Knowledge Aware Automated Health Claims Processing with Medical Ontologies and Large Language Models

Sheng Jie Lui[1], **Cheng Xiang**[1], **Shonali Krishnaswamy**[2]
[1]National University of Singapore
[2]AIDA Technologies
luishengjie@u.nus.edu, elexc@nus.edu.sg, shonali@aidatech.io

Abstract

One of the major challenges in automating health insurance claims processing lies in the complexity involved in validating an incoming claim's medical diagnoses against its policy Underwriting (UW) exclusions. Termed UW Exclusion Detection, this process ensures claims are only paid out if their diagnoses are not medically associated with conditions excluded under the policy. Medical diagnoses in health insurance claims are typically represented by the International Classification of Disease (ICD) codes, established by the World Health Organization. For example, given a policy that excludes "all respiratory illness". A claim with the ICD code J45 (Asthma) will be subject to rejection as J45 is a respiratory-related diagnosis that falls within the scope of the policy's UW exclusion. The key challenge in automating this process lies in the wide range of available ICD codes. The ICD-10-CM coding scheme consists of over 40,000 codes, which often results in scenarios where codes encountered during inference are absent from the training data. These unseen ICD codes limit the effectiveness of data-driven approaches, which depend on the training data to discern medically relevant associations between UW exclusions and ICD codes. This underscores the need to supplement data-driven approaches with additional domain knowledge. We hypothesize that integrating implicit medical domain knowledge inherent in Large Language Models (LLMs) with explicit domain knowledge from medical ontologies, will enhance data-driven approaches for UW Exclusion Detection. Thoroughly validated on real-world health insurance claims data, our proposed approach proved effective in accurately establishing medically relevant associations between UW exclusions and ICD codes.

Introduction

Traditionally, processing health insurance claims has been a manual and time-consuming task. This process requires assessors to meticulously review claim details like diagnoses, medical procedures, incurred amount, and other relevant information, against policy details and historical claim records to ascertain their validity. Recognizing the need to streamline this process, health insurance organizations are increasingly adopting automated claims processing solutions. These systems automatically process routinely approved claims, allowing assessors to focus their resources on more complex cases. In the context of automated health insurance claims processing, Straight-Through Processing (STP), refers to the automatic approval of valid claims without manual intervention. A high STP rate minimizes the need for human oversight, expediting both the processing and reimbursement of claims. This ultimately enhances the overall customer experience.

To achieve STP, a claim must undergo a series of critical checks, where Underwriting (UW) Exclusion is particularly crucial. This process identifies claims whose diagnoses are associated with medical conditions specified in the policy's UW exclusion. As these diagnoses are not covered by the policy, they are directed to manual review where a human assessor may potentially reject them. Medical diagnoses in health insurance claims are typically documented using the International Classification of Disease (ICD) codes, developed by the World Health Organization (WHO) (WHO 2004). The ICD-10-CM codes, recognized as an industry standard, consists of over 40,000 unique codes, updated bi-annually (CMS 2023). This facilitates the comprehensive categorization of medical diagnoses. UW exclusions, denoted in free-text, specify medical conditions not covered by the policy. For example, given a policy with the UW exclusion "Exclude all eye-related diseases", all subsequent claims with eye-related disorders like H25 (Age-related cataract) or H25 (Glaucoma) will be excluded. Table 1 illustrates how the relationship between UW exclusions and ICD codes influences the claim status.

In recent years, data-driven machine learning (ML) approaches have been adopted by insurance organizations to streamline claims processing and enhance fraud detection capabilities (Jones and Sah 2023). However, conventional data-driven ML algorithms often underperform in real-world UW Exclusion Detection (Lui, Xiang, and Krishnaswamy 2024). The key challenges in automating UW Exclusion Detection utilizing data-driven ML approaches stem from:

- **New and Unseen ICD Codes**: The extensive range of available ICD codes often results in scenarios where codes encountered during inference are not present in the training dataset. For instance, if the ICD code $U07.1$ (COVID-19) is not available during training but frequently encountered during inference, data-driven ML models may struggle to accurately establish relevant associations for such codes.

UW Exclusion	ICD Code	Claim Status	Reason
Exclude all eye-related diseases.	H25 (Age-related cataract)	Rejected	H25 is an eye-related disease.
Exclude all eye-related diseases.	H40 (Glaucoma)	Rejected	H40 is an eye-related disease.
Exclude all eye-related diseases.	H81.0 (Meniere's disease)	Not Rejected	H81.0 results from fluid build-up in the inner ear chambers. It is not an eye-related disease.

Table 1: Examples illustrating the relationship between UW Exclusions and ICD codes in the context of UW Exclusion Detection: a claim is rejected if the UW Exclusion is associated with the corresponding ICD code.

- **Medical Synonyms and Abbreviations**: Different descriptions in free-text UW exclusions can represent similar medical conditions. For example, UW exclusions for 'myocardial infarction' and 'heart attack' are interchangeable and are both associated to ICD code I21.9. If the training data lacks diverse synonyms and abbreviations alongside their corresponding ICD codes, ML models may struggle to accurately capture their equivalence. This discrepancy may potentially result in underperformance during inference.

Additionally, linguistic challenges, such as negation terms may further complicate the interpretation of UW exclusions. For example, an UW exclusion that "Excludes all eye diseases except cataract" should not result in the rejection of claims with ICD codes like H26 (Other cataract). If the training datasets do not adequately capture these complexities, ML models may struggle to accurately interpret these complex associations.

These complexities limit conventional data-driven approaches from effectively learning the necessary associations between UW exclusions and ICD codes for UW Exclusion Detection, resulting in their underperformance. This underscores the critical need to supplement data-driven approaches with additional domain knowledge to effectively establish medically relevant associations between UW exclusions and ICD codes. In this work, we categorize domain knowledge into two distinct categories:

1. **Implicit Domain Knowledge**: This refers to domain knowledge embedded within models trained on comprehensive datasets. For instance, Large Language Models (LLMs), trained on extensive datasets that include medical literature, offer a promising solution due to their ability to interpret medical terminologies in a zero-shot and few-shot manner (Liu et al. 2023).
2. **Explicit Domain Knowledge**: This refers to verified sources of knowledge that have been curated by domain experts. For example, medical ontologies like the Disease Ontology (DO) (Baron et al. 2023) provide a comprehensive categorization of disease and their associations.

We hypothesize that supplementing data-driven ML approaches with implicit domain knowledge inherent in LLMs and explicit domain knowledge from medical ontologies, will facilitate more accurate inferences, that extend beyond the scope of the training data. This addresses the limitations faced by conventional data-driven approaches for UW Exclusion Detection.

