ESP IJAST
ESP International Journal of Advancements in Science & Technology
ISSN: 2583-9233/ Volume 4 Issue 2 April 2026
Paper Id: IJAST-V4I2P111/Page No: 81–100

Original Article

A Hybrid Knowledge-Infused Neural Framework for Non-Taxonomic and Ternary Relations Discovery from Patient-generated Texts

Jael Gudu[1] Joseph Balikuddembe[2], Johnson Mwebaze[3]
[1,2,3] *Department of Computer Networks, Makerere University, Uganda*

Received Date: 20 June 2026 Revised Date: 25 June 2026 Accepted Date: 30 June 2026

Abstract: *Non-taxonomic and ternary relations are necessary for comprehensive semantic understanding. However, they have remain underexplored, as current research has focused on identifying taxonomic relations. This study addresses this through semantic alignment based on non-taxonomic and ternary relational components. We develop and evaluate a knowledge-infused neural framework for cross-domain ontology integration that supports capturing and representing non-taxonomic and ternary relationships beyond general taxonomic relations. A key contribution is the implementation of the delayed fusion strategy that balances contextual neural learning, the interpretable rule-based dictionary knowledge, and incorporating BioBERT as a relations validator to ensure domain factual grounding. The framework was evaluated on 27,183 key phrases. Two strategies prioritized conservative prediction and wider domain concepts coverage for the knowledge graph construction. The framework achieved an accuracy of 98.91% and an F1 score of 77.6%, a 10% improvement. It also validated 384 semantic relations with a validation rate of 92.7%. Of these, 240 were ternary relations that captured multiple contexts of interactions. Non-taxonomic relations were 144, organized into semantic categories. The framework transforms unstructured patient-generated texts into structured interoperable knowledge, while preserving clinical contexts. This advances semantic interoperability by improving clinical decision making and biomedical knowledge reuse.*

Keywords: *Knowledge-Infused Neural Framework, Ontology Integration, Non-Taxonomic, Ternary, Semantic Interoperability*

I. INTRODUCTION

The need for scalable, relation-aware ontology integration that can manage relationships that go beyond IS-A, and equivalent across several ontologies[1], [2] is the motivation to this study. The limitation of current approaches in capturing other rich semantic relationships often leads to uneven and disjointed knowledge integration[1]. Our goal is to improve the quality of an ontology networks by in terms of the level of expressiveness, and scalability. Ontologies need strategies in place to guarantee that all relationships are transparent, maintain interoperability, and scale as the knowledge grows [3]. This creates a need for a change in strategy towards ontology integration methods that can automatically detect relations while also being aware of how they connect and add them to the existing ontologies regardless of the schema of the data [4].

Due to the inherent nature of patient-generated texts, they need significant pre-processing and feature extraction efforts. Even when processed, existing and state-of-the-art approaches predominantly focus on capturing taxonomical broader/narrower and equivalence relations. This leads to a failure in capturing other rich semantic relationships including non-taxonomic, ternary, and associative relations from unstructured texts [5], [6]. This greatly hinders semantic interoperability and automated reasoning on data that is available. We infer that these non-taxonomic and ternary relations are key in achieving a complete and comprehensive semantic understanding of domain knowledge.

This study is cognizant to the fact that, these relations must be genuine ontological relations, holding and true between entities in reality, regardless of any methods used to discover them within these texts [7], [8]. Furthermore, the existing methods are designed for and work well with clean, structured data. They struggle with noisy, sequential and incomplete data, which is a common characteristic of patient-generated texts. Consequently, this affects the reliability of the extracted knowledge structure. This therefore places a demand for integration framework that is interpretable and extends beyond the broader/narrower and equivalence relations to other comprehensive semantic relationships across multiple ontologies simultaneously [6], [9]. Based on this context, we propose an ontology alignment framework for mental health cross-domain or multi-domain ontology. The framework functions by carrying out an exhaustive examination and classification of non-taxonomic and ternary relationships between concepts.

A. Research Contribution

The primary artefact in this study is the hybrid knowledge-infused neural semantic integration framework. The artefact is a functional system that is meant to extract, classify and structure medical domain knowledge from patient-generated texts. The key characteristics including architecture and performance of the framework are summarized in this section.

- The framework adopts a dual-stream feature-fusion [10] architecture with a delay fusion strategy based on four key layers namely: data acquisition and ontology pre-processing layer, semantic processing and reasoning layer, ontology merging and link analysis layer, and the evaluation and quality assurance layer. The architectural choice facilitates interpretability of the framework's learning behaviour
- The function of the framework is to extract relevant clinical concepts for each stream, classify the relations between them into ternary and non-taxonomic, within the relevant local domain network. The functions are executed by different components including the semantic interpreter and the semantic reasoner. The relations empowered the creation of a semantically rich local knowledge graph that provides a multidimensional view of information, through the ontology generator component. It then aligns the local graph to global, established and authoritative biomedical standards on BioPortal, specifically SNOMED and MeSH.
- The framework demonstrated high effectiveness, achieving a test accuracy of 98.91%, with a high recall of 92.8% and an F1 score of 77.6%. This highlights the framework's reliability and robustness in identifying semantically relevant medical concepts existing in real-world noisy patient-generate texts.

The framework as it is serves a practical tool for data engineering and semantic data design that can be used to transform real-world unstructured patient-generated text into structured and interoperable semantic knowledge. This addresses the core problem of semantic interoperability of mental health (anxiety and depression) data. Additionally, the framework is best suited for high recall information retrieval and screening of clinical concepts in the real-world context.

II. RELATED WORKS

Existing biomedical ontologies including SNOMED CT, LOINC, ICD scale very well concerning discovery and integrating taxonomic relation and structured texts. However, they do not do the same with non-taxonomic relations and patient-generated texts.

A. Characteristics of Patient-Generated Texts

Patient-generated texts, often found in digital platforms, have the following inherent characteristics that make them challenging to analyse without proper advanced techniques..

a) Uncontrolled Vocabulary

Uncontrolled vocabulary refers to informal, everyday expressions that are clinically meaningful, but are not fit for use with standard biomedical standards such SNOMED CT and MeSH [11]. This creates a lexical-semantic gap between patient informal vocabulary and standardized structured clinical vocabularies found in these standards. The standards employ professional vocabulary that are clinically sound [12]. For example, "could barely sleep" is a somatic patient expression for loss of sleep, while "burning up" is an expression for "fever" [12]. These patient vocabularies are difficult to capture using traditional methods leading to discrepancies and inconsistency in terminology usage, thus affecting the discoverability of domain knowledge [13], [14]. Approaches relying on exact lexical matching during discovery can fail to bridge this gap. This consequently leads to low recall and missed relevant concept extraction. Therefore, a framework designed to address this gap and integrate patient vocabularies to standard, structured domain ontologies cannot rely on exact or lexical equivalent only.

b) Semantic-dependency and Cross-reference Complexity

In patient-generated text, semantic relationships are defined based on context. Context refers to the information surrounding a concept that give details of a particular medical event [15]. Within such texts, context is not confined to a single sentence, in some instances it spreads across multiple sentences [16]. For example, in these sentences

"An individual has suffered from anxiety for over a year and a half. It has mainly affected their sleep".

The concept 'anxiety' is very clear within the first sentence, but in the subsequent sentence, the 'it' needs more surrounding concepts for comprehensive understanding of these medical events [17]. This characteristic of patient-generated texts is one of the primary causes of mapping inconsistencies, and a barrier to seamless semantic interoperability. Without proper tools and mechanisms that reason and learn context, cross-dependency and co-reference in patient-generated texts fail. This leads to missed concepts, misclassification of concepts, and mapping inaccuracies.

These requirements greatly informed the methodological, architectural approach, specifically, adopting a hybrid framework of BiLSTM and BioBERT frameworks in this work. The BiLSTM with attention mechanism is adopted to capture the sequential dependencies of the narratives and identify local potential features [18]. The BiLSTM thus leverages on its sequential processing capabilities and resource efficiency to identify candidate semantic relationships across the different sentences in the text with high semantic sensitivity to relevant domain concepts [19]. On the other hand, BioBERT is adopted as a semantic relations and dependency validator, leveraging its existing pre-trained biomedical domain knowledge to perform the following: Verify the cross-dependencies identified or missed by the BiLSTM, and validate the semantic soundness of the non-taxonomic and ternary relations. This is to ensure that the identified relations remais faithful to the available biomedical domain knowledge and contexts. Consequently, this approach improves the quality of the relations. This hybrid framework approach allows the proposed framework to address the cross-sentence dependencies and co-reference challenges in patient-generated texts, while balancing the computational demands and contextual accuracy.

B. Types of Semantic Relationships

Within patient-generated texts, semantic relations can be expressed explicitly or in some instances be implicit in meaning and in words. These relations are often difficult to realize as they are hidden in between different sentences and are spread across a paragraph [20], [21]. This study adopts a two-dimensional categorization approach to semantic relationships focusing on the *semantic types* and *Arity*:

a) Semantic type

This dimension defines the content of the relations. It encompasses *Taxonomic (hierarchical and equivalent)* and *Non-taxonomic (Associative) relations* [22]. Taxonomic relations are represented by the hierarchical arrangement of concepts based on the broader/narrower relations. They often describe IS-A relationships between concepts [23]. On the contrary, non-taxonomic relations represent a wide variety of associative connections beyond IS-A relation that are meant to enrich the ontology's level of expressiveness [9], [24]. They include: temporal, causal, conditional, nested/overlapping relations [25].

b) Arity or Degree of Relationship

This dimension defines the instances of concepts participating in the relationship within the text [26]. It encompasses *Binary and N-ary* relations. *Binary Relation* where only two concepts participate in the relation are the primary focus of most relation extraction works. *N-ary Relations* describe ternary and higher-ary relations, where n is greater than two [27]. For Example in this sentence,

"A patient has suffered anxiety for a long time now, it has affected their sleep. They previously tried cognitive behavioural therapy to help but it failed after a few months. Instead, the situation became worse. The person then consulted a doctor who then prescribed mirtazapine. The medication worked effectively and the person was able to sleep. However, the effect of the medication after a few days has been dreadful causing loss of energy. The person is aware that these are some of the symptoms"[17].

