96]. These include the integration of logic constraints, retrieval-based context expansion, and minimal domain adaptation via unlabeled data. Moreover, we emphasized the importance of transparency, interpretability, and adherence to clinical governance standards, given the high-stakes environment in which concept normalization operates. With continued advances in foundation models and a growing emphasis on data-efficient solutions, we anticipate that prompt-based zero-shot normalization will play a central role in the modernization and automation of clinical documentation, data interoperability, and next-generation healthcare analytics. [97]

References

[1] M. T. Caton, W. F. Wiggins, S. R. Pomerantz, and K. P. Andriole, "The composite severity score for lumbar spine mri: a metric of cumulative degenerative disease predicts time spent on interpretation and reporting," *Journal of digital imaging*, vol. 34, pp. 811–819, 5 2021.

[2] A. Gupta, A. Singh, D. Bharadwaj, and A. K. Mondal, "Humans and robots: A mutually inclusive relationship in a contagious world," *International Journal of Automation and Computing*, vol. 18, pp. 185–203, 1 2021.

[3] K. V. Garvey, K. J. T. Craig, R. G. Russell, L. L. Novak, D. Moore, A. M. Preininger, G. P. Jackson, and B. M. Miller, “The potential and the imperative: the gap in ai-related clinical competencies and the need to close it,” *Medical science educator*, vol. 31, pp. 1–6, 9 2021.

[4] M. Yuan and A. Vlachos, “Zero-shot fact-checking with semantic triples and knowledge graphs,” in *Proceedings of the 1st Workshop on Knowledge Graphs and Large Language Models (KaLLM 2024)*, pp. 105–115, 2024.

[5] T. A. Miller, A. J. McMurry, J. Jones, D. Gottlieb, and K. D. Mandl, “The smart text2fhir pipeline.,” *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2023, pp. 514–, 3 2023.

[6] D. Bertsimas and B. Öztürk, “Global optimization via optimal decision trees,” *Journal of Global Optimization*, 7 2023.

[7] B. Bialecki, J. Park, and M. Tilkin, “Using object storage technology vs vendor neutral archives for an image data repository infrastructure.,” *Journal of digital imaging*, vol. 29, pp. 460–465, 2 2016.

[8] R. M. Reeves, L. M. Christensen, J. R. Brown, M. Conway, M. Levis, G. T. Gobbel, R. U. Shah, C. A. Goodrich, I. Ricket, F. F. Minter, A. R. Bohm, B. E. Bray, M. E. Matheny, and W. W. Chapman, “Adaptation of an nlp system to a new healthcare environment to identify social determinants of health.,” *Journal of biomedical informatics*, vol. 120, pp. 103851–103851, 6 2021.

[9] Y. Tang, Y. Tang, Y. Peng, K. Yan, M. Bagheri, B. Redd, C. Brandon, Z. Lu, M. Han, J. Xiao, and R. M. Summers, “Automated abnormality classification of chest radiographs using deep convolutional neural networks,” *NPJ digital medicine*, vol. 3, pp. 70–70, 5 2020.

[10] X. Li, Y. Wang, D. Wang, W. Yuan, D. Peng, and Q. Mei, “Improving rare disease classification using imperfect knowledge graph,” *BMC medical informatics and decision making*, vol. 19, pp. 1–10, 12 2019.

[11] Z. He, C. Tao, J.-G. Bian, R. Zhang, and J. Huang, “Introduction: selected extended articles from the 2nd international workshop on semantics-powered data analytics (sepda 2017).,” *BMC medical informatics and decision making*, vol. 18, pp. 56–, 7 2018.

[12] S. Yang, H. Xiong, K. Xu, L. Wang, J. Bian, and Z. Sun, “Improving covariance-regularized discriminant analysis for ehr-based predictive analytics of diseases,” *Applied Intelligence*, vol. 51, pp. 377–395, 8 2020.

[13] A. Zhang, L. Xing, J. Zou, and J. C. Wu, “Shifting machine learning for healthcare from development to deployment and from models to data.,” *Nature biomedical engineering*, vol. 6, pp. 1330–1345, 7 2022.