In this work, we present KOAAL (**K**nowledge aware **O**ntology **A**lignment and **A**ccess with **L**anguage models), a novel framework that aligns disjoint medical ontologies into a unified reference ontology. KOAAL harnesses the domain knowledge encapsulated within the reference ontology for UW Exclusion Detection. To achieve this, KOAAL integrates the implicit domain knowledge embedded in LLMs, explicit domain knowledge inherent in medical ontologies, and user feedback from labeled health insurance claims data into a unified and comprehensive reference ontology. During inference, KOAAL leverages the domain knowledge encapsulated in the reference ontology. It extracts relevant medical terms from the UW exclusion, and map these terms, along with the ICD codes, to their corresponding nodes within the reference ontology. This allows the association between UW exclusions and ICD codes to be clearly established through the structural properties of the reference ontology. In this work, we demonstrate the efficacy and practical utility of the KOAAL framework for UW Exclusion Detection, highlighting three major contributions:

1. **Align Medical Ontologies with Domain Knowledge from LLMs**: We introduce a novel approach that transforms implicit domain knowledge from LLMs into an explicit format. This facilitates the establishment of complex inter- and intra-relationships between nodes across distinct ontologies. By aligning multiple ontologies into a unified and comprehensive reference ontology, our approach facilitates the establishment of complex associations between UW exclusions and ICD codes.
2. **Integrate User Feedback via Edge Weights**: We propose an approach that integrates feedback from claims assessors, captured through labeled health insurance claims data, into our reference ontology as edge weights. This approach minimizes the need for frequent model retraining and facilitates the fine-tuning of our reference ontology based on client-specific requirements.
3. **Utilize Language Models to Harness Domain Knowledge from Reference Ontology**: We introduce an approach that leverages language models, like LLMs and SBioBERT (Lui, Xiang, and Krishnaswamy 2024), to harness the domain knowledge stored in our reference ontology for UW Exclusion Detection. First, an LLM is utilized to extract valid medical terms from UW exclu-

sions, addressing the challenges associated with medical synonyms, abbreviations and negation terms. Next, a fine-tuned SBioBERT encoder is employed to map the extracted terms to their corresponding nodes within the reference ontology, based on their medical concepts. This mapping facilitates the establishment of medically relevant associations between UW exclusions and ICD codes through interconnected paths within the reference ontology. It enhances explainability by explicitly delineating the relationships that link UW exclusions to ICD codes, providing clear and traceable insights into their connections.

We conduct extensive experimental evaluations on two real-world health insurance claims datasets to validate the effectiveness of our proposed KOAAL framework. When integrated into deployed automated health insurance claims processing systems, KOAAL demonstrated enhanced capabilities for UW Exclusion Detection. It effectively establishes complex associations between UW exclusions and ICD codes that had been overlooked by previous implementations. This underscores the real-world efficacy and practical utility of the KOAAL framework.

The remainder of this paper is organized as follows: We begin by providing an overview of prior approaches for UW Exclusion Detection. Next, we introduce our proposed KOAAL framework, which leverages the domain knowledge from LLMs, medical ontologies, and feedback from claim assessors for UW Exclusion Detection. Following this, we experimentally validate KOAAL using real-world health insurance claims data and discuss the practical implications of deploying our framework. Lastly, we conclude the paper by summarizing our key findings and outlining potential directions for future research.

Related Work

This section reviews prior work that establishes medically relevant associations between free-text medical documents and ICD codes focusing on the application use case of UW Exclusion Detection. The majority of existing research that maps medical text to ICD codes treats this task as a multi-label classification challenge. However, these studies often restrict their analysis to the most frequently occurring ICD codes (Mullenbach et al. 2018; Wang et al. 2020; Pascual, Luck, and Wattenhofer 2021), neglecting the challenges posed by unseen ICD codes.

In recent years, the adoption of pre-trained language models (PLMs) have gained significant attention. When fine-tuned, PLMs excel in assimilating domain-specific knowledge. Notably, variants of BERT (Devlin et al. 2019) like BioBERT (Lee et al. 2020) and BioClinicalBERT (Alsentzer et al. 2019) that were fine-tuned on bio-medical datasets, demonstrated improved performance in recognizing medical entities within electronic health records (EHRs). These models outperformed the baseline BERT(Devlin et al. 2019) model, trained on generic datasets from Wikipedia (Turchin, Masharsky, and Zitnik 2023). While PLMs, when fine-tuned on domain-specific datasets, excel in specialized applications, they often struggle with health insurance claims data, which frequently includes unseen ICD codes (Lui, Xiang, and Krishnaswamy 2024). The absence of these codes in the training data impedes the model's ability to establish relevant associations, resulting in underperformance in tasks such as UW Exclusion Detection.

LLMs like GPT 3.5 (OpenAI 2023), trained on broader and more diversified datasets, can interpret a wide range of medical terminologies. However, these models are susceptible to model hallucination, where they may generate non-existent medical codes (Ong et al. 2023; Zhou et al. 2023). Such errors could inadvertently result in the approval of unnecessary treatments (Pal, Umapathi, and Sankarasubbu 2023). Additionally, the specifics of the training datasets, particularly regarding medical codes, remain undisclosed for models like GPT-3.5 (OpenAI 2023) and LLaMA (Touvron et al. 2023). This lack of transparency raises concerns about the accuracy and relevance of the medical content these models are trained on. Another major barrier in adopting LLMs for real-world applications is the substantial cost and computational resources required to fine-tune these models with additional information (Zhou et al. 2023). Moreover, fine-tuning an LLM post-training can lead to challenges like catastrophic forgetting (Luo et al. 2023). This poses a significant challenge as medical knowledge continuously evolves, necessitating frequent updates to maintain its accuracy and relevance.

Medical ontologies, typically structured as Directed Acyclic Graphs (DAGs), provide a verified and systematic representation of medical terms, diagnoses, and their interconnections. For example, medical ontologies such as SNOMED CT (SNOMED International 2023), RxNorm (Liu et al. 2005), and the Disease Ontology (DO) (Baron et al. 2023) provide a comprehensive list of medical terminologies, characterizing various medical diagnoses (Sahoo et al. 2022; Kulmanov et al. 2020). Additionally, these ontologies facilitate the incorporation of domain knowledge into ML models, enriching them with additional contextual knowledge (Kulmanov et al. 2020). For instance, the Neural Concept Recognizer (NCR) (Arbabi et al. 2019) introduces an unsupervised Named Entity Recognition (NER) approach that aligns medical terms in free-text documents with medical concepts from a reference ontology.

To address the challenges associated with unseen ICD codes in UW Exclusion Detection, the KAMEL framework (Lui, Xiang, and Krishnaswamy 2024), was introduced. KAMEL comprises two key components: (i) an implicit domain knowledge inference component, which utilizes a sentence encoder fine-tuned on health insurance claims data to discern the similarity between UW exclusions and ICD codes based on their textual descriptions; (ii) an explicit domain knowledge inference component, which leverages a reference ontology to establish medically relevant connections between UW exclusions and ICD codes through their interconnected paths. The responses from both components are aggregated to obtained the final prediction. This work also introduces the ICD Ontology (ICDO), which represents all currently available ICD codes as a DAG. This structure enables the KAMEL framework to make more accurate inference on unseen ICD codes for UW Exclusion Detection.