In the first two sentences, three concepts "anxiety, affected sleep, and cognitive behavioural therapy" are participating in a ternary relationship. However, as the sentences increase it moves to an N-ary relationship with 6 concepts involved [17]. These complexities underline the necessity for advanced frameworks that are capable of handling n-ary relations. These relations would also facilitate the creation of more comprehensive biomedical knowledge graphs, thereby enhancing the level of expressiveness (scope) and effectiveness of integration [27].

C. A review of Existing Concept and Relation Extraction Techniques for Natural Language Processing

Concept and relationship extraction for purposes of structuring information found in patient-generated texts is useful in many tasks including knowledge graph population and completion [28], [29], ontology alignment [30], [31] and information retrieval [32]. However, constructing high-quality comprehensive knowledge graphs remains complex due to the gaps identified in section 2.2.

In various domains such as healthcare, security, software, and education, concept and relations extraction from unstructured texts has largely been driven by deep learning (CNNs and RNNs) and pre-trained language frameworks (BERTs and LLMs) for contextual embedding. They have become the core technology for Named Entity Recognition and knowledge graph construction [27], [28].

a) Rule-based Approaches

Rule-based approaches rely on manually defined extraction rules to match the text and extract the relations manually coded and explicitly defined linguistic rules. This makes this approach highly flexible, interpretable and transparent in identifying domain related concepts [33], [34], [35]. It is effective for specific domains where rules are well-defined because

they can efficiently handle specific linguistic nuances by directly tailoring rules to unique exceptions [34]. However, they still faces certain challenges and shortcomings including: First, building the rules is not only time-consuming and difficult to cover all rules, but also has poor scalability and portability. Second, it depends on predefined patterns and can only identify explicit aspect that match the patterns that have been specified. This leads to inadequate results when faced with complex document structures, and implicit dependencies typically encountered in patient-generated texts [36]. Third, when used on their own often have low domain concept coverage and lacks capacity to preserve context in a narrative. There is therefore a need to identify an approach that can lead to accuracy and a wider coverage in retrieving relevant domain related information. Recent attempts to address these shortcomings have integrated machine learning and pre-trained frameworks techniques [37], [38].

b) Dictionary-based Approaches

Dictionary-based methods use language resources such as dictionaries, thesauruses to mine the similarity between entities to improve the quality of matching [39]. Through these techniques, concepts within a narrative can be identified and enriched. Additionally, lexical and semantic relationships between entities can be inferred based on the entity labels [40]. Despite their strengths, they still have limitations in vocabulary coverage due to the complicated nature of medical terminologies [41], [42]. As a result, they remain inadequate in handling non-taxonomic relations and contexts that demand more than lexical level extraction. To improve the performance of the matcher a hybridization approach is recommended by [43].

c) Machine Learning Approaches

The development of machine learning approaches has greatly improved context-aware concept and relation extraction [44]. These techniques employ the usage of algorithms as well as artificial neural networks to learn and make decisions from texts to detect patterns and circumstances that influence relations [45], [46]. This is attributed to their ability to process complex linguistic patterns and contextual relationships including capturing long-range dependencies even with limited annotated data. This makes them most ideal for running semantic analyses that require complex reasoning across unstructured texts [47].

D. Deep Learning Approaches

There are various deep learning techniques including Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Long-Short Term Memory (LSTMs), and Gated Recurrent Unit (GRU) [48], [49]. These techniques often require more training to match the precision and ensure domain-specific adaptability of rule-based systems [49]. Additionally, they still require substantial computational resources than rule-based and dictionary based approaches [34]. Frameworks such as CNN and Long Short-Term Memory (LSTM) shows significant performance over rule-based and dictionary-based approaches in incorporating contextual information [49]. The advantage is that they can work independently or in tandem with explicit semantic domain knowledge to ensure tasks are executed. They also have the ability to reason over unspecified relationships, by generalizing a sentence(s) without requiring any patient or situation-specific training data, thus offering a more scalable approach [48], [50].

While CNNs and RNNs are effective at capturing local features within a sentence, they struggle with long-distance dependencies in multiple sentences or across sentences. Despite handling longer sequences better than RNN, LSTMs and BiLSTM face issues with very long-range dependencies. This contributes to the persistent semantic interoperability gap in unstructured texts that have long-range, cross-sentences dependencies leading to missed concepts and non-taxonomic relations. For this task, BiLSTM leverages on its sequential processing and labeling richness is well suited for the linear format with clear, concise, logical flow of information found in patient expressions [50]. Additionally, the BiLSTM leverages on its ability to capture local-feature and mid-range dependencies based on its low computing resource consumption architecture. A comparative study results by [51] reveals that in one training, the train epoch of BiLSTM is 178.41 seconds, while a transformer training epoch is 1657.75 seconds. This makes a BiLSTM suitable for devices and scenes that require low or less computing resource consumption. This creates the need for a more efficient fusion scheme(s) that balances the performance and compute cost.

Therefore, the design choice of a hybrid knowledge-infused neural framework is a deliberate design consideration, not to outperform any of the efficient frameworks, but to achieve a balance of the semantic interoperability demands, practicability, and interpretability of the study target output. This approach ensures that every output can be explained based on patterns in the dictionary, a lexical rule that can be validated, or a pattern that has been learnt by the neural network. This allows the domain knowledge to guide and constrain the neural concept and relation extraction process.

E. Transformer based Approaches

The emergence of transformer based approaches including Generative Pre-trained Transformers such as Bidirectional Encoder Representation from Transformers (BERT) and Large Language Models (LLMs) has transformed the concept and relation extraction landscape, with their exceptional state-of-the-art performance [52]. They have introduced new possibilities for generating and evaluating relationships found in different types of data, including unstructured text. Biomedical adaptations of BERT like BioBERT [53] and ClinicalBERT [54] has enabled more accurate interpretation of clinical narratives thus advancing clinical decision support, and disease prediction. These frameworks leverage on the innate capabilities of their architecture consisting of an encoder and a decoder to understand contextual relationships within text [55].

Transformer-based frameworks have a tendency to generate content that may be irrelevant, made-up, or inconsistent with the input data, a problem known as hallucination. A hallucination may lead to a wrong concept or relationship prediction. It may also lead to an output that is not compliant with the domain requirements [56]. These consequently have the potential to spread incorrect or misleading information [57]. The problem of reliability of information is particularly important in medicine, where such errors may have serious consequences in the outcome of a patient. Hallucination mitigation actions taken in this study include: first, adopting a knowledge-infused integration approach by using an expanded dictionary, which ensures factual grounding. Second, BioBERT was adopted both as a context analyzer and as a verifier-in-the-loop measure ensure no fabricated concepts and relations were reported.

III. METHODOLOGY

This study adopts Design Science Research (DSR) approach guided by the aim of improving semantic interoperability by developing a context and relation-aware ontology integration framework.

A. Dataset

The study derives its corpus from online medical forums including: patient.info health forum [58], Health Unlocked [59], and Mental Health Forum [60], Depression Dataset on Kaggle [61]. It was a combination of anxiety disorder (AnxD) and depression. Simple random sampling was used to avoid selection bias on the texts. The dataset was a collection of 38,115 sentences with 27,183 key phrases. From the key phrases 21,746 were used as the training set, and 5,437 as the test set. The division of the dataset is shown in Table 1.

Table 1: Size of datasets

Dataset	Total Documents	Relevant Set	Training Set	Testing Set
AnxD and Depression	38,115	27,183	21,746	5,437

B. Ontology Dictionary Construction

The dictionaries constructed include symptom, disease, treatment dictionaries. The goal is to construct a controlled defining vocabulary of key terminologies (domain-level semantics) and the variations of the same that exist. External medical concepts resources such as UMLS, SNOMED CT, ICD 10, and BioPortal guided the construction of the dictionary. This is majorly because these external resources provide large number of precise medical concepts essential domain knowledge expressed explicitly and can be used in supervised ontology learning to define seed elements [62]. The dictionary evaluates the words based on the natural language meaning of those words (for example, by comparing whether two English-language words are synonyms), and linking the words based on their meaning, semantic and grammatical relations [63].

C. Reverse Index Mapping

The expanded local dictionaries were then subjected to a flattened reverse index mapping procedure. This was meant to support O(1) phrase lookup to identify the documents containing the concepts. This was to ensure cleaning of the noisy patient vocabularies and ensuring consistency in the concept representation across the dataset. It was also important because it ensured that synonyms were not separated into different extracted relations.

D. Concept Classification

The categorization criteria used:

- Disease: If the term is relevant to identifying a specific medical problem a patient has been diagnosed with,
- Treatment and Drugs: If a term is relevant to curing diseases, treating any medical conditions, relieving any symptoms of diseases, or preventing any diseases [64],
- Symptom: If a term is relevant to departures from normal functioning or feelings of individuals, which may express the phenomenon affected by diseases [65].

The three classes of interest correspond to MeSH [64], SNOMED CT [66], and UMLS categories. These categories provide a consistent way of finding concepts within the domain [64].

IV. FRAMEWORK ARCHITECTURE

The framework presents four coordinated layers of service, as shown in Figure 1. First layer is the data acquisition and ontology pre-processing layer. The second layer is the semantic processing and reasoning layer. The third layer is the ontology merging and link analysis layer, and the fourth layer is the evaluation and quality assurance layer. The outcome was an integrated global ontology network supporting knowledge sharing and decision-making.