[14] S. Sheikhalishahi, R. Miotto, J. T. Dudley, A. Lavelli, F. Rinaldi, and V. Osmani, “Natural language processing of clinical notes on chronic diseases: Systematic review,” *JMIR medical informatics*, vol. 7, pp. e12239–, 4 2019.

[15] X. Yang, H. Yang, T. Lyu, S. Yang, Y. Guo, J.-G. Bian, H. Xu, and Y. Wu, “Ichi - a natural language processing tool to extract quantitative smoking status from clinical narratives,” *Proceedings. IEEE International Conference on Healthcare Informatics*, vol. 2020, pp. 1–2, 2020.

[16] E. V. Garcia, J. L. Klein, and A. J. Taylor, “Clinical decision support systems in myocardial perfusion imaging.,” *Journal of nuclear cardiology : official publication of the American Society of Nuclear Cardiology*, vol. 21, pp. 427–439, 1 2014.

[17] M. Jaberi-Douraki, E. Meyer, J. E. Riviere, N. I. M. Gedara, J. Kawakami, G. J. Wyckoff, and X. Xu, “Pulmonary adverse drug event data in hypertension with implications on covid-19 morbidity.,” *Scientific reports*, vol. 11, pp. 13349–13349, 6 2021.

[18] L. Wang, Y. Wang, F. Shen, M. Rastegar-Mojarad, and H. Liu, “Discovering associations between problem list and practice setting,” *BMC medical informatics and decision making*, vol. 19, pp. 69–69, 4 2019.

[19] R. Avula, “Assessing the impact of data quality on predictive analytics in healthcare: Strategies, tools, and techniques for ensuring accuracy, completeness, and timeliness in electronic health records,” *Sage Science Review of Applied Machine Learning*, vol. 4, no. 2, pp. 31–47, 2021.

[20] S. M. Meystre, Y. Kim, G. T. Gobbel, M. E. Matheny, A. Redd, B. E. Bray, and J. H. Garvin, “Congestive heart failure information extraction framework for automated treatment performance measures assessment.,” *Journal of the American Medical Informatics Association : JAMIA*, vol. 24, pp. e40–e46, 7 2016.

[21] S. G. Finlayson, P. LePendu, and N. H. Shah, “Building the graph of medicine from millions of clinical narratives.,” *Scientific data*, vol. 1, pp. 140032–140032, 9 2014.

[22] I. Spasic, Özlem Uzuner, and L. Zhou, "Emerging clinical applications of text analytics.," *International journal of medical informatics*, vol. 134, pp. 103974–103974, 9 2019.

[23] J. R. Machireddy, "Harnessing ai and data analytics for smarter healthcare solutions," *International Journal of Science and Research Archive*, vol. 08, no. 02, pp. 785–798, 2023.

[24] R. D. Parker, M. Abram, and K. Mancini, "Using natural language processing to understand unstructured healthcare data," *SSRN Electronic Journal*, 1 2022.

[25] M. Wankhade, C. S. R. Annavarapu, and A. Abraham, "Mapa bilstm-bert: multi-aspects position aware attention for aspect level sentiment analysis," *The Journal of Supercomputing*, vol. 79, pp. 11452–11477, 2 2023.

[26] Q. Li, S. A. Spooner, M. Kaiser, N. Lingren, J. Robbins, T. Lingren, H. Tang, I. Solti, and Y. Ni, "An end-to-end hybrid algorithm for automated medication discrepancy detection," *BMC medical informatics and decision making*, vol. 15, pp. 37–37, 5 2015.

[27] R. S. Evans, J. Benuzillo, B. D. Horne, J. F. Lloyd, A. Bradshaw, D. Budge, K. Rasmusson, C. Roberts, J. Buckway, N. Geer, T. Garrett, and D. L. Lappé, "Automated identification and predictive tools to help identify high-risk heart failure patients: pilot evaluation," *Journal of the American Medical Informatics Association : JAMIA*, vol. 23, pp. 872–878, 2 2016.