However, despite demonstrating significant improvements in precision when evaluated against data-driven multi-label classifiers, KAMEL exhibits low recall scores. This underperformance stems from:

1. Overfitting Due to Limited Data: The limited number of UW exclusion and ICD code pairs available during training limits the sentence encoder's ability to generalize to new data.
2. Limitations of Sentence Encoder for Medical Concept Alignment: UW exclusions are segmented into key phrases, which are mapped to medical concepts within the ontology. However, as the sentence encoder is not optimized for such mappings, it struggles to accurately align medical terms with ontology nodes based on their associated medical concepts.
3. Disjoint Sources of Explicit Domain Knowledge: KAMEL utilizes the DO (Baron et al. 2023) and introduces the ICDO as sources of explicit domain knowledge. However, these ontologies are represented as separate, unconnected entities. This impedes the framework's ability to accurately establish relevant associations when nodes related to UW exclusions and ICD codes are distributed across these two disjoint ontologies.

In the subsequent section, we demonstrate how our proposed approach, KOAAL, addresses these limitations by utilizing the same reference ontologies, DO and ICDO, to establish more complex associations between UW exclusions and ICD codes. This exemplifies its enhanced capabilities in harnessing the domain knowledge encapsulated within these ontologies for UW Exclusion Detection.

KOAAL Framework Overview

In this section, we present KOAAL, a domain-knowledge-driven framework for UW Exclusion Detection. KOAAL integrates implicit domain knowledge from LLMs, explicit domain knowledge from medical ontologies, and feedback from claim assessors into a comprehensive reference ontology. To effectively harness the domain knowledge encapsulated within the reference ontology for UW Exclusion Detection, KOAAL extracts relevant medical terms from UW exclusions and map the extracted terms and ICD codes from the health insurance claim to their corresponding nodes within the reference ontology using a fine-tuned sentence encoder and node attributes, respectively. By representing UW exclusions and ICD codes as nodes within the reference ontology, KOAAL establishes their associations based on the structural properties of the ontology.

The KOAAL framework addresses key limitations of existing ontology-based UW Exclusion Detection frameworks, like KAMEL (Lui, Xiang, and Krishnaswamy 2024). Unlike KAMEL, which primarily relies on the SBioBERT encoder to discern associations between UW exclusions and ICD codes based on their textual description, KOAAL employs both the SBioBERT encoder and leverages the implicit domain knowledge inherent in LLMs, like GPT-3.5, through its advanced generative capabilities. To address the challenges that stem from disjoint medical ontologies, KOAAL

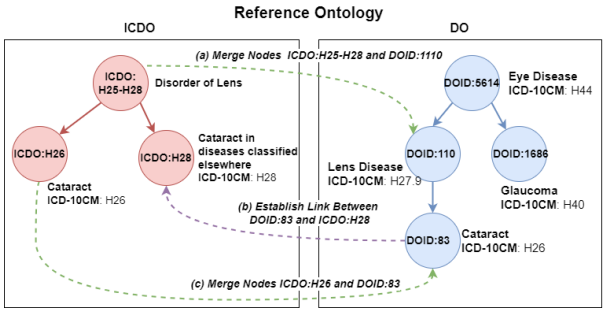

Figure 1: Given the base ontologies ICDO (Lui, Xiang, and Krishnaswamy 2024) and DO (Baron et al. 2023), the objective is to establish medically relevant links within and between them.

leverages the medical domain knowledge inherent in LLMs to effectively combine these separate ontologies into a unified, comprehensive reference ontology. This integration facilitates the establishment of medically relevant associations that were previously unavailable.

To address the limitations of the sentence encoder for medical concept alignment, KOAAL employs an incremental fine-tuning strategy to train the encoder on diverse medical concepts from various medical ontologies. Unlike KAMEL, which trains the sentence encoder primarily on UW exclusions and ICD code descriptions, KOAAL exposes the encoder to a broader range of medical concepts. This enhances the model's generalization capabilities, reducing the risk of overfitting and the challenges associated with medical synonyms and abbreviations. Moreover, it improves the framework's ability to more accurately map medical terms from UW exclusions to their corresponding nodes within the reference ontology, based on their associated medical concepts. Furthermore, our proposed approach of leveraging LLMs to extract relevant medical terms from UW exclusions addresses the challenges that stem from negation terms.

In the subsequent subsections, we outline the KOAAL framework, which consists of three key steps: First, we construct a comprehensive reference ontology by harnessing the implicit domain knowledge inherent in LLMs to align disjoint medical ontologies. Next, we integrate feedback from claim assessors as edge weights within the reference ontology, refining the established linkages. Lastly, we utilize the reference ontology to establish medically relevant associations between UW exclusions and ICD codes for UW Exclusion Detection.

In this paper, we validate our approach using OpenAI's GPT-3.5-Turbo (OpenAI 2023), leveraging its ability to effectively harness relevant medical domain knowledge. Our choice of utilizing this GPT-based model was influenced by the ease of integration provided by the OpenAI API, which facilitates the incorporation of LLMs without extensive infrastructure requirements. However, it is important to emphasize that our proposed methodology is model-agnostic and not restricted to GPT-based models.

Algorithm 1: Aligning Medical Ontologies with LLMs

Input: DO (O_{DO}), ICDO (O_{ICD}), SBioBERT, and cosine similarity threshold t (Equation 2).
Output: Reference ontology (O_{ref}), fine-tuned SBioBERT.

Fine-tune SBioBERT on medical concepts c_{DO} in O_{DO} where $c_{DO} \in n_{DO}$ and $n_{DO} \in O_{DO}$ using the BatchAllTripletLoss.
Initialize $O_{ref} \leftarrow O_{DO} \cup O_{ICD}$.
Iterate over nodes $n_{DO} \in O_{DO}$ and $n_{ICD} \in O_{ICD}$.
for each medical concept $c_{DO} \in n_{DO}$ **do**
  for each medical concept $c_{ICD} \in n_{ICD}$ **do**
    $e_{DO} = SBioBERT(c_{DO})$
    $e_{ICD} = SBioBERT(c_{ICD})$
    if $F_{similar}(e_{DO}, e_{ICD}, t)$ **then**
      $r = GPTRelation(c_{DO}, c_{ICD})$
      if $r =$ *parent-child* **then**
        Add edge between n_{DO} and n_{ICD}.
      else if $r =$ *child-parent* **then**
        Add edge between n_{ICD} and n_{DO}.
      else if $r =$ *equivalence* **then**
        Nodes n_{DO} and n_{ICD} are merged.
      else if $r =$ *no-relation* **then**
        No action taken
      end if
    end if
  end for
end for
Further fine-tune SBioBERT on medical concepts c_{ref} in O_{ref} where $c_{ref} \in n_{ref}$ and $n_{ref} \in O_{ref}$ using the BatchAllTripletLoss.
return O_{ref}, SBioBERT