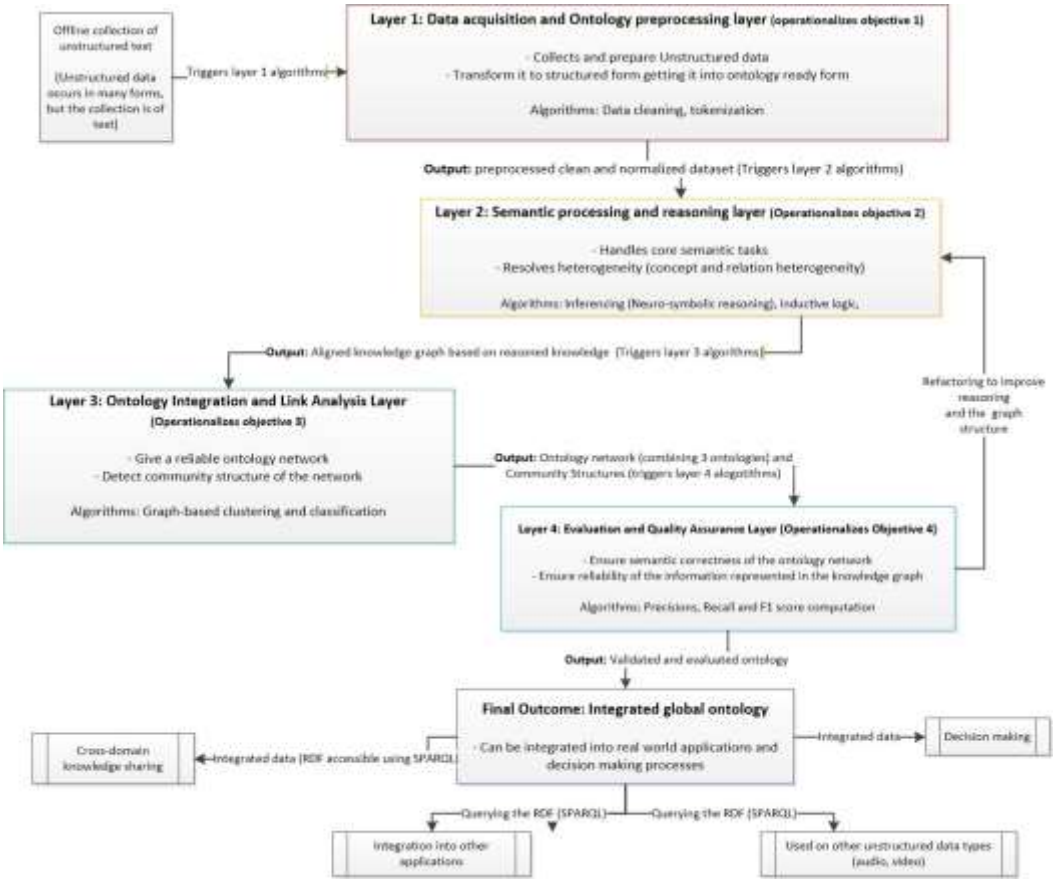


Figure 1: Layered Architecture of Proposed Framework Adopted With Modifications From [67]

A. Framework Layers Design

a) Data Acquisition and Ontology Pre-processing Layer

This layer collects dataset and cleans up the imperfections and inherent irregularities common in the patient-generated texts. Offline versions of the dataset were then stored in Comma Separated Values (CSV) files as a simple and lightweight storage option of the data [68], [69]. The core components are the data receiver, and the data pre-processor as shown in Figure 2.

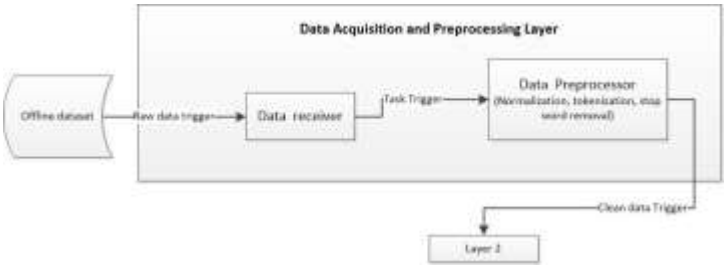
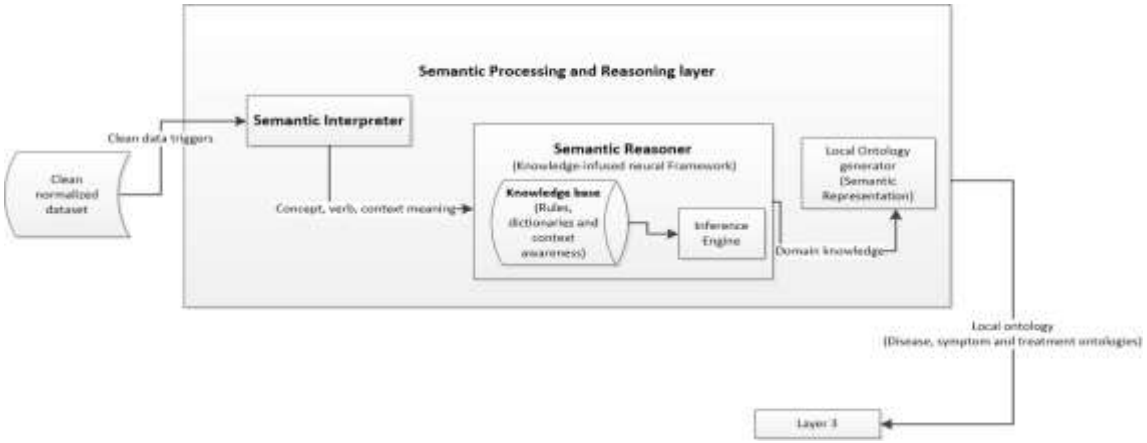


Figure 2: Data Acquisition and Pre-Processing Components

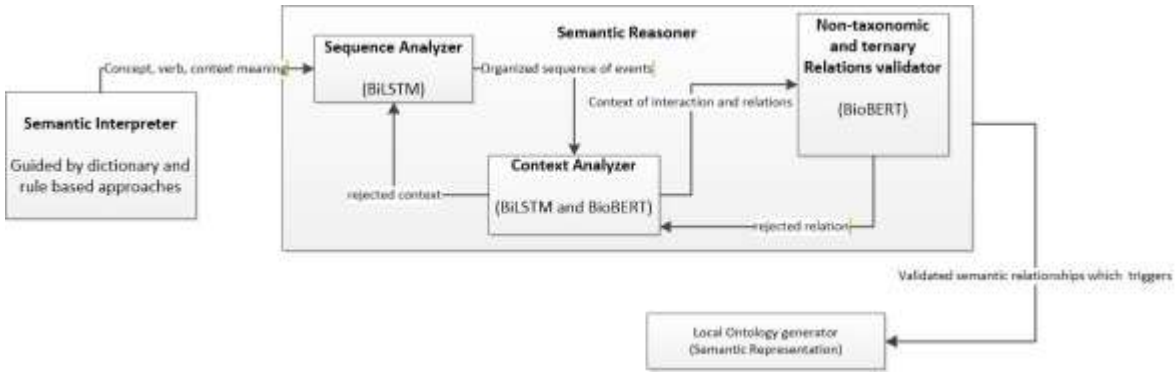
The data receiver's function is to collect the raw data and prepares it for pre-processing. The data pre-processor component identified key phrases in the corpus. It also converts the text into forms that can be used to draw actionable and valuable insights.

b) Semantic Processing and Reasoning Layer

The semantic processing and reasoning layer receives the clean standardized dataset from the data acquisition and pre-processing layer as input. This is the core analytical engine of the framework and it is the contribution of this study. It ensures unstructured patient narratives are converted into structured knowledge that can be integrated with existing medical standards. The primary function of the semantic processing and reasoning layer is to ensure local concepts for diseases, symptom and treatment are extracted based on inferred domain knowledge. The core components are the semantic interpreter, the semantic reasoner, and the ontology generator as shown in Figure 3.



(a) Semantic Processing and Reasoning Components



(b) Core components of the semantic reasoner

Figure 3: Semantic Processing and Reasoning layer Components

The semantic interpreter handles defining the semantics in a way that the information within the dataset becomes easy to reason out and understand. It is achieved using a hybrid of dictionary expanded with WordNet and rule-based algorithms, to provide potential lexical and semantic interpretations of words in the dataset. This is done to ensure interpretability and factual grounding. The interpreter produces a set of explainable structured data that can inform further logical reasoning over the data.

The semantic reasoner implements the core knowledge-infusion neural framework. Its main function is to reason out contextual knowledge based on the predefined rules and conditions [70]. It reasons out the logical sequences of sentences based on different contexts to identify non-taxonomic and ternary relations from the text guided by what exists in external domain ontologies [70], [71]. It contains the text sequence analyser (BiLSTM), context analyser (for both intra and cross sentence dependencies) (BiLSTM), and the relationship validator (BioBERT).

The ontology generator component offers the capability to generate a highly descriptive structured representation for the domain knowledge that supports inconsistencies checking, information retrieval and knowledge network creation. It helps visualize the knowledge by representing the different concepts along with the relations that connect them as a visual network.

c) Ontology Integration and Link Analysis Layer

The ontology integration and link analysis layer combines the three ontologies (disease, symptom, and treatment) into one linked local knowledge network. The core components are the *integration engine* and the *link analyser*. The integration engine's function is to ensure that the output network is complete, concise and consistent [72]. The link analyser's function is to traverses the graph and accesses the links of each node gathering structural information [73].

d) Evaluation and Quality Assurance Layer

The evaluation and quality assurance exists to check the syntactic and semantic quality of the unified network and community structures with an ultimate goal of ensuring syntactic (graph structure) and semantic integrity of the network [74], [75]. This is assessed and constantly improved through evaluation metrics such as precision, recall and F1-score.

e) Final Product

The outcome of the evaluation layer is a unified local ontology comprising of heterogeneous disease, symptom and treatment data merged into a holistic, cohesive semantic network. The final product is then integrated to existing medical domain ontologies, specifically BioPortal to confirm external applicability and semantic interoperability with different domain systems and ontologies.

V. EXPERIMENT RESULTS

Table 2: Performance Evaluation Results for the Local Concept Dictionary

Dataset	Method	Precision (%)	Recall (%)	F1-Score (%)
AnxD and Depression Dataset	Dictionary_NoWordNet	65%	23%	34%

The results show a lower recall of 23%, a precision of 65%, and an F1-score of 34% in comparison to other methods used in this study.

A. Expanding Local Dictionaries with WordNet

The local concept dictionaries were then expanded and enriched using 5 WordNet synonyms for each word variant to prevent the explosion of the synonyms. The results reveal a significant improvement with precision getting to 89%, the recall to 93%, and F1 score to 89% as shown in Table 3. The high recall (93%) meant that the expansion gave more lexical coverage of concepts allowing the framework to identify more concept variants and synonyms often found in patient-generated texts.