[28] A. Sharma and K. Forbus, "Modeling the evolution of knowledge in learning systems," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, pp. 669–675, 2012.

[29] A. K. Saxena, "Evaluating the regulatory and policy recommendations for promoting information diversity in the digital age," *International Journal of Responsible Artificial Intelligence*, vol. 11, no. 8, pp. 33–42, 2021.

[30] B. K. Bellows, S. L. DuVall, A. W. C. Kamauu, D. Supina, T. Babcock, and J. LaFleur, "Healthcare costs and resource utilization of patients with binge-eating disorder and eating disorder not otherwise specified in the department of veterans affairs.," *The International journal of eating disorders*, vol. 48, pp. 1082–1091, 5 2015.

[31] Y. Zhang, C. Tao, Y. Gong, K. Wang, and Z. Zhao, "The international conference on intelligent biology and medicine 2018: Medical informatics thematic track (medicalinfo2018).," *BMC medical informatics and decision making*, vol. 19, pp. 21–21, 1 2019.

[32] E. Massaad and P. Cherfan, "Social media data analytics on telehealth during the covid-19 pandemic," *Cureus*, vol. 12, pp. e7838–, 4 2020.

[33] Y. C. Choi, D. Zhang, and J. E. Tyczynski, "Comparison between health insurance claims and electronic health records (ehrs) for metastatic non-small-cell lung cancer (nsclc) patient characteristics and treatment patterns: A retrospective cohort study," *Drugs - real world outcomes*, vol. 8, pp. 577–587, 8 2021.

[34] J. Li and Z. Lu, "Automatic identification and normalization of dosage forms in drug monographs," *BMC medical informatics and decision making*, vol. 12, pp. 9–9, 2 2012.

[35] N. Kapoor, R. Lacson, L. Cochon, M. M. Hammer, I. K. Ip, G. W. Boland, and R. Khorasani, "Radiologist variation in the rates of follow-up imaging recommendations made for pulmonary nodules.," *Journal of the American College of Radiology : JACR*, vol. 18, pp. 896–905, 2 2021.

[36] R. Ghosh, D. Kempf, A. Pufko, L. F. B. Martinez, C. M. Davis, and S. Sethi, "Automation opportunities in pharmacovigilance: An industry survey," *Pharmaceutical medicine*, vol. 34, pp. 7–18, 2 2020.

[37] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Federated query processing for big data in data science," in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 6145–6147, IEEE, 2019.

[38] S. Wunnava, X. Qin, T. Kakar, C. Sen, E. A. Rundensteiner, and X. Kong, "Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding," *Drug safety*, vol. 42, pp. 113–122, 1 2019.

[39] J. Yuan, Z. Jin, H. Guo, H. Jin, X. Zhang, T. Smith, and J. Luo, "Constructing biomedical domain-specific knowledge graph with minimum supervision," *Knowledge and Information Systems*, vol. 62, pp. 317–336, 3 2019.

[40] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Context-aware query performance optimization for big data analytics in healthcare," in *2019 IEEE High Performance Extreme Computing Conference (HPEC-2019)*, pp. 1–7, 2019.

[41] J. H. Caufield, Y. Zhou, A. O. Garlid, S. P. Setty, D. A. Liem, Q. Cao, J. M. Lee, S. Murali, S. Spendlove, W. Wang, L. Zhang, Y. Sun, A. A. T. Bui, H. Hermjakob, K. E. Watson, and P. Ping, "A reference set of curated biomedical data and metadata from clinical case reports," *Scientific data*, vol. 5, pp. 180258–180258, 11 2018.

[42] D. A. Sippo, G. I. Warden, K. P. Andriole, R. Lacson, I. Ikuta, R. L. Birdwell, and R. Khorasani, "Automated extraction of bi-rads final assessment categories from radiology reports with natural language processing.," *Journal of digital imaging*, vol. 26, pp. 989–994, 7 2013.