Step 1: Align Medical Ontologies with LLMs

Medical ontologies like the DO (Baron et al. 2023) and the ICDO (Lui, Xiang, and Krishnaswamy 2024) have proven to be valuable sources of explicit domain knowledge for UW Exclusion Detection (Lui, Xiang, and Krishnaswamy 2024). However, their isolated structures impede the establishment of medically relevant associations between them. While the DO includes generic medical concepts that are well-aligned with the medical terms present in UW exclusions, it encompasses only 2,427 unique ICD codes, representing a small fraction of all available ICD codes. On the other hand, the ICDO contains an exhaustive list of all currently available ICD codes, but its descriptions often lack the layman terms typically used in UW exclusions.

To overcome this limitation, we focus on developing a more comprehensive reference ontology that establishes key relationships both within and across medical ontologies like DO and ICDO. For instance, as illustrated in Figure 1, medical concepts between edge (a) are merged as both nodes refer to lens-related diseases. On the other hand, edge (b) establishes a directed connection from node DOID:83 to ICDO:H28, indicating that 'cataract' is a parent concept to 'cataract in diseases classified elsewhere'. Lastly, medical concepts associated with edge (c) are merged as both nodes ICDO:H26 and DOID:83 refer to the diagnosis 'cataract'. This updated ontology structure establishes a clear connection between the medical concept 'eye disease' and ICD code H28 through the path: *DOID:5614* $\rightarrow$ *DOID:110* $\rightarrow$ *DOID:83* $\rightarrow$ *ICDO:H28*. Without aligning the DO and ICDO, such connections would remain unestablished, highlighting the critical need to align these disjoint medical ontologies to establish complex medical associations.

To achieve this, we propose an approach that leverages the implicit domain knowledge in LLMs to align medical ontologies, as illustrated in Figure 2. First, we utilize the SBioBERT sentence encoder, which adopts the SBERT architecture (Reimers and Gurevych 2019) and is initialized with BioBERT's (Lee et al. 2020) pre-trained weights. This encoder identifies semantically similar medical concepts within DO and the ICDO. SBioBERT is initially fine-tuned on medical concepts in the DO using the BatchAllTripletLoss (Hermans, Beyer, and Leibe 2017), an extension of the triplet loss function. This loss function considers all triplets in the batch and aims to minimize the distance, d, between the embeddings of an anchor, a, and a positive example, p, while maximizing the distance between the embeddings of the anchor, a, and a negative example, n, as defined in Equation 1. For instance, consider the triplet ('Severe asthma attack', 'Asthma with status asthmaticus', 'Cataract'). Both 'Severe asthma attack' (anchor) and 'Asthma with status asthmaticus' (positive) are asthma-related conditions and should have similar embeddings. On the other hand, 'Severe asthma attack' (anchor) and 'Cataract' (negative) are unrelated medical concepts and their embeddings should be dissimilar. Initially, SBioBERT model is fine-tuned on 37,424 medical concepts in the DO, enabling the encoder to establish medically relevant associations within the ontology.

$$L(a, p, n) = \sum \max(d(a, p) - d(a, n) + \textit{margin}, 0) \quad (1)$$

Next, medical concepts from DO and ICDO are processed using the SBioBERT encoder to assess their semantic similarity, as detailed in Equation 2. If the cosine similarity between these embeddings exceed the predefined threshold t, these concepts are considered semantically similar and are subsequently processed by GPT-Relation, a prompt-based model designed to discern the directional relationships between medical concepts. The generated relationships are categorized into one of the four types: (i) parent-child, (ii) child-parent, (iii) equivalence, or (iv) no relation. Based on the type of relationship identified, the framework may establish directed edges, merge nodes, or disregard the connection if no valid relationship exists. This process facilitates the establishment of medically relevant inter- and intra-relationships between nodes within the reference ontology. Subsequently, the SBioBERT is further fine-tuned on the updated reference ontology which consists of 94,989 medical concepts. This approach is formally detailed in Algorithm 1.

$$F_{similar}(M_A, M_B, t) = \frac{M_A \cdot M_B}{\|M_A\| \|M_B\|} > t \quad (2)$$

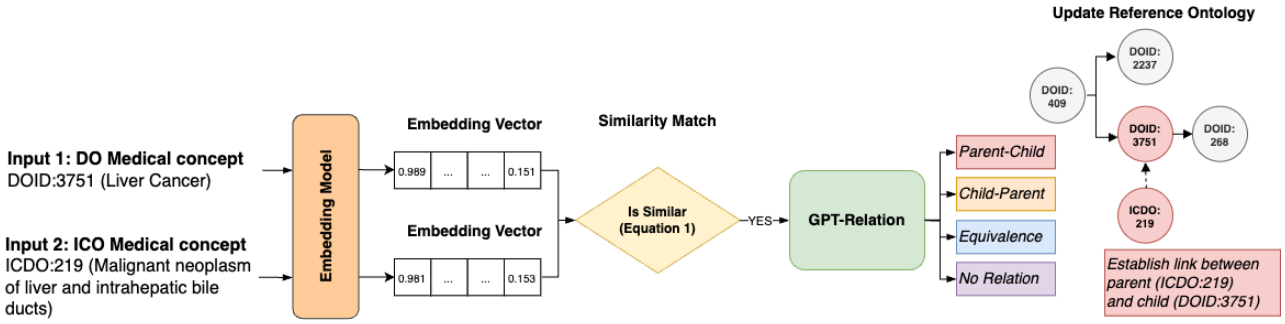

Figure 2: Align Medical Ontologies with LLMs: Nodes from disjoint medical ontologies are aligned based on medically relevant associations. A sentence encoder is used to determine if their descriptions are semantically similar by evaluating their cosine similarity. GPT-Relation is then employed to identify directed associations between these concepts. The relationships established are then used to update the reference ontology.

Step 2: Integrate Feedback via Edge Weights

The approach detailed in the previous subsection establishes inter- and intra-relationships between nodes from DO and ICDO to construct a reference ontology. However, not all connections may be relevant in the context of UW Exclusion Detection. For example, consider the generated path: diabetes→obesity→sleep apnea. Although diabetes is commonly associated with obesity, and obesity is a significant risk factor for sleep apnea, an exclusion for diabetes should not directly lead to the rejection of claims for sleep apnea.