Table 3: Performance Evaluation of the Expanded Dictionary with WordNet

		Performance		
Dataset	**Method**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
AnxD and Depression Dataset	Dictionary + WordNet	89%	93%	89%

The recall (93%) outperforms the UMLS recall of 87% reported by [76], but was slightly lower than 97.9% reported by [77] that used specialized concept dictionary integrated in a neural network..

B. Concept Extraction

To evaluate the extent of domain concept coverage two experiments using two different strategies were adopted. The hybrid per class adaptive strategy extracted 113 unique concepts, with an average of 7 domain concepts per document. This gave priority to a more conservative prediction, making it ideal for use as the final performance evaluation measure of the framework. The hybrid union extracted 222 unique concepts, with an average of 14 domain concepts per document (in cases where the document was concept dense). This gave priority to wider coverage of domain concepts, making it ideal for the construction of a rich semantic knowledge graph. Table 4 gives a breakdown of the concepts extracted per strategy.

Table 4: Concepts extracted per strategy

Class type	Hybrid per class adaptive strategy	Hybrid Union
Symptom	54	54
Disease	19	19
Treatment	40	40
Unmatched	0	60
Total	113	222

C. Concept Classification

A concept was only classified in any of the main classes if it was found and identified as a true positive match to what existed in the dictionary. Figure 4 shows the top three symptom mentions by count with loss of motivation being the most frequently mentioned (10520), while the anxiety was the most mentioned disease (4201). The most mentioned treatment was mood stabilizers (18612).

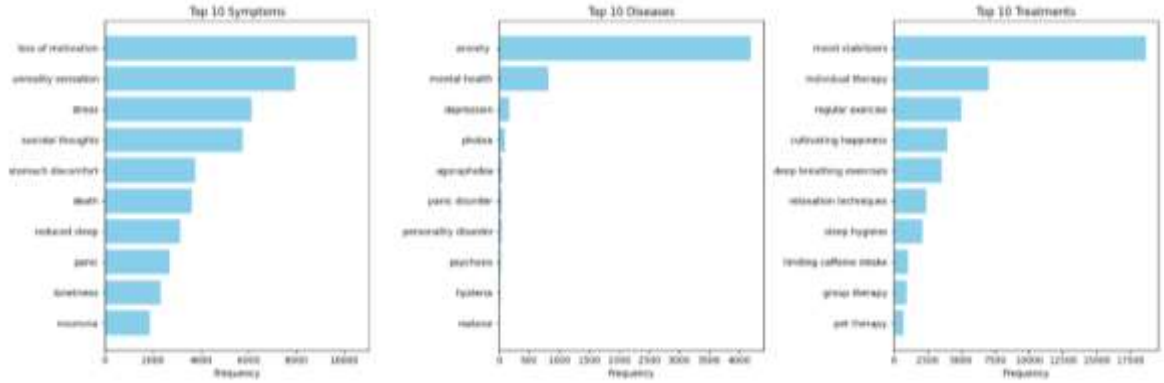
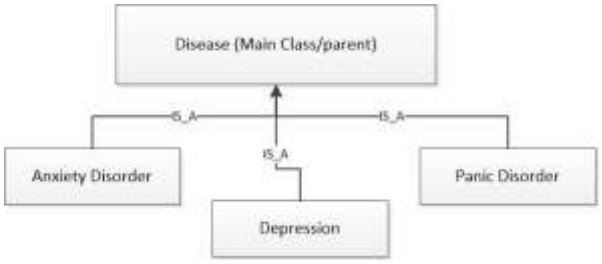
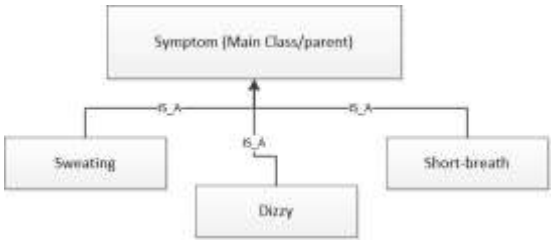


Figure 4: Top 10 Concept Interpretation per class

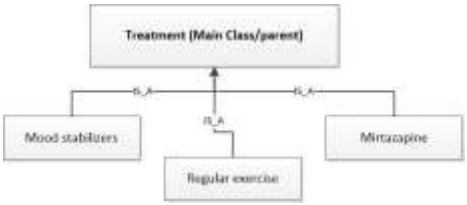
Top-level classification used inductive logic to infer class membership of concepts based on the IS_A relationships. This was necessary to show that the concept belongs to a main class within the medical domain as shown in Figure 5.



(a) Disease classification



(b) Symptom classification



(c) Treatment classification

Figure 5: Concept Classification into Domain Categories

D. Bidirectional LSTM (BiLSTM) with Attention Mechanism

The addition of BiLSTM further enhanced the framework's performance. This improvement stems from its ability to utilize bidirectional contextual information to have a complete understanding of medium range dependencies [78], [79]. It effectively captured nuanced semantic cues that are vital in classification task, while incurring less computational costs during pre-processing and training of the framework [80].

a) Framework Architecture

The BiLSTM's architecture was deliberately designed with relatively low complexity with 2 input layers receiving total parameters 7,019,457 (26.78 MB). The 26.78MB comprised of 4,019,101(15.33MB) trainable parameters and 3,000,356 (11.45MB) non-trainable parameters. The practicability of this is that the total parameter size of 26.78MBs represents a compact neural network architecture. The framework implements a dual-stream feature-fusion architecture with a delayed fusion strategy [10]. This balances the need for contextual neural learning and dictionary and rule-based knowledge (knowledge-infusion method) integration. As seen in Figure 6, the framework receives two inputs *seq_input* (sequential text input) and *dict_feat* that contains engineered features obtained from medical dictionaries, WordNet which are processed

independently. The *seq_input* undergoes linguistic processing of the text sequence, while *dict_feat* is prepared for attention weighting.

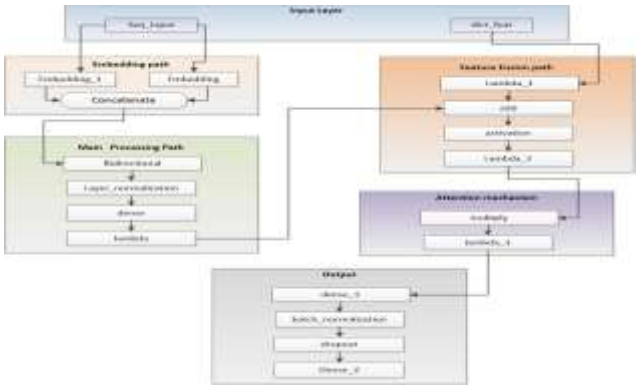


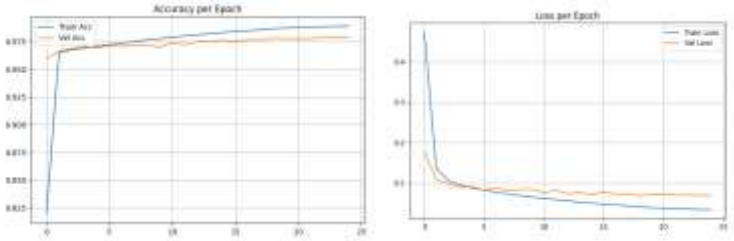
Figure 6: Model's Architectural Design

The study adopts a hybrid weight initialization strategy to foster framework stability, enhance the rate of convergence, consequently influencing the accuracy and generalizability of the framework [81].

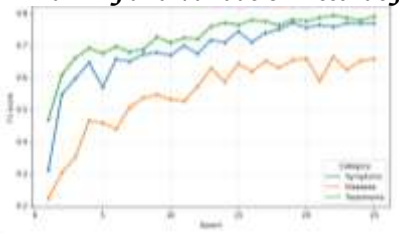
b) Framework Configuration and Training

The framework was trained for 25 epochs each providing the framework with a learning opportunity from the data, incrementally improving its understanding of the underlying patterns in the training set. The results show, that the training was successful in predicting the target labels and implicit instances. To mitigate overfitting during training, the performance was monitored on a separate validation set and an adaptive learning rate. Hyperparameters were tuned based on this validation performance. This resulted to the F1 Scores per class improving steadily with the best of the F1 realized between epoch 19-25 (Symptom 0.768-0.768, Disease 0.653-0.657, and treatment 0.779-0.789). In demonstrating the learning dynamics, the framework showed a relatively low accuracy of 0.6448 at the beginning due to the random weights. However, the framework showed a quick improvement on the accuracy at epoch 2 (0.9642).

The adaptive learning rate scheduler (ReduceLROnPlateau) activated at epochs 18, 23, 25 allowed for necessary adjustments of the random weights and supported consistent gradual improvement in the framework's performance and accuracy up to 0.9788 at epoch 25. The validation metrics tracked the training performance closely, indicating strong generalization capabilities; meaning the framework was able to correctly classify new data found in the dataset and did not just memorize the training set. Similarly, for the training and validation loss reinforces this trend. As the number of iterations increased, the loss values decreased, reaching its lowest point around epoch 17, and gradually stabilized. This trend indicated ongoing optimization of the framework with increasing iterations, leading to progressively stable performance. The small gap between training and validation curves confirms lower overfitting. Figure 7 (a) illustrates the Training and Validation Accuracy (left) and Training and Validation Loss (right) of the framework over the epochs.



(a) Overall Training and Validation Accuracy over Epochs



(a) Per category F1-score over the epochs

Figure 7: Framework Performance over the Epochs

F1 score Per Class Performance

The progression of the F1 score across the 25 training epochs reveals the continuous refinement efforts as shown in Figure 8. The markers in the figure show the values for each epoch with lines connecting successive epochs during the training.

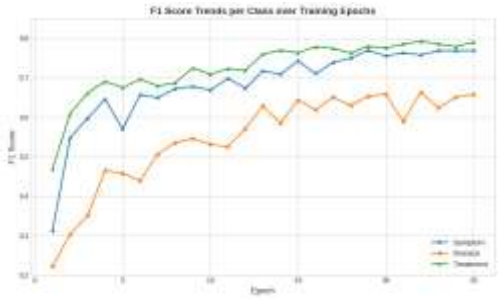


Figure 8: F1 Score Performance Over The 25 Epochs

The treatment concepts achieved the highest F1 score (78.9%) followed by symptoms (76.8%) and disease (65.7%).

c) Relation Extraction

The initial extractor, powered by the BiLSTM identified 414 non-taxonomic and ternary relations. Table 5 shows the extraction results of the different relations extracted.