[43] A. H. Zadeh, H. M. Zolbanin, R. Sharda, and D. Delen, "Social media for nowcasting flu activity: Spatio-temporal big data analysis," *Information Systems Frontiers*, vol. 21, pp. 743–760, 1 2019.

[44] B. Zaidat, J. Tang, V. Arvind, E. A. Geng, B. Cho, A. H. Duey, C. Dominy, K. D. Riew, S. K. Cho, and J. S. Kim, "Can a novel natural language processing model and artificial intelligence automatically generate billing codes from spine surgical operative notes?," *Global spine journal*, vol. 14, pp. 2022–2030, 3 2023.

[45] S. Prasanna, N. Premnath, S. Angraal, R. Sedhom, R. Khera, H. Parsons, S. M. Q. Hussaini, P. T. Johnson, E. Lou, M. S. Beg, I. M. Subbiah, P. Rao, and A. Gupta, "Sentiment analysis of tweets on prior authorization.," *Journal of Clinical Oncology*, vol. 39, pp. 322–322, 10 2021.

[46] M. Arenson, J. Hogan, L. Xu, R. Lynch, H. Lee, J. Choi, J. Sun, A. Adams, and R. Patzer, "215.3: Prediction of post-transplant hospitalization among kidney transplant recipients with clinical notes and electronic healthcare record data," *Transplantation*, vol. 106, pp. S41–S41, 9 2022.

[47] C. Zheng, M.-S. Lee, N. Bansal, A. S. Go, C. Chen, T. N. Harrison, D. Fan, A. Allen, E. Garcia, B. Lidgard, D. Singer, and J. An, "Identification of recurrent atrial fibrillation using natural language processing applied to electronic health records.," *European heart journal. Quality of care & clinical outcomes*, vol. 10, pp. 77–88, 3 2023.

[48] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Q. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of big data*, vol. 8, pp. 1–74, 3 2021.

[49] A. Tabaie, E. W. Orenstein, S. Kandaswamy, and R. Kamaleswaran, "Integrating structured and unstructured data for timely prediction of bloodstream infection among children.," *Pediatric research*, vol. 93, pp. 969–975, 7 2022.

[50] D. Xu, C. Wang, A. Khan, N. Shang, Z. He, A. S. Gordon, I. J. Kullo, S. N. Murphy, Y. Ni, W.-Q. Wei, A. G. Gharavi, K. Kiryluk, C. Weng, and I. Ionita-Laza, "Quantitative disease risk scores from ehr with applications to clinical risk stratification and genetic studies.," *NPJ digital medicine*, vol. 4, pp. 116–, 7 2021.

[51] M. Muniswamaiah, T. Agerwala, and C. Tappert, "Data virtualization for analytics and business intelligence in big data," in *CS & IT Conference Proceedings*, vol. 9, CS & IT Conference Proceedings, 2019.

[52] A. Sharma, *Structural and network-based methods for knowledge-based systems*. PhD thesis, Northwestern University, 2011.

[53] S. Chowdhury, X. Dong, L. Qian, X. Li, Y. Guan, J. Yang, and Q. Yu, "A multitask bi-directional rnn model for named entity recognition on chinese electronic medical records.," *BMC bioinformatics*, vol. 19, pp. 499–499, 12 2018.

[54] E. Gustafson, J. A. Pacheco, F. Wehbe, J. I. Silverberg, and W. K. Thompson, "Ichi - a machine learning algorithm for identifying atopic dermatitis in adults from electronic health records," *Proceedings. IEEE International Conference on Healthcare Informatics*, vol. 2017, pp. 83–90, 9 2017.

[55] F. Shen, D. W. Larson, J. M. Naessens, E. B. Habermann, H. Liu, and S. Sohn, "Detection of surgical site infection utilizing automated feature generation in clinical notes," *Journal of healthcare informatics research*, vol. 3, pp. 267–282, 11 2018.

[56] X. Yang, J.-G. Bian, Y. Gong, W. R. Hogan, and Y. Wu, "Madex: A system for detecting medications, adverse drug events, and their relations from clinical notes," *Drug safety*, vol. 42, pp. 123–133, 1 2019.