To ensure the relevance of links within the reference ontology, we incorporate feedback from claim assessors, captured through labeled health insurance claims data, as edge weights. This data includes medical terms from UW exclusions, ICD codes, and their corresponding claim statuses (labels). During the fine-tuning process, medical terms and ICD codes are mapped to specific nodes within the ontology, where their paths are generated. Next, the edge weights in the reference ontology are initialized as β and subsequently adjusted based on the frequency of positive and negative labels associated with these paths. To maintain contextual relevance, the threshold α moderates the impact of edges with high centrality, particularly those with more negative than positive associations. The incorporation of these edge weights allows the ontology to be fine-tuned based on feedback from claim assessors. The specifics of this approach are detailed in Algorithm 2.

Algorithm 2: Integrate Feedback via Edge Weights

Input: Reference Ontology (O_{ref}), training data (N_{UW}, N_{ICD}, L) where:
N_{UW}: Ontology nodes associated with the UW exclusion.
N_{ICD}: Ontology nodes associated with the ICD code.
L: Labels '1' for rejected claims and '0' otherwise.
α: Hyper-parameter that prevents edges with high centrality from being penalized.
β: Default value of edge weights.
Output: Updated ontology O_{ref}, augmented with edge weights.

Let E denote the set of edges where $(N, E) \in O_{ref}$
For each edge $e \in E$, initialize the edge weights $e_w = \beta$, $e_{pos} = 0$, $e_{neg} = 0$.
for $(n_{UW}, n_{ICD}, l) \in (N_{DO}, N_{ICD}, L)$ **do**
  $P \leftarrow getSimplePath(n_{UW}, n_{ICD})$
  $len_{path} \leftarrow \|P\|$
  for $i = 1$ to $len_{path} - 1$ **do**
    if *HasPath*$(P[i], P[i+1])$ **then**
      if $l = 1$ **then**
        $e_{pos(P[i],P[i+1])} \leftarrow e_{pos(P[i],P[i+1])} + 1$
      else
        $e_{neg(P[i],P[i+1])} \leftarrow e_{neg(P[i],P[i+1])} + 1$
      end if
    end if
  end for
end for
for $(e_{pos}, e_{neg}, e_w) \in E$ **do**
  $score \leftarrow \frac{e_{pos}}{e_{pos}+e_{neg}}$
  if $score \leq \alpha$ or $score \geq \beta$ **then**
    $e_w \leftarrow score$
  end if
end for
return O_{ref}

Step 3: Establish Medically Relevant Associations

In this subsection, we outline our proposed approach to establish medically relevant associations between UW exclusions and ICD codes for UW Exclusion Detection. As illustrated in Figure 3, KOAAL evaluates policy-specific UW exclusions and ICD codes from incoming claims to determine whether these claims should be rejected based on their associations. KOAAL maps both the UW exclusions and ICD codes to corresponding nodes within the reference ontology where their associations are delineated through their connected paths.

To achieve this, we first utilize GPT-MER, a prompt-based model designed to extract medical terms from free-text UW exclusions with minimal examples. This approach significantly reduces reliance on extensively annotated data, offering an advantage over traditional supervised learning

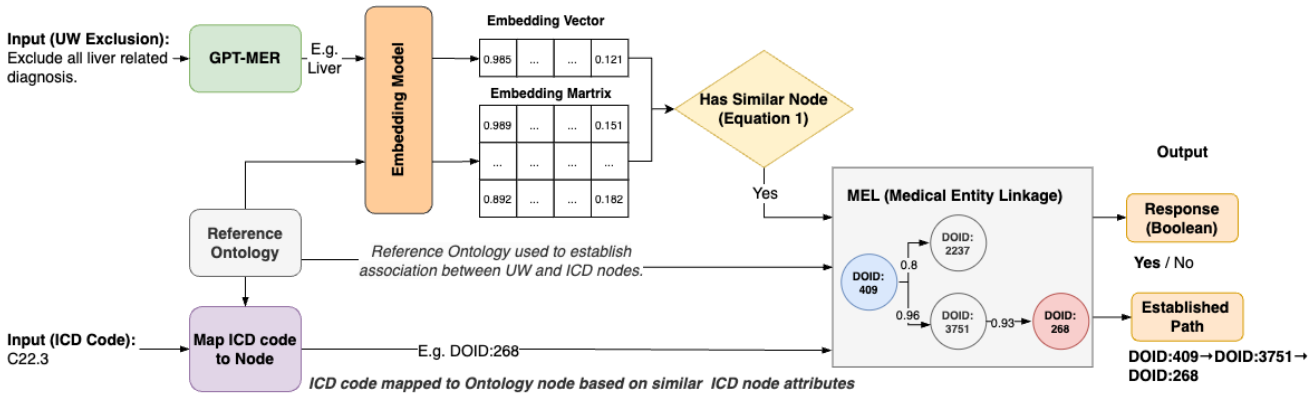

Figure 3: Establish Medically Relevant Associations: KOAAL validates the relationship between an input UW exclusion and ICD code. In this example, the UW exclusion mapped to DOID:409 (liver disease) and the ICD code mapped to DOID:268 (liver angiosarcoma) are associated through DOID:3751 (liver cancer).

methodologies. Next, a post-processing step is implemented to ensure that the medical terms generated by the model exists in the UW exclusion. This mitigates potential scenarios where the model hallucinates and generates non-existent medical terms. Subsequently, the extracted terms are encoded by the fine-tuned SBioBERT model (in step 1) and mapped to their corresponding nodes within the reference ontology. This mapping is achieved by computing the cosine similarity between the embeddings of the extracted terms and all medical concepts in the reference ontology, as described in Equation 2. If the minimum edge weight, e_{min}, of a path between the mapped UW and ICD nodes is greater than or equal to threshold β, the UW exclusion and ICD code is considered to be associated and the claim is flagged for potential rejection. Additionally, the generated paths delineate the associations between UW exclusions and ICD codes, enhancing explainability by providing users with clear, interpretable connections. This approach is formalized in Algorithm 3.

Details of the prompt-based models employed in the KOAAL framework and the specifics of the evaluation datasets, discussed in the subsequent section, can be assessed at https://github.com/luishengjie/koaal.

Algorithm 3: Medical Entity Linkage (MEL)

Input: Reference ontology (O_{ref}), UW node (n_{UW}), ICD node (n_{ICD}), and threshold β.
Output:
$flag_{out}$: "yes" if a valid connection exists between n_{UW} and n_{ICD}, "no" otherwise.
$path_{out}$: connected path between n_{UW} and n_{ICD}.

