Integrating NANDA-I Diagnoses with Functional Health Patterns: A Semantic Similarity Approach

Takahiro Kubo
Graduate School of Data Science, Faculty of Nursing Musashino University
Tokyo, Japan
ta_kubo@musashino-u.ac.jp

Virach Sornlertlamvanich
Faculty of Data Science
Musashino University
Tokyo, Japan
httpss://orcid.org/0000-0002-6918-8713

Thatsanee Charoenporn
Faculty of Data Science
Musashino University
Tokyo, Japan
https://orcid.org/0000-0002-9577-9082

***Abstract*—This study proposes a data-driven method to map NANDA-I nursing diagnoses to Gordon's 11 Functional Health Patterns (FHP) using text similarity techniques. The goal is to support nurses in objectively identifying appropriate care needs without relying solely on individual clinical experience. The method leverages Sentence-BERT (SBERT) for contextual vectorization, WordNet for semantic expansion, and an individual threshold method for matching. Evaluation using 56 NANDA-I diagnosis instances revealed that the SBERT and WordNet configuration achieved the highest F1 score. SBERT significantly outperformed the classical TF-IDF method due to its ability to capture nuanced semantic relationships. While WordNet improved TF-IDF performance, its impact on SBERT was limited. The individual threshold method performed better than the weighted integration method, emphasizing the importance of threshold selection based on application goals. Despite challenges in mapping due to conceptual differences and linguistic ambiguity between NANDA-I and FHP, the study demonstrates that data-driven approaches can improve the accuracy and consistency of nursing assessments. Future work should explore larger datasets and diverse modeling techniques.**

Keywords*—*Nursing Diagnoses, Functional Health Patterns, BERT

I. INTRODUCTION

In recent years, shortages in nursing staff and increased workloads have made it challenging to maintain high-quality nursing care [1]. Quality nursing practice requires accurate patient assessment through comprehensive data collection and analysis, alongside appropriate nursing diagnoses. Nursing diagnoses are critical as they underpin the direction of nursing interventions, significantly impacting patient outcomes [2]. However, the accuracy of nursing diagnoses often relies heavily on individual nurses' experience and judgment, resulting in significant variability among nurses even when interpreting the same patient data. This variability poses particular challenges for inexperienced nurses, highlighting the need to standardize nursing diagnoses.

Given these issues, there is growing demand for objective and comprehensive assessment algorithms. Recent studies indicate that nursing diagnosis support systems can improve diagnostic accuracy. However, limitations in the comprehensiveness and inference accuracy of proposed diagnoses hinder their practical clinical applications [3]. Additionally, problems such as incomplete data recording, difficulty tracing analytical processes, and inadequate reflection of patient-specific information have been reported in electronic health record (EHR)-integrated support systems due to facility-specific customizations [4].

Moreover, a significant gap exists between Gordon's 11 Functional Health Patterns (FHP) [8], a widely used nursing assessment framework, and NANDA-I nursing diagnoses, a standardized nursing diagnosis classification. Currently