[57] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, vol. 19, pp. 171–209, 1 2014.

[58] L. Wang, Z. Xue, C. F. Ezeana, M. Puppala, S. Chen, R. L. Danforth, X. Yu, T. He, M. L. Vassallo, and S. T. C. Wong, "Preventing inpatient falls with injuries using integrative machine learning prediction: a cohort study.," *NPJ digital medicine*, vol. 2, pp. 127–127, 12 2019.

[59] W. He, T. M. Scharadin, M. Saldana, C. Gellner, S. Hoang-Phou, C. L. Takanishi, G. L. Hura, J. A. Tainer, K. L. Carraway, P. T. Henderson, and M. A. Coleman, "Cell-free expression of functional receptor tyrosine kinases," *Scientific reports*, vol. 5, pp. 12896–12896, 8 2015.

[60] K. Pothemont, S. Quinton, M. Jayoushe, S. Jedel, A. Bedell, S. B. Hanauer, E. Mutlu, and T. Taft, "Patient perspectives on medical trauma related to inflammatory bowel disease," *Journal of clinical psychology in medical settings*, vol. 29, pp. 1–12, 7 2021.

[61] A. Zirikly, B. Desmet, D. Newman-Griffis, E. E. Marfeo, C. McDonough, H. Goldman, and L. Chan, "Information extraction framework for disability determination using a mental functioning use-case.," *JMIR medical informatics*, vol. 10, pp. e32245–e32245, 3 2022.

[62] B. I. Reiner, "Strategies for medical data extraction and presentation part 2: Creating a customizable context and user-specific patient reference database," *Journal of digital imaging*, vol. 28, pp. 249–255, 4 2015.

[63] A. E. Krumm, E. Ötleş, K. M. Marcotte, B. L. Spencer, S. Izadi, B. C. George, and B. Zendejas, "Strategies for evaluating predictive models: examples and implications based on a natural language processing model used to assess operative performance feedback," *Global Surgical Education - Journal of the Association for Surgical Education*, vol. 3, 12 2023.

[64] N. S. Zheng, V. E. Kerchberger, V. A. Borza, H. N. Eken, J. C. Smith, and W.-Q. Wei, "An updated, computable medication-indication resource for biomedical research.," *Scientific reports*, vol. 11, pp. 18953–, 9 2021.

[65] W. Branch-Elliman, J. Strymish, V. Kudesia, A. K. Rosen, and K. Gupta, "Natural language processing for real-time catheter-associated urinary tract infection surveillance: Results of a pilot implementation trial.," *Infection control and hospital epidemiology*, vol. 36, pp. 1004–1010, 5 2015.

[66] T. Cai, L. Zhang, N. Yang, K. K. Kumamaru, F. J. Rybicki, T. Cai, and K. P. Liao, "Extraction of emr numerical data: an efficient and generalizable tool to extend clinical research.," *BMC medical informatics and decision making*, vol. 19, pp. 1–7, 11 2019.

[67] Y. Zhao, J. Zhang, and M. Wu, "Finding users' voice on social media: An investigation of online support groups for autism-affected users on facebook," *International journal of environmental research and public health*, vol. 16, pp. 4804–, 11 2019.

[68] R. Jiang, S. Gulbranson, M. Bhattacharya, F. Lievano, P. B. Komarnitsky, M. M. Koch, R. Potluri, and J. E. Tyczynski, "Abstract 4228: Incidence of pericardial and pleural effusions in patients with small cell lung cancer treated with third-line or later-line therapy: An analysis of electronic health records data," *Cancer Research*, vol. 78, pp. 4228–4228, 7 2018.

[69] Z. He, J.-G. Bian, C. Tao, and R. Zhang, "Selected articles from the third international workshop on semantics-powered data analytics (sepda 2018).," *BMC medical informatics and decision making*, vol. 19, pp. 148–148, 8 2019.