$flag_{out} \leftarrow no$, $path_{out} \leftarrow null$
$P \leftarrow getAllSimplePaths(n_{UW}, n_{ICD})$
for path $p \in P$ **do**
  $e_{min} \leftarrow \min\{e_w \mid e_w \in p\}$
  if $e_{min} \geq \beta$ **then**
    $flag_{out} \leftarrow yes$
    $path_{out} \leftarrow p$
  end if
end for
return $flag_{out}, path_{out}$

Experimental Validation of the KOAAL Framework

In this section, we evaluate the effectiveness of the KOAAL framework for UW Exclusion Detection using real-world health insurance claims data. First, we conduct an ablation study to validate the effectiveness of the individual components within the KOAAL framework. Next, we compare KOAAL against traditional multi-label classification approaches, two GPT-prompt-based models: GPT-UW and GPT-UW-Desc, and the KAMEL framework (Lui, Xiang, and Krishnaswamy 2024).

In these experiments, we utilize the following hyperparameters within the KOAAL framework: (i) we initialize the cosine similarity threshold $t = 0.8$, as outlined in Equation 2. A higher cosine similarity ensures more precise mapping of medical terms to relevant nodes within the reference ontology. This threshold was selected to optimize the accuracy and relevance of these mappings. (ii) The hyper-parameters for generating the edge weights are set to $\alpha = 0.1$ and $\beta = 0.5$, as specified in Algorithm 2. The low α threshold minimizes the penalization of high-centrality edges with a slightly more negative than positive label ratio. The β value of 0.5 was selected a balanced midpoint for initializing the edge weights.

Evaluation Metrics

In our experiments, the F1 score is used as the primary metric to assess the effectiveness of each approach, supplemented by precision and recall scores. These metrics are crucial for evaluating the system's ability to accurately identify valid claims while minimizing false rejections, which is critical for automating health insurance claims processing. For instance, an approach with high precision but low recall

	Health Insurance I	Health Insurance II	MIMIC-IV-ICD-10-N3
Total Claims	76197	255706	32318
Train Data	36077	197101	22622
Test Data	40039	58312	9696

Table 2: Summary of the datasets used in the experiments. For more detailed descriptions, refer to the KAMEL study (Lui, Xiang, and Krishnaswamy 2024).

	O_{DO}	O_{ICD}	$O_{DO} \cup O_{ICD}$	O_{ref}
Nodes:	10,960	2,194	13,154	12,750
Edges:	15,155	2,172	17,327	17,953
Medical Concepts:	37,424	57,266	94,690	94,989
ICD Codes:	2,427	45,530	45,530	45,530

Table 3: Overview of the ontologies used: (i) DO (O_{DO}), (ii) ICDO (O_{ICD}), (iii) Unaligned Reference Ontology ($O_{DO} \cup O_{ICD}$), (iv) Aligned Reference Ontology (O_{ref}).

may inadvertently approve invalid claims, resulting in significant financial loss. On the other hand, an approach with high recall but low precision could result in many claims being incorrectly flagged for rejection. This increases processing time as each claim would require manual verification by claim assessors (Lui, Xiang, and Krishnaswamy 2024). To provide a comprehensive overview of the system's performance, the overall accuracy is also measured.

Datasets

To facilitate a direct comparison with existing methodologies, we evaluate the KOAAL framework using three real-world datasets previously employed in the KAMEL study (Lui, Xiang, and Krishnaswamy 2024). These datasets include: Health Insurance I, Health Insurance II, and MIMIC-IV-ICD-10-N3. Health Insurance I and II consists of actual health insurance claims data while MIMIC-IV-ICD-10-N3 is a publicly available dataset curated specifically for simulating UW Exclusion Detection. Each dataset includes free-text UW exclusions, along with their corresponding ICD codes, descriptions, and claim statuses (Lui, Xiang, and Krishnaswamy 2024). Table 2 includes a summary of the experimental datasets used.

Ablation Study

In this subsection, we conduct an ablation study to evaluate the impact of individual components within the KOAAL framework. The performance is assessed across four distinct configurations of KOAAL where the details of the ontologies used in this experiment are presented in Table 3:

1. **KOAAL$_{\text{basic}}$**: Operates with disjoint ontologies O_{DO} and O_{ICD}.
2. **KOAAL$_{\text{OA}}$**: Aligns O_{DO} and O_{ICD} with implicit domain knowledge from GPT-relation (detailed in Algorithm 1).
3. **KOAAL**: Implements the full KOAAL framework, which includes aligning O_{DO} and O_{ICD} with implicit domain knowledge from GPT-relation (Algorithm 1) and incorporating user feedback via edge weights (Algorithm 2).
4. **KOAAL$_{\text{SW}}$**: Employs the sliding window approach to extract medical entities, as utilized by KAMEL (Lui, Xiang, and Krishnaswamy 2024), replacing GPT-MER. This approach segments the UW exclusions into key phrases, which are validated against medical concepts within the reference ontology. For example the UW exclusion "exclude eye diseases" would be segmented into key phrases: 'exclude eye diseases', 'exclude eye', 'eye diseases', 'exclude', 'eye', and 'diseases'.

The results of the ablation study are presented in Table 4 where progressive improvements in F1 score is observed as components were incrementally integrated into the KOAAL framework. These enhancements are primarily attributed to improvements in recall, highlighting the effectiveness of each component in establishing medically relevant associations between UW exclusions and ICD codes. The performance improvement from KOAAL$_{\text{basic}}$ to KOAAL$_{\text{OA}}$ demonstrates the efficacy of our approach, which integrates medical ontologies with implicit domain knowledge from LLMs to establish complex, medically relevant associations necessary for UW Exclusion Detection. Further improvements in F1 score from KOAAL$_{\text{OA}}$ to KOAAL underscore the benefits of incorporating feedback from claim assessors, ensuring that the connections between UW exclusions and ICD codes are relevant.

To evaluate the effectiveness of our approach in extracting relevant medical terms from UW exclusions, we compare the performance of KOAAL and KOAAL$_{\text{SW}}$. Based on the results obtained in Table 4, KOAAL consistently outperforms KOAAL$_{\text{SW}}$ in precision across all three datasets. This demonstrates the advantage of utilizing the implicit domain knowledge inherent in LLMs to identify valid medical terms in UW exclusions. Table 5 presents examples of the medical terms extracted by KOAAL and KOAAL$_{\text{SW}}$. These findings are aligned with our experimental results where the medical terms extracted by KOAAL are more complete and contain less noise. Additionally, KOAAL effectively handles negation term, enhancing its robustness. However, as KOAAL post-processes the output from GPT-MER to verify that the extracted medical terms exist in the input text, misspellings in the generated responses prevent the extraction of relevant medical terms. For example, given an UW exclusion for 'gynecological disorders', if GPT-MER returns the medical term as 'gynaecological disorders', the discrepancy in spelling causes the medical term to be ignored by the system for further processing. Such scenarios contribute to the lower recall observed when evaluated on Health Insurance II and the MIMIC-IV-ICD-10-N3 dataset.