Table 5: Relations Extracted from Anxiety and Depression Dataset		
Type of relationship	**Relationship Count**	**Sample relations**
Ternary	55	*'treatedWith', 'managesConditionContext', 'relievedInContext',*
Non-Taxonomic	359	*interacts_with, similar_to, co_symptom_of, participates_in*
Total	414	

This captured higher levels of relations, especially the ternary, that provided in-depth understanding to different patient experiences, medical conditions and treatments.

d) Overall Performance Evaluation

Table 6 presents the precision, recall, F1 score performance along with the Hamming loss and Jaccard similarity measure for multi-label classification.

Table 6: Test Dataset Results for the BiLSTM with Attention Mechanism						
		Performance				
Dataset	**Method**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Hamming loss**	**Jaccard**
AnxD and Depression Dataset	BiLSTM with attention mechanism	61.33	75.68	67.75	0.038	0.60

BiLSTM achieved an F1 score of 67.75%, which indicated a reasonable performance in generalization even with new data, previously unknown to it. The results reveal a trade-off, with the priority being on semantic sensitivity (identifying correct domain concepts), yielding a recall of 75.68% (28,950 true positive concepts) over the precision of 61.33% (18251 false positives and 9301 false negatives). The Hamming loss (0.038) further validated classification result, revealing that 3.27% of the concepts were wrongly labelled and 96.9% were correctly labelled. The Jaccard (0.60) further supports consistency in classifying the concepts effectively, thus providing a richer coverage of the domain concepts classified across its training cycles.

E. BioBERT as a Relations Validator and Verifier-in-the-loop

BioBERT acted as a biomedical relations precision sieve serving as a cleaner and refiner of the proposed relations, filtering out what was not semantically and factually sound [82]. Adopting the zero-shot validation strategy allowed the model to perform the validation by understanding relational information and semantic similarity based on BioBERT's innate pre-trained knowledge, without the need for additional training [83]. Table 7 shows sampled templates used for the different non-taxonomic relationship types.

Table 7: Summary of sampled relations template		
Relationship category	**Predicate/ Relation Pattern**	**Template**
Causal	caused_by	{a} is caused by {b}
	leads_to	{a} lead to {b}
	triggered_by	{a} is triggered by {b}
	results_in	{a} results in {b}
Risk	risk_factor_for	{a} is a risk factor for {b}
	increases_risk_of	{a} increases the risk of {b}
Therapeutic	treated_by	{a} is treated by {b}
	managed_by	{a} lead to {b}
	alleviated_by	{a} is alleviated by {b}
	responsive_to	{a} is responsive to {b}
	prevented_by	{a} is prevented by {b}
	resistant_to	{a} is resistant to {b}
Functional	interacts_with	{a} interacts with {b}
	modulate	{a} modulates {b}
	inhibits	{a} inhibits {b}
	activates	{a} activates {b}
Comparative	similar_to	{a} is similar to {b}
	alternative_to	{a} is an alternative to {b}
Clinical context	common_in	{a} is common in {b}
	symptom_of	{a} is a symptom of {b}
	complication_of	{a} is a complication of {b}
Temporal	precedes	{a} precedes {b}
	follows	{a} follows {b}
	occurs_during	{a} occurs during {b}
Associative	associated_with	A} is associated with {b}
	linked_to	{a} is linked to {b}
	co_occurs_with	{a} co-occurs with {b}
	correlated_with	{a} is correlated with {b}

The table shows the wide array and diversity of the relations covered based on different relation categories. Each relation type from the BiLSTM output had a template defined that expressed the relation in a declarative canonical sentence form.

a) Confidence Score Threshold Selection

For this work, the confidence threshold was set to 0.4, a low threshold that was focused on giving priority to recall and ensuring as many candidate relations as possible were detected and accepted for further validation processing. This threshold was set purely as a cutoff parameter. The input given to the zero-shot validator were 414 candidate relations of which 384 met the 0.4 threshold and were retained as the final validation output. Of the 414 candidate relations, 30 relations failed to meet the threshold and were discarded.

b) BioBERT Validation Results

Table 8: Summary of Biomedical validated Relations			
Type of relation	**BiLSTM Relations**	**BioBERT Validated Relations**	**Sample validated relations**
Ternary	55	240	*{'symptom': 'low energy', 'disease': 'anxiety', 'treatment': 'negative thought challenge', 'relation': 'managesConditionContext'}*
Non-taxonomic	359	144	*Depression, treatedWith, fluoxetine, fatigue, suggests,anxiety*
Total	414	384	Validation rate:92.7%

BioBERT validated all the ternary relations (100% validation rate) initially extracted by BiLSTM. However, due to the pre-training on biomedical literature BioBERT was able to capture more valid ternary relations contexts, therefore

identifying 240 ternary relations. This was 185 more entries than what the BiLSTM had extracted. This does not indicate a failure by the BiLSTM, but instead confirms that BiLSTM was recall focused, and was limited in precision (section 5.3.5). The 100% validation rate shows that the BiLSTM performed significantly well in extracting complex relations from the unstructured text; that align to existing biomedical knowledge. In contrast, BioBERT validated 144 of the 359 non-taxonomic relations, with a 40.1% validation rate. This drop in relations can be attributed to the fact that BioBERT acted as a biomedical relations precision optimizer. In addition, predicate label semantic mismatch between patient context labels and standard biomedical predicate labels also contributed to the drop.

i) Ternary Relation Extraction

Unlike binary non-taxonomic associations that seek general connections between concepts, in ternary relation structure the relation covers a multi-concept clinical context. The multi-concept clinical context here refers to clinical integral truth conditions (with all the concepts from all the three classes, disease, symptom and treatment participating). This context justified the specific situation why, where and when the relation holds true, without which the relation is invalid.

The validated ternary relations had a consistent pattern and structure that comprised of a subject concept, relation (explaining the nature of the link), object concept with the context attached to it. This was expressed as:

- concept, associated_with, concept | {clinical context}
- Table 10 explicitly demonstrates this ternary structure from the text given.

A person has suffered anxiety for a long time now, it has affected their sleep. They had. previously tried cognitive behavioural therapy to help but it failed after a short while. Instead, the situation became worse. The person then consulted a doctor who prescribed mirtazapine the person's first time on such a medication. The medication worked effectively improving their sleep. However, the effect of the medication after a few days has been dreadful causing loss of energy. The person is aware that these are some of the symptoms.

From the sentence, the framework successfully extract and validate the following:

Table 10: Sample Results of Ternary Relations

Subject Concept	Subject type	Relation	Object Concept	Object type
Anxiety	*Disease*	*affects*	*sleep*	*Symptom | {for a long time >1 year, treated using CBT (treatment)}*
loss of energy	*Symptom*	*caused_by*	*mirtazapine*	*Treatment | {Treated anxiety (disease) temporarily, dreadful effect after few days}*
Anxiety	*Disease*	*treated_by*	*Cognitive Behavioural Therapy (CBT)*	*Treatment | {unable to sleep (symptom), for a long time}*
Anxiety	*Disease*	*treated_by*	*mirtazapine*	*Treatment | {improved sleep, temporarily effective}*

ii) Distribution and Semantic categorization of Non-Taxonomic Relation Extraction

The non-taxonomic relations were further divided into subcategories including: Functional non-taxonomic relations, Associative non-taxonomic relations, Causal non-taxonomic relations, Risk factors non-taxonomic relations, and Statistical non-taxonomic relations. Table 9 gives the results

Table 9: Distribution and Semantic Categorization of Non-taxonomic Relations

BiLSTM Relation category	BioBERT Mapped to	Relations patterns/Predicate labels	Validated Relations Count	Percentage (%)	Sample relations
associated (Comparative: similar_to) temporal	Associative	associatedWith	54	37.50	{'raw': 'D_anxiety', 'label': 'anxiety', 'type': 'disease'} --associatedWith--> {'raw': 'D_heart palpitations, 'label': 'cardiovascular disease', 'type': 'disease'} {'raw': 'D_anxiety', 'label': 'anxiety', 'type': 'disease'} --associatedWith--> {'raw': 'D_respiratory_conditions',

					'label': 'respiratory conditions', 'type': 'disease'}
Associated (Functional: interacts_with)	Functional	responsiveTo,alleviatedBy, selfManageableBy	45	31.25	{'raw': 'D_anxiety', 'label': 'anxiety', 'type': 'disease'} --responsiveTo--> {'raw': 'T_anti_anxiety_medications', 'label': 'anti anxiety medications', 'type': 'treatment'} {'raw': 'S_irritability', 'label': 'irritability', 'type': 'symptom'} --alleviatedBy--> {'raw': 'T_aromatherapy', 'label': 'aromatherapy', 'type': 'treatment'}
Conditional	Causal	causedBy	20	13.89	{'raw': 'S_panic', 'label': 'panic', 'type': 'symptom'} --causedBy--> {'raw': 'D_anxiety', 'label': 'anxiety', 'type': 'disease'} {'raw': 'S_irritability', 'label': 'irritability', 'type': 'symptom'} --causedBy--> {'raw': 'D_anxiety', 'label': 'anxiety', 'type': 'disease'}
None directly identified	Risk Factors	riskFactorFor	20	13.89	{'raw': 'S_panic', 'label': 'panic', 'type': 'symptom'} --riskFactorFor--> {'raw': 'D_anxiety', 'label': 'anxiety', 'type': 'disease'} {'raw': 'S_panic', 'label': 'panic', 'type': 'symptom'} --riskFactorFor--> {'raw': 'D_respiratory_conditions', 'label': 'respiratory conditions', 'type': 'disease'}
None directly identified	Statistical	coOccursWith	5	3.47	{'raw': 'S_panic', 'label': 'panic', 'type': 'symptom'} --coOccursWith--> {'raw': 'S_insomnia ', 'label': 'insomnia', 'type': 'symptom'} {'raw': 'S_panic', 'label': 'panic', 'type': 'symptom'} --coOccursWith--> {'raw': 'S_irritability', 'label': 'irritability', 'type': 'symptom'}

c) Confidence Score Analysis

Table 10 gives the confidence range for the relations together with average confidence for the different categories.