[70] C. A. McCarty, R. L. Chisholm, C. G. Chute, I. J. Kullo, G. P. Jarvik, E. B. Larson, R. Li, D. R. Masys, M. D. Ritchie, D. M. Roden, J. P. Struewing, and W. A. Wolf, "The emerge network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies.," *BMC medical genomics*, vol. 4, pp. 13–13, 1 2011.

[71] C. K. M. Orr, A. Bauza, P. D. Langer, M. A. Zarbin, and N. Bhagat, "Open-globe injuries with motor vehicle accidents: a 12-year review," *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie*, vol. 253, pp. 1313–1317, 1 2015.

[72] A. Sharma and K. D. Forbus, "Graph-based reasoning and reinforcement learning for improving q/a performance in large knowledge-based systems," in *2010 AAAI Fall Symposium Series*, 2010.

[73] L. Wang, G. Wang, and C. A. Alexander, "Natural language processing systems and big data analytics," *International Journal of Computational Systems Engineering*, vol. 2, no. 2, pp. 76–, 2015.

[74] D. Zhang, C. Yin, J. Zeng, X. Yuan, and P. Zhang, "Combining structured and unstructured data for predictive models: a deep learning approach," *BMC medical informatics and decision making*, vol. 20, pp. 280–280, 10 2020.

[75] R. Avula *et al.*, "Data-driven decision-making in healthcare through advanced data mining techniques: A survey on applications and limitations," *International Journal of Applied Machine Learning and Computational Intelligence*, vol. 12, no. 4, pp. 64–85, 2022.

[76] A. Medici, A. Marshall-Colon, E. Ronzier, W. Szponarski, R. Wang, A. Gojon, N. M. Crawford, S. Ruffel, G. M. Coruzzi, and G. Krouk, "Atnigt1/hrs1 integrates nitrate and phosphate signals at the arabidopsis root tip.," *Nature communications*, vol. 6, pp. 6274–6274, 2 2015.

[77] R. E. Greenblatt, E. Zhao, S. E. Henrickson, A. J. Apter, R. A. Hubbard, and B. E. Himes, "Factors associated with exacerbations among adults with asthma according to electronic health record data," *Asthma research and practice*, vol. 5, pp. 1–11, 1 2019.

[78] A. K. Saxena, R. R. Dixit, and A. Aman-Ullah, “An lstm neural network approach to resource allocation in hospital management systems,” *International Journal of Applied Health Care Analytics*, vol. 7, no. 2, pp. 1–12, 2022.

[79] R. Tang, F. Acevedo, C. R. Lanahan, S. B. Coopey, A. Yala, R. Barzilay, C. Li, A. S. Colwell, A. J. Guidi, C. L. Cetrulo, J. Garber, B. L. Smith, M. A. Gadd, M. C. Specht, and K. S. Hughes, “Incidental breast carcinoma: incidence, management, and outcomes in 4804 bilateral reduction mammoplasties.,” *Breast cancer research and treatment*, vol. 177, pp. 741–748, 7 2019.

[80] S. R. Earnshaw and S. L. Dennett, “Integer/linear mathematical programming models: a tool for allocating healthcare resources.,” *PharmacoEconomics*, vol. 21, no. 12, pp. 839–851, 2003.

[81] J. R. Machireddy, “Automation in healthcare claims processing: Enhancing efficiency and accuracy,” *International Journal of Science and Research Archive*, vol. 09, no. 01, pp. 825–834, 2023.

[82] C.-Y. Chen, P. Lee, V. M. Castro, J. Minnier, S. N. Murphy, V. S. Gainer, T. Cai, P. Sklar, R. H. Perlis, and J. W. Smoller, “Genetic validation of bipolar disorder identified by automated phenotyping using electronic health records,” *European Neuropsychopharmacology*, vol. 29, pp. S968–S969, 2019.

[83] C. Dominy, V. Arvind, J. E. Tang, C. P. Bellaire, S. D. Pasik, J. S. Kim, and S. K. Cho, “Scoliosis surgery in social media: a natural language processing approach to analyzing the online patient perspective.,” *Spine deformity*, vol. 10, pp. 239–246, 10 2021.