DATASET	METRIC	$KOAAL_{basic}$	$KOAAL_{OA}$	$KOAAL_{SW}$	$KOAAL$
HEALTH INSURANCE I	Acc	87.59%	90.32%	88.51%	**91.30%**
	F1	40.92%	60.36%	52.71%	**66.34%**
	Prec	**92.47%**	91.33%	80.60%	90.34%
	Rec	26.27%	45.07%	39.16%	**52.42%**
HEALTH INSURANCE II	Acc	82.23%	88.15%	85.10%	**89.23%**
	F1	64.23%	71.19%	66.35%	**74.80%**
	Prec	82.78%	**91.30%**	76.54%	90.60%
	Rec	52.47%	58.34%	**85.55%**	63.69%
MIMIC-IV-ICD-10-N3	Acc	63.24%	68.16%	**84.60%**	81.10%
	F1	67.95%	73.58%	**89.60%**	86.44%
	Prec	**98.38%**	97.59%	90.91%	93.67%
	Rec	51.90%	59.05%	**88.32%**	80.25%

Table 4: This table compares the performance of different configurations of KOAAL, highlighting the impact of individual components within the framework.

Approach	UW Exclusion	Extracted Terms
$KOAAL$	cysts / growths / lesions of the breasts, gallstones and the biliary system, cysts/ growths/ lesions of the liver.	uw-term:[‘biliary system’, ‘breasts’, ‘gallstones’, ‘liver’], excluded-term:[]
	breasts, fracture of the left foot and its sequelae, disorders of the urinary system (including kidneys).	uw-term:[‘kidneys’, ‘disorders of the urinary system’, ‘fracture of the left foot’, ‘breasts’], excluded-term:[]
	gynecological disorders excluding breasts.	uw-term:[‘gynecological disorders’], excluded-term:[‘breasts’]
$KOAAL_{SW}$	cysts / growths / lesions of the breasts, gallstones and the biliary system, cysts/ growths/ lesions of the liver.	uw-term:[‘and the biliary’, ‘liver’]
	breasts, fracture of the left foot and its sequelae, disorders of the urinary system (including kidneys).	uw-term:[‘fracture of the’, ‘of the urinary’]
	gynecological disorders excluding breasts.	uw-term:[‘gynecological disorders’, ‘breasts’]

Table 5: Examples of medical terms extracted by $KOAAL$ and $KOAAL_{SW}$, highlighting the differences in their accuracy and relevance.

Benchmark Validation

In this subsection, we validate the effectiveness of our proposed KOAAL framework against five benchmark approaches:

1. **KAMEL**: The original KAMEL approach introduced in (Lui, Xiang, and Krishnaswamy 2024).
2. **PLM-ICD**: A multi-label classifier referenced in the KAMEL study (Lui, Xiang, and Krishnaswamy 2024).
3. **CAML**: Another multi-label classifier referenced in the KAMEL study (Lui, Xiang, and Krishnaswamy 2024).
4. **GPT-UW**: A prompt-based model, based on OpenAI's GPT-3.5-Turbo, designed to discern the relationship between UW exclusions and ICD codes.
5. **GPT-UW-Desc**: A prompt-based model, based on OpenAI's GPT-3.5-Turbo designed to discern the relationship between UW exclusions and the textual descriptions of ICD codes.

Table 6 presents the results of our benchmark validation, where KOAAL consistently outperformed all baseline approaches across all evaluation datasets in terms of F1 score. This demonstrates its effectiveness in establishing medically relevant associations between UW exclusions and ICD codes. These improvements, primarily due to increased recall, underscore KOAAL's ability to establish more complex and medically relevant associations than the benchmark approaches. Notably, KOAAL addressed the limitations of KAMEL (Lui, Xiang, and Krishnaswamy 2024), as demonstrated by the improved recall and F1 score across all datasets. This improvement underscores KOAAL's enhanced capability to leverage the domain knowledge encapsulated within the same base ontologies, DO (Baron et al. 2023) and ICDO (Lui, Xiang, and Krishnaswamy 2024), for UW Exclusion Detection. Although KOAAL exhibited lower recall compared to PLM-ICD on the Health Insurance I dataset, its significantly higher precision resulted in an overall improvement in F1 score.

DATASET	METRIC	CAML*	PLM-ICD*	GPT-UW	GPT-UW-Desc	KAMEL*	$KOAAL$
HEALTH INSURANCE I	Acc	53.10%	41.05%	84.06%	89.98%	**97.43%**	91.30%
	F1	23.52%	25.93%	27.14%	59.96%	39.24%	**66.34%**
	Prec	16.04%	16.32%	53.70%	86.59%	**93.61%**	90.34%
	Rec	44.10%	**63.10%**	18.16%	45.85%	24.82%	52.42%
HEALTH INSURANCE II	Acc	54.15%	41.91%	85.12%	84.71%	85.34%	**89.23%**
	F1	36.10%	33.46%	62.26%	64.11%	64.23%	**74.80%**
	Prec	27.75%	23.48%	85.66%	77.93%	82.78%	**90.60%**
	Rec	51.63%	58.24%	48.90%	54.46%	52.47%	**63.69%**
MIMIC-IV-ICD-10-N3	Acc	28.22%	28.81%	41.29%	67.74%	77.50%	**81.10%**
	F1	0.00%	12.25%	36.08%	72.90%	77.75%	**86.44%**
	Prec	0.00%	82.11%	**98.77%**	98.69	77.28%	93.67%
	Rec	0.00%	6.21%	22.07%	57.80%	78.23%	**80.25%**

Table 6: Benchmark validation results. This table compares the performance of KOAAL with benchmarks approaches. Benchmarks from the KAMEL study (Lui, Xiang, and Krishnaswamy 2024) are indicated with an asterisk.

Real-World Application of KOAAL

Key components of the KOAAL framework have been deployed as part of an automated health claims processing system, widely adopted across Singapore and the ASEAN region. These systems process approximately 1 million claims annually, demonstrating the robustness and effectiveness of our framework in high-volume environments. In production environments, KOAAL has achieved a $10\% - 20\%$ improvement in recall, while maintaining precision on par with prior approaches. Furthermore, KOAAL's ability to incorporate user feedback through the adjustment of edge weights streamlines the fine-tuning process. This significantly reduces the turnaround time, effectively bridging the gap between back-testing results and real-world deployment. Moreover, by integrating domain knowledge from various sources into a comprehensive reference ontology, the relationship between UW exclusions and ICD codes can be established based on their associated paths within the ontology. For instance, given an UW exclusion that excludes "fits and seizures" and the ICD code R56.8 (Other and unspecified convulsions), their relationship can be establish through the connected path: seizures→convulsions→Other and unspecified convulsions (R56.8). This enhances explainability by providing justifiable, traceable, and verifiable reasons for each prediction, allowing claim assessors to understand why these claims are flagged for rejection.