Table 10: Summary of the Confidence Scores per Relation Category

Relation Type	Confidence Range	Average Confidence (%)
Associative	0.917 – 0.984	96
Functional	0.697 – 0.965	88
Causal	0.790 – 0.962	92
Risk Factors	0.704 – 0.947	86

Statistical	0.608 – 0.851	75
Overall Average Confidence		**87.4**

The non-taxonomic relations confidence range was between 0.60-0.98, with an overall confidence average of 87.4%. The results reveal that all the confidence scores were above the 0.4 threshold that was set. The lowest confidence score was 0.60, which was still higher than the threshold that was set. This shows that the 0.4 threshold made it possible to validate a wider variety of relations, while prioritizing recall without compromising the quality of the relations. An analysis of the results reveals that over 90.5% of the validated relations were above 90%, which indicates that BioBERT zero-shot validator captured high quality relations that existed within the pre-existing biomedical knowledge, confirming its role in the framework as a precision focused optimizer.

F. Hybrid Knowledge-Infused Neural Framework

The hybrid framework demonstrated a robust performance improvement achieving an F1 score of 77.6%. The semantic knowledge, fused with contextual learning, inductive reasoning capabilities and domain based validation, gave the approach a significantly stronger recall of 92.8% as shown in Table 11. It demonstrated the framework's expressive, consistent, specialized and contextually aware ability to identify relevant domain concept and capturing complex patterns.

Table 11: Summary of the overall performance of the framework

	Performance			**Hybrid Completeness and Consistency Check**		
Dataset	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Hamming loss**	**Jaccard**	
AnxD and Depression Dataset	66.7	92.8	77.6	0.029	0.667	

When compared to the BiLSTM results with a recall of 75.6%, precision of 61.3%, and F1 score of 67.7%, the hybrid framework achieved a 10% improvement on the F1 score. It also achieved a 17.2% improvement on the recall and a 5.4% improvement on precision. The improvement on the precision and recall indicates the significance of infusing domain knowledge and BioBERT to the framework. The framework focused on optimizing the precision and ensuring the relations were semantically relevant and valid to the biomedical domain. The Jaccard of 0.667, a value closer to 1 means the predicates and true relation labels existing in the domain were nearly identical [84], resulting in reliable relations. This mean the results produced by the model are reliable.

VI. DISCUSSIONS

Symptoms had the highest number of mentions, which is normal considering most of the texts have been written from the patient-perspective. The patient-reported outcome encompasses high number of symptoms that gave unique insights, experiences, and their views regarding their health condition and treatment. This knowledge offers significant value in explaining and treating symptoms [85], [86]. Semantic coherence was necessary for this study because it created unity within concepts found in the different classes providing an easy-to-determine context to aid in the resolution of ambiguity and in the narrowing to a specific meaning of concepts.

The highest confidence score range was 0.917-0.984, assigned to '*associatedWith*'. This indicates that BioBERT was highly confident in validating this type of non-taxonomic relations between different disease, symptom and treatment concepts. The consistent high confidence score within the 54 associative relation category made them a more reliable semantic link within the knowledge graph. The relatively low confidence scores largely for statistical relations (0.60) and risk factor relations (0.70) suggests relations that are indicative and explain factors that influence the occurrence of disorders, but might not be the final or controlling factor to the existence of the disorder [87]. The moderate confidence scores (0.79 - 0.85) predominantly for causal and some statistical relations, indicated that these relations were reliable for inclusion in the knowledge graph, but might require more evidence and verification to increase the level of certainty.

The hybrid knowledge-infused neural approach, integrating symbolic baseline with a BiLSTM framework fused with an attention mechanism demonstrated high levels of efficacy in the extraction and classification of complex medical relations found in free texts. In experimental evaluation, the hybrid framework achieved an accuracy of 98.91% on the test set, a significant improvement over traditional RNNs and standalone BiLSTM algorithms [88]. The framework's strong performance on other cross-validation metrics including a high recall (~93%) and F1 (~78%) highlights the reliability and robustness of the hybrid framework. The results affirm the potential that the framework has in clinical decision support, by refining accuracy and efficiency in information retrieval, and information association discovery.

A. Practicability of Hybrid Framework

The practicability of the framework in real-world clinical and data management is connected to:

- Its reliable levels of accuracy,
- Response speed in providing classification results in a very short time.

This is particularly important in situations where clinicians and key decision makers need to identify key data or concepts and their associations. The practical value of the BiLSTM-based framework fused with an attention mechanism is evident in several data engineering and knowledge management aspects.

The non-taxonomic and ternary relations demonstrated the framework's ability to structure the representation of knowledge from unstructured patient-generated texts. The framework created a structured web of knowledge categorized into functional, causal, associative, and statistical relations, all representing knowledge beyond the broader, narrower and equivalent relations. This provided an automated way of integrating the vocabularies to standardized global domain concepts found in ontologies such as SNOMED CT and MeSH.

The ternary relations preserve clinical contextual structures during semantic integration address a semantic interoperability challenge: the ambiguity of isolated clinical facts. Any system supporting a clinical or patient monitoring decision is equipped with the context (all relevant and valid information) to guide its reasoning towards a semantically consistent interpretation across different health systems.

B. Limitations

The performance of the framework is dependent upon and is influenced by several factors including, balance, and the size of the dataset [89].

C. Imbalanced Dataset

This bias is confirmed by the resulting class distribution where symptoms (54 concepts) was the majority (prevalent) class, followed by treatment (40 concepts), while disease (19) was the minority (rare) class. This mirrored how often the balance is unreached in real-world datasets. With this type of distribution, it was expected that the framework would have failed to make meaningful predictions for the minority classes [90]. The F1 score results of 0.789 for treatment, 0.768 for symptoms, and 0.657 for disease, confirmed the expectation and showed the difficulty the framework faced in accomplishing its objective of accurately predicting the minority class of interest. This was likely due to the fact that the framework had fewer disease patterns to learn from, compared to the symptom and treatment patterns. This still represented the shortcoming of the study approach that can be addressed in future works. Further research should therefore explore techniques for addressing class imbalanced data with promising results including data sampling and cost-sensitive learning, neural network feature learning as demonstrated in prior research works by [91] and [92].

D. Dataset Size

Classification is an inherently complex and becomes more challenging when dealing with smaller datasets potentially leading to a biased classification framework [93], [94]. In this study, the data size affected the framework's ability to learn the different class patterns. An increase in the number of the disease data samples to learn from would have given the framework a good discriminative power between classes of interest. Future research should therefore focus on enhancing the technique by adopting strategies such as increasing the data size discussed by [94] and refining techniques like adaptive learning rate, to improve the robustness of hybrid framework when trained on limited data. By recognizing and tackling these limitations and challenges, future studies can significantly enhance the accuracy, robustness and applicability of the hybrid framework in a real-world medical environment.

VII. CONCLUSION

This makes the framework useful for knowledge extraction tasks requiring comprehensive semantic coverage. This is critical in ensuring information needed for: knowledge representation, informed decision making, semantic dependency building is available. In such applications and tasks, overlooking key concepts can be more costly than retrieving occasional false positives making the trade-off a justifiable choice for this study.

- **Funding:** This work was supported by METEGA.
- **Ethics declarations:** Competing Interests
- **Ethics Approval and Consent to Participate:** Not applicable.
- **Consent for Publication:** Not applicable.

VIII. REFERENCES

[1] L. Tao, "Extending OWL with Custom Relations for Knowledge-Driven Intelligent Agents," in German Conference on Multiagent System Technologies, Springer, 2017, pp. 156–166.

[2] H. Borgelt, F. G. Kitel, and N. Kockmann, "Ontology-Based Data Pipeline for Semantic Reaction Classification and Research Data Management," Computers, vol. 14, no. 8, p. 311, 2025.

[3] B. Kass, "Top 5 Tips for Managing and Versioning an Ontology," Enterprise Knowledge. Accessed: Mar. 22, 2025. [Online]. Available: https://enterprise-knowledge.com/top-5-tips-for-managing-and-versioning-an-ontology/

[4] C. Jonquet and M. Poveda-Villalón, "About versioning ontologies or any digital objects with clear semantics," in DaMaLOS 2023 - 3rd Workshop on Metadata and Research (objects) Management for Linked Open Science, Hersonissos (Crète), Greece: L. J. Castro and S. Schimmler and J. Dierkes and D. Dessì and D. Rebholz-Schuhmann, May 2023. doi: 10.4126/FRL01-006444994.

[5] M. Sundgren, R. Mistry, and G. Maeztu, "Harnessing Unstructured Data and Hospital Interoperability," Appl. Clin. Trials, vol. 33, no. 11, 2024.

[6] H. Akremi and S. Zghal, "Holistic Ontology Alignment Using Ontologies Embeddings," in Intelligent Systems and Pattern Recognition, A. Bennour, A. Bouridane, S. Almaadeed, B. Bouaziz, and E. Edirisinghe, Eds., Cham: Springer Nature Switzerland, 2025, pp. 203–213.

[7] M. Abdullahi, S. Aliyu, and A. F. Kana, "A Novel Model for the Semantic Enrichment of an Ontology Alignment System," FUOYE J. Eng. Technol., vol. 7, Jul. 2022, doi: 10.46792/fuoyejet.v7i3.819.

[8] C. Trojahn, R. Vieira, D. Schmidt, A. Pease, and G. Guizzardi, "Foundational ontologies meet ontology matching: A survey," Semantic Web, vol. 13, no. 4, pp. 685–704, 2022, doi: 10.3233/SW-210447.

[9] A. Lehnert, "Building Healthy Relationships (in Ontologies)," synaptica. [Online]. Available: https://synaptica.com/building-healthy-relationships-in-ontologies/

[10] S. Fan et al., "Graph-Enhanced Dual-Stream Feature Fusion with Pre-Trained Model for Acoustic Traffic Monitoring," in ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2025, pp. 1–5.