[84] V. G. V. Vydiswaran, Y. Zhang, Y. Wang, and H. Xu, “Special issue of bmc medical informatics and decision making on health natural language processing,” *BMC medical informatics and decision making*, vol. 19, pp. 76–76, 4 2019.

[85] C. Xiang and M. Abouelyazid, “The impact of generational cohorts and visit environment on telemedicine satisfaction: A novel investigation,” 2020.

[86] Y. Zhao, N. J. Fesharaki, H. Liu, and J. Luo, “Using data-driven sublanguage pattern mining to induce knowledge models: application in medical image reports knowledge representation,” *BMC medical informatics and decision making*, vol. 18, pp. 61–61, 7 2018.

[87] K. D. Forbus, C. Riesbeck, L. Birnbaum, K. Livingston, A. B. Sharma, and L. C. Ureel, “A prototype system that learns by reading simplified texts.,” in *AAAI Spring Symposium: Machine Reading*, pp. 49–54, 2007.

[88] J. Wang, J. S. de Vale, S. Gupta, P. Upadhyaya, F. A. Lisboa, S. A. Schobel, E. A. Elster, C. J. Dente, T. G. Buchman, and R. Kamaleswaran, “Clotcatcher: a novel natural language model to accurately adjudicate venous thromboembolism from radiology reports.,” *BMC medical informatics and decision making*, vol. 23, pp. 262–, 11 2023.

[89] D. Valdez and P. Goodson, “Neutral or framed? a sentiment analysis of 2019 abortion laws.,” *Sexuality research & social policy : journal of NSRC : SR & SP*, vol. 19, pp. 936–945, 1 2022.

[90] H. Tian, Y. Tao, S. Pouyanfar, S.-C. Chen, and M.-L. Shyu, “Multimodal deep representation learning for video classification,” *World Wide Web*, vol. 22, pp. 1325–1341, 5 2018.

[91] L. Zhou, J. M. Plasek, L. M. Mahoney, F. Y. Chang, D. Dimaggio, and R. A. Rocha, “Mapping partners master drug dictionary to rxnorm using an nlp-based approach,” *Journal of biomedical informatics*, vol. 45, pp. 626–633, 11 2011.

[92] J. Song, M. Ojo, K. H. Bowles, M. V. McDonald, K. Cato, S. C. Rossetti, V. Adams, S. Chae, M. Hobensack, E. Kennedy, A. Tark, M.-J. Kang, K. Woo, Y. Barrón, S. Sridharan, and M. Topaz, “Detecting language associated with home healthcare patient’s risk for hospitalization and emergency department visit.,” *Nursing research*, vol. 71, pp. 285–294, 2 2022.

[93] B. Kim, M. Merchant, C. Zheng, A. A. Thomas, R. Contreras, S. J. Jacobsen, and G. Chien, “Second prize: A natural language processing program effectively extracts key pathologic findings from radical prostatectomy reports,” *Journal of endourology*, vol. 28, no. 12, pp. 1474–1478, 2014.

[94] R. Avula, “Optimizing data quality in electronic medical records: Addressing fragmentation, inconsistencies, and data integrity issues in healthcare,” *Journal of Big-Data Analytics and Cloud Computing*, vol. 4, no. 5, pp. 1–25, 2019.

[95] E. Bertino, “Security of cellular networks position paper,” *CCF Transactions on Networking*, vol. 3, pp. 93–96, 7 2020.

[96] K. D. Forbus, K. Lockwood, A. B. Sharma, and E. Tomai, “Steps towards a second generation learning by reading system.,” in *AAAI Spring Symposium: Learning by Reading and Learning to Read*, pp. 36–43, 2009.

[97] Q. He and E. Agu, “Context-aware probabilistic models for predicting future sedentary behaviors of smartphone users,” *Journal of healthcare informatics research*, vol. 6, pp. 1–41, 11 2021.