Conclusion and Future Work

This work underscores the critical role of domain knowledge in enhancing the decision-making processes for automated health insurance claims processing. We demonstrated that combining the implicit domain knowledge inherent in LLMs, the explicit domain knowledge from medical ontologies, and feedback from claim assessors provides a comprehensive and robust source of domain knowledge for UW Exclusion Detection. Our proposed approach leverages the domain knowledge within the reference ontology to establish complex, medically relevant associations between UW exclusions and ICD codes that were that were previously unattainable with existing methodologies. Furthermore, KOAAL enhances explainability by clearly delineating the associations between UW exclusions and ICD codes through their connected paths within the reference ontology. Extensive experimental evaluations using two real-world health insurance claims datasets validated the efficacy of our approach in establishing medically relevant associations between UW exclusions and ICD codes. This work paves the way for future research focused on harnessing domain knowledge inherent in medical ontologies and LLMs to address real-world challenges in complex domains such as health insurance.

Future work will focus on scaling the KOAAL framework to include more complex ontologies, such as SNOMED CT (SNOMED International 2023). This expansion seeks to enhance the framework's utility by establishing more comprehensive medically relevant associations across medical codes, diagnoses and surgical procedures.

References

Alsentzer, E.; Murphy, J. R.; Boag, W.; Weng, W.-H.; Jin, D.; Naumann, T.; and McDermott, M. B. A. 2019. Publicly Available Clinical BERT Embeddings. *ArXiv*, abs/1904.03323.

Arbabi, A.; Adams, D. R.; Fidler, S.; and Brudno, M. 2019. Identifying Clinical Terms in Medical Text Using Ontology-Guided Machine Learning. *JMIR Medical Informatics*, 7.

Baron, J. A.; Johnson, C. S.-B.; Schor, M. A.; Olley, D.; Nickel, L.; Felix, V.; Munro, J.; Bello, S.; Bearer, C.; Lichenstein, R.; Bisordi, K.; Koka, R.; Greene, C.; and Schriml, L. 2023. The DO-KB Knowledgebase: a 20-year journey developing the disease open science ecosystem. *Nucleic Acids Research*, gkad1051.

CMS. 2023. 2024 ICD-10-CM. https://www.cms.gov/

medicare/coding-billing/icd-10-codes/2024-icd-10-cm. Accessed: 2023-07-10.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Burstein, J.; Doran, C.; and Solorio, T., eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Hermans, A.; Beyer, L.; and Leibe, B. 2017. In Defense of the Triplet Loss for Person Re-Identification. *ArXiv*, abs/1703.07737.

Jones, K.; and Sah, S. 2023. The Implementation of Machine Learning In The Insurance Industry With Big Data Analytics. *International Journal of Data Informatics and Intelligent Computing*, 2: 21–38.

Kulmanov, M.; Smaili, F. Z.; Gao, X.; and Hoehndorf, R. 2020. Semantic similarity and machine learning with ontologies. *Briefings in Bioinformatics*, 22(4): bbaa199.

Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C. H.; and Kang, J. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4): 1234–1240.

Liu, S.; Ma, W.; Moore, R.; Ganesan, V.; and Nelson, S. 2005. RxNorm: prescription for electronic drug information exchange. *IT Professional*, 7(5): 17–23.

Liu, X.; McDuff, D.; Kovacs, G.; Galatzer-Levy, I.; Sunshine, J.; Zhan, J.; Poh, M.-Z.; Liao, S.; Di Achille, P.; and Patel, S. 2023. Large Language Models are Few-Shot Health Learners.

Lui, S. J.; Xiang, C.; and Krishnaswamy, S. 2024. KAMEL: Knowledge Aware Medical Entity Linkage to Automate Health Insurance Claims Processing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21): 22797–22805.

Luo, Y.; Yang, Z.; Meng, F.; Li, Y.; Zhou, J.; and Zhang, Y. 2023. An Empirical Study of Catastrophic Forgetting in Large Language Models During Continual Fine-tuning. *ArXiv*, abs/2308.08747.

Mullenbach, J.; Wiegreffe, S.; Duke, J.; Sun, J.; and Eisenstein, J. 2018. Explainable Prediction of Medical Codes from Clinical Text. In Walker, M.; Ji, H.; and Stent, A., eds., *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 1101–1111. New Orleans, Louisiana: Association for Computational Linguistics.

Ong, J.; Kedia, N.; Harihar, S.; Vupparaboina, S. C.; Singh, S. R.; Venkatesh, R.; Vupparaboina, K.; Bollepalli, S. C.; and Chhablani, J. 2023. Applying large language model artificial intelligence for retina International Classification of Diseases (ICD) coding. *Journal of Medical Artificial Intelligence*, 6(0).

OpenAI. 2023. ChatGPT. https://openai.com/chatgpt. Accessed: 10/12/2023.

Pal, A.; Umapathi, L. K.; and Sankarasubbu, M. 2023. MedHALT: Medical Domain Hallucination Test for Large Language Models. arXiv:2307.15343.

Pascual, D.; Luck, S.; and Wattenhofer, R. 2021. Towards BERT-based Automatic ICD Coding: Limitations and Opportunities. In *Workshop on Biomedical Natural Language Processing*.

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Conference on Empirical Methods in Natural Language Processing*.

Sahoo, S.; Kobow, K.; Zhang, J.; Buchhalter, J.; Dayyani, M.; Upadhyaya, D.; Prantzalos, K.; Bhattacharjee, M.; Blümcke, I.; Wiebe, S.; and Lhatoo, S. 2022. Ontology-based feature engineering in machine learning workflows for heterogeneous epilepsy patient records. *Scientific Reports*, 12: 19430.

SNOMED International. 2023. SNOMED CT. Accessed: [Date you accessed the information].

Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; Rodriguez, A.; Joulin, A.; Grave, E.; and Lample, G. 2023. LLaMA: Open and Efficient Foundation Language Models. *ArXiv*, abs/2302.13971.

Turchin, A.; Masharsky, S.; and Zitnik, M. 2023. Comparison of BERT implementations for natural language processing of narrative medical documents. *Informatics in Medicine Unlocked*, 36: 101139.

Wang, S.-M.; hsuan Chang, Y.; Kuo, L.-C.; Lai, F.; Chen, Y.-N. V.; yun Yu, F.; Chen, C.-W.; wei Li, Z.; and Chung, Y.-F. 2020. Using Deep Learning for Automatic Icd-10 Classification from Free-Text Data.

WHO. 2004. ICD-10 : international statistical classification of diseases and related health problems : tenth revision.

Zhou, H.; Gu, B.; Zou, X.; Li, Y.; Chen, S. S.; Zhou, P.; Liu, J.; Hua, Y.; Mao, C.; Wu, X.; Li, Z.; and Liu, F. 2023. A Survey of Large Language Models in Medicine: Progress, Application, and Challenge. *ArXiv*, abs/2311.05112.