[11] Benedictine University Library, Benedictine University Library. Accessed: Aug. 01, 2025. [Online]. Available: https://researchguides.ben.edu/c.php?g=1196526&p=8764909

[12] P. Li and Y. Hu, "Evaluating online doctor–patient communication quality and its effects: Text analysis approach," Digit. Health, vol. 11, p. 20552076251395547, Apr. 2025, doi: 10.1177/20552076251395547.

[13] K. Samanta and D. Rath, "Controlled Terms Versus Uncontrolled Terms in Resource Description," DESIDOC J. Libr. Inf. Technol., vol. 44, pp. 158–167, May 2024, doi: 10.14429/djlit.44.3.19554.

[14] T. Stoeckel, T. Ishii, Y. A. Kim, H. T. Ha, N. T. P. Ho, and S. McLean, "A comparison of contextualized and non-contextualized meaning-recall vocabulary test formats," Res. Methods Appl. Linguist., vol. 2, no. 3, p. 100075, Dec. 2023, doi: 10.1016/j.rmal.2023.100075.

[15] P. Kowalski and A.-L. Jousselme, "Context-awareness for information correction and reasoning in evidence theory," Int. J. Approx. Reason., vol. 153, pp. 29–48, 2023, doi: https://doi.org/10.1016/j.ijar.2022.11.009.

[16] T. Xie et al., "ByteScience: Bridging Unstructured Scientific Literature and Structured Data with Auto Fine-tuned Large Language Model in Token Granularity," in 2024 IEEE International Conference on Data Mining Workshops (ICDMW), Los Alamitos, CA, USA: IEEE Computer Society, Dec. 2024, pp. 907–911. doi: 10.1109/ICDMW65004.2024.00126.

[17] J. Gudu, J. Balikuddembe, J. Mwebaze, and D. Opiyo, "Extracting Non-Taxonomic and Ternary Relations from Patient-Generated Texts for Semantic Interoperability," 2026.

[18] Y. Yu and X. Liu, "Research on enterprise text classification methods of BiLSTM and CNN based on BERT," in Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence, 2023, pp. 491–495.

[19] B. Jang, M. Kim, G. Harerimana, S. Kang, and J. W. Kim, "Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism," Appl. Sci., vol. 10, no. 17, p. 5841, 2020.

[20] R. Alam, "Understanding Semantic Relations: Types, Examples, and Applications," Linkedin. Accessed: Aug. 02, 2025. [Online]. Available: https://www.linkedin.com/pulse/understanding-semantic-relations-types-examples-md-rabby-alam-ndewf/

[21] J. C. Zemla, "Knowledge representations derived from semantic fluency data," Front. Psychol., vol. 13, p. 815860, 2022.

[22] Coalesce Catalog, "Semantic Ontology: Understanding Data Relationships and Hierarchies," Catalog from Coalesce. [Online]. Available: https://www.castordoc.com/data-strategy/semantic-ontology-understanding-data-relationships-and-hierarchies#:~:text=Types%20of%20Data%20Relationships%20in,context%2C%20providing%20flexibility%20in%20interpretation.

[23] N. V. Loukachevitch, "Automatic Methods for Extracting Taxonomic Relationships from Texts," Pattern Recognit. Image Anal., vol. 33, no. 3, pp. 398–406, Sep. 2023, doi: 10.1134/S1054661823030276.

[24] R. Du, H. An, K. Wang, and W. Liu, "A short review for ontology learning from text: Stride from shallow learning, deep learning to large language models trend," ArXiv Prepr. ArXiv240414991, 2024.

[25] H. Gharagozlou, J. Mohammadzadeh, A. Bastanfard, and S. S. Ghidary, "Semantic Relation Extraction: A Review of Approaches, Datasets, and Evaluation Methods With Looking at the Methods and Datasets in the Persian Language," ACM Trans Asian Low-Resour Lang Inf Process, vol. 22, no. 7, Jul. 2023, doi: 10.1145/3592601.

[26] M. Lentschat, P. Buche, J. Dibie-Barthelemy, and M. Roche, "A new method to extract n-Ary relation instances from scientific documents," Expert Syst. Appl., vol. 209, p. 118332, Dec. 2022, doi: 10.1016/j.eswa.2022.118332.

[27] J. L. M. Fernandes, "Extracting n-ary Relations from Biomedical Literature using Deep-Learning Techniques," PhD Thesis, UNIVERSIDADE DE LISBOA, 2025.

[28] S. Choi and Y. Jung, "Knowledge Graph Construction: Extraction, Learning, and Evaluation," Appl. Sci., vol. 15, no. 7, 2025, doi: 10.3390/app15073727.
[29] J. McCusker, "LOKE: Linked Open Knowledge Extraction for Automated Knowledge Graph Construction." 2023. [Online]. Available: https://arxiv.org/abs/2311.09366
[30] S. Khorshidi et al., "ODKE+: Ontology-Guided Open-Domain Knowledge Extraction with LLMs," ArXiv Prepr. ArXiv250904696, 2025.
[31] R. Kondo, S. Kundu, M. Imai, K. Tomoda, Y. Takazawa, and R. Hisano, "Constraint-Aware Ontology Engineering for Information Extraction from Financial Contracts," 2025.
[32] X. Zhao et al., "A Comprehensive Survey on Relation Extraction: Recent Advances and New Frontiers," ACM Comput Surv, vol. 56, no. 11, Jul. 2024, doi: 10.1145/3674501.
[33] T. Aastha, "A Review Study of Natural Language Processing Techniques for Text Mining," Int. J. Eng. Res. Technol. IJERT, vol. 10, no. 9, 2021.
[34] X. X. Qian, P. H. Chau, D. Y. T. Fong, M. Ho, and J. Woo, "Development and Validation of a Rule-Based Natural Language Processing Algorithm to Identify Falls in Inpatient Records of Older Adults: Retrospective Analysis," JMIR Aging, vol. 8, p. e65195, Jul. 2025, doi: 10.2196/65195.
[35] G. Yu, "A combined rule-based and machine learning approach for blackout analysis using natural language processing," Masters, 2022.
[36] S. V. Mahadevkar, S. Patil, K. Kotecha, L. W. Soong, and T. Choudhury, "Exploring AI-driven approaches for unstructured document analysis and future horizons," J. Big Data, vol. 11, no. 1, p. 92, Jul. 2024, doi: 10.1186/s40537-024-00948-z.
[37] S. J. Skeen, S. S. Jones, C. M. Cruse, and K. J. Horvath, "Integrating Natural Language Processing and Interpretive Thematic Analyses to Gain Human-Centered Design Insights on HIV Mobile Health: Proof-of-Concept Analysis.," JMIR Hum. Factors, vol. 9, no. 3, p. e37350, Jul. 2022, doi: 10.2196/37350.
[38] M. Tayefi et al., "Challenges and opportunities beyond structured data in analysis of electronic health records," WIREs Comput. Stat., vol. 13, no. 6, p. e1549, Nov. 2021, doi: 10.1002/wics.1549.
[39] K. Ramar and G. Gurunathan, "Technical Review on Ontology Mapping Techniques," Asian J. OfInformation Technol., vol. 15, no. 4, pp. 676-688, 2016.
[40] A. Chauhan, L. Sliman, and others, "Ontology matching techniques: a gold standard model," ArXiv Prepr. ArXiv181110191, 2018.
[41] O. Araque, G. Zhu, and C. A. Iglesias, "A semantic similarity-based perspective of affect lexicons for sentiment analysis," Knowl.-Based Syst., vol. 165, pp. 346–359, Feb. 2019, doi: 10.1016/j.knosys.2018.12.005.
[42] A. H. Muhammad, I. Oyong, and others, "Revisiting the challenges and surveys in text similarity matching and detection methods.," J. Inform., vol. 16, no. 3, 2022.
[43] X. Liu, Q. Tong, X. Liu, and Z. Qin, "Ontology Matching: State Of The Art, Future Challenges, And Thinking Based On Utilized Information," IEEE Access, vol. 9, 2021.
[44] V. Pichiyan, S. Muthulingam, S. G, S. Nalajala, A. Ch, and M. N. Das, "Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis," 3rd Int. Conf. Evol. Comput. Mob. Sustain. Netw. ICECMSN 2023, vol. 230, pp. 193–202, Jan. 2023, doi: 10.1016/j.procs.2023.12.074.
[45] D. Seenivasan, "ETL in a World of Unstructured Data: Advanced Techniques for Data Integration," Int. J. Manag. Eng., vol. 11, no. 1, pp. 127–145, 2021, doi: http://dx.doi.org/10.2139/ssrn.5148188.
[46] S. S. Vichare, "Probabilistic Ensemble Machine Learning Approaches For Unstructured Textual Data Classification," Masters, THE PURDUE UNIVERSITY, West Lafayette, Indiana, 2024.
[47] E. Anderson et al., "The design of an llm-powered unstructured analytics system," ArXiv Prepr. ArXiv240900847, 2024.
[48] Y. Li, Z. Luan, Y. Liu, H. Liu, J. Qi, and D. Han, "Automated information extraction model enhancing traditional Chinese medicine RCT evidence extraction (Evi-BERT): algorithm development and validation.," Front. Artif. Intell., vol. 7, p. 1454945, 2024, doi: 10.3389/frai.2024.1454945.
[49] T. Zengeya and J. V. Fonou Dombeu, "A Review of State of the Art Deep Learning Models for Ontology Construction," IEEE Access, vol. PP, pp. 1–1, Jan. 2024, doi: 10.1109/ACCESS.2024.3406426.
[50] J. Ji, B. Chen, and H. Jiang, "Fully-connected LSTM–CRF on medical concept extraction," Int. J. Mach. Learn. Cybern., vol. 11, Sep. 2020, doi: 10.1007/s13042-020-01087-6.
[51] H. Lei, "The Difference between BiLSTM and Transformer-The Discussion about Suitable Scenarios for BiLSTM and Transformer," Sci. Technol. Eng. Chem. Environ. Prot., vol. 1, no. 1, 2025.
[52] H. N. Cho et al., "Task-specific transformer-based language models in health care: scoping review," JMIR Med. Inform., vol. 12, p. e49724, 2024.
[53] J. Lee et al., "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.
[54] G. Wang et al., "Optimized glycemic control of type 2 diabetes with reinforcement learning: a proof-of-concept trial," Nat. Med., vol. 29, no. 10, pp. 2633–2642, 2023.
[55] V. Visweswaraiah, "Everyday AI: Real-World Applications of Transformer- Based Language Models," Int. J. Comput. Trends Technol., vol. 73, pp. 19–27, Sep. 2025, doi: 10.14445/22312803/IJCTT-V73I9P103.
[56] M. Chelli et al., "Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis," J Med Internet Res, vol. 26, p. e53164, May 2024, doi: 10.2196/53164.

[57] T. V. Ivanisenko, P. S. Demenkov, and V. A. Ivanisenko, “An Accurate and Efficient Approach to Knowledge Extraction from Scientific Publications Using Structured Ontology Models, Graph Neural Networks, and Large Language Models,” Int. J. Mol. Sci., vol. 25, no. 21, 2024, doi: 10.3390/ijms252111811.
[58] Navigate Health, “Anxiety disorders,” Navigate Health. Accessed: Jul. 26, 2025. [Online]. Available: https://community.patient.info/tag/anxiety%20disorders/
[59] HealthUnlocked, “Anxiety and Depression Support,” HealthUnlocked. Accessed: Jul. 26, 2025. [Online]. Available: https://healthunlocked.com/anxiety-depression-support
[60] Together 4 Change Limited, “Mental Health Forum.” [Online]. Available: https://www.mentalhealthforum.net/forum/threads/do-you-experience-anxiety-approved-research.394720/
[61] D. S. França, “Depression Dataset.” Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/diegosilvadefrana/depression-dataset
[62] M. Lubani, S. A. M. Noah, and R. Mahmud, “Ontology population: Approaches and design aspects,” J. Inf. Sci., vol. 45, no. 4, pp. 502–515, Aug. 2019, doi: 10.1177/0165551518801819.
[63] N. Kuusik and J. Vain, “Ontology Merging Using the Weak Unification of Concepts,” Big Data Cogn. Comput., vol. 8, no. 9, 2024, doi: 10.3390/bdcc8090098.
[64] National Library of Medicine, “MeSH Terminology Standards and Tools,” Subject Heading. Accessed: Oct. 04, 2025. [Online]. Available: https://www.nlm.nih.gov/mesh/intro_trees.html
[65] SNOMED International, “SNOMED CT Expressions,” SNOMED CT Expressions. Accessed: Feb. 11, 2026. [Online]. Available: https://docs.snomed.org/snomed-ct-practical-guides/snomed-ct-starter-guide/7-snomed-ct-expressions
[66] IHTSDO, “SNOMED Clinical Terms® User Guide.” The International Health Terminology Standards Development Organisation, 2008. [Online]. Available: https://www.nlm.nih.gov/research/umls/Snomed/SNOMED_CT_User_Guide_20080731.pdf
[67] J. Mökander, J. Schuett, H. R. Kirk, and L. Floridi, “Auditing large language models: a three-layered approach,” AI Ethics, vol. 4, no. 4, pp. 1085–1115, Nov. 2024, doi: 10.1007/s43681-023-00289-2.
[68] S. Chen, “CSV Files: Use cases, Benefits, and Limitations,” OneSchema. Accessed: Sep. 27, 2025. [Online]. Available: https://www.oneschema.co/blog/csv-files#advantages-of-csv-files-3
[69] J. Mitlöhner, S. Neumaier, J. Umbrich, and A. Polleres, Characteristics of Open Data CSV Files. 2016, p. 79. doi: 10.1109/OBD.2016.18.
[70] A. Pliatsios, K. Kotis, and C. Goumopoulos, “A systematic review on semantic interoperability in the IoE-enabled smart cities,” Internet Things, vol. 22, p. 100754, Jul. 2023, doi: 10.1016/j.iot.2023.100754.
[71] G. Lan, T. Liu, X. Wang, X. Pan, and Z. Huang, “A semantic web technology index,” Sci. Rep., vol. 12, no. 1, p. 3672, Mar. 2022, doi: 10.1038/s41598-022-07615-4.
[72] M. S. M. Rudwan and J. V. Fonou-Dombeu, “A Novel Algorithm for Multi-Criteria Ontology Merging through Iterative Update of RDF Graph,” Big Data Cogn. Comput., vol. 8, no. 3, 2024, doi: 10.3390/bdcc8030019.
[73] J. Li et al., “A comprehensive review of community detection in graphs,” Neurocomputing, vol. 600, p. 128169, 2024.
[74] V. Iyer, L. M. Sanagavarapu, and R. Reddy, “A framework for syntactic and semantic quality evaluation of ontologies.” 2021. [Online]. Available: https://arxiv.org/abs/2112.08543
[75] T. Kupiec, D. Wojtowicz, and K. Olejniczak, “Structures and functions of complex evaluation systems: comparison of six Central and Eastern European countries,” Int. Rev. Adm. Sci., vol. 89, p. 002085232110269, Aug. 2021, doi: 10.1177/00208523211026964.
[76] A. Abbas et al., “Clinical Concept Extraction with Lexical Semantics to Support Automatic Annotation,” Int. J. Environ. Res. Public. Health, vol. 18, no. 20, 2021, doi: 10.3390/ijerph182010564.
[77] G. Berge, O.-C. Granmo, T. Tveit, A. Ruthjersen, and J. Sharma, “Combining unsupervised, supervised and rule-based learning: the case of detecting patient allergies in electronic health records,” BMC Med. Inform. Decis. Mak., vol. 23, Sep. 2023, doi: 10.1186/s12911-023-02271-8.
[78] E. da Silva Farias and E. Palhares Junior, “Application of Attention Mechanism with Bidirectional Long Short-Term Memory (BiLSTM) and CNN for Human Conflict Detection using Computer Vision,” ArXiv E-Prints, p. arXiv-2502, 2025.
[79] M. Jin, S.-M. Choi, and G.-W. Kim, “COMCARE: A Collaborative Ensemble Framework for Context-Aware Medical Named Entity Recognition and Relation Extraction,” Electronics, vol. 14, no. 2, 2025, doi: 10.3390/electronics14020328.
[80] M. M. Lawal and A. Abdulrauf, “Fake News Detection Using Bi-LSTM Architecture: A Deep Learning Approach on the ISOT Dataset,” J. Comput. Theor. Appl., vol. 3, no. 2, pp. 104–114, 2025.
[81] C. Desai, “Impact of Weight Initialization Techniques on Neural Network Efficiency and Performance: A Case Study with MNIST Dataset,” vol. 13, p. 26115, Apr. 2024, doi: 10.18535/ijecs/v13i04.4809.
[82] K. Nastou, M. Koutrouli, S. Pyysalo, and L. J. Jensen, “Improving dictionary-based named entity recognition with deep learning,” Bioinformatics, vol. 40, no. Supplement_2, pp. ii45–ii52, 2024.
[83] X. Dong, D. Zhao, J. Meng, B. Guo, and H. Lin, “SyRACT: zero-shot biomedical document-level relation extraction with synergistic RAG and CoT,” Bioinformatics, vol. 41, no. 7, p. btaf356, 2025.
[84] M. Jadeja, “Jaccard Similarity Made Simple: A Beginner’s Guide to Data Comparison,” Medium. [Online]. Available: https://mayurdhvajsinhjadeja.medium.com/jaccard-similarity-34e2c15fb524
[85] M. Missel et al., “Understanding symptoms in the lives of adult patients with acute or chronic illness: a phenomenological study of patient experiences,” Int. J. Qual. Stud. Health Well-Being, vol. 20, no. 1, p. 2534871, Dec. 2025, doi: 10.1080/17482631.2025.2534871.
[86] M. Tringale, G. Stephen, A.-M. Boylan, and C. Heneghan, “Integrating patient values and preferences in healthcare: a systematic review of qualitative evidence.,” BMJ Open, vol. 12, no. 11, p. e067268, Nov. 2022, doi: 10.1136/bmjopen-2022-067268.

[87] F. Albuquerque, E. Monteiro, and M. A. B. Rodrigues, “The Explanatory Factors of Risk Disclosure in the Integrated Reports of Listed Entities in Brazil,” Risks, vol. 11, no. 6, 2023, doi: 10.3390/risks11060108.
[88] Z. Ruiye, “Volleyball training video classification description using the BiLSTM fusion attention mechanism,” Heliyon, vol. 10, no. 15, p. e34735, Aug. 2024, doi: 10.1016/j.heliyon.2024.e34735.
[89] B. He, Y. Yang, L. Wang, and J. Zhou, “The text classification method based on BiLSTM and multi-scale CNN,” Comput. Life, vol. 12, no. 2, pp. 43–49, 2024.
[90] R. S. Alkhawaldeh, J. AlShaqsi, B. Al-Ahmad, S. M. Alkhawaldeh, and O. Drogham, “An attention-based multi-residual and BiLSTM architecture for early diagnosis of autism spectrum disorder,” Sci. Rep., vol. 15, no. 1, p. 33608, Sep. 2025, doi: 10.1038/s41598-025-19006-6.
[91] J. Leo, E. Luhanga, and K. Michael, “Machine Learning Model for Imbalanced Cholera Dataset in Tanzania,” Sci. World J., vol. 2019, no. 1, p. 9397578, Jan. 2019, doi: 10.1155/2019/9397578.
[92] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” J. Big Data, vol. 6, no. 1, p. 27, Mar. 2019, doi: 10.1186/s40537-019-0192-5.
[93] M. Figeys, F. Koubasi, D. Hwang, A. Hunder, A. Miguel-Cruz, and A. Ríos Rincón, “Challenges and promises of mixed-reality interventions in acquired brain injury rehabilitation: A scoping review,” Int. J. Med. Inf., vol. 179, p. 105235, Nov. 2023, doi: 10.1016/j.ijmedinf.2023.105235.
[94] D. Rajput, W.-J. Wang, and C.-C. Chen, “Evaluation of a decided sample size in machine learning applications,” BMC Bioinformatics, vol. 24, no. 1, p. 48, Feb. 2023, doi: 10.1186/s12859-023-05156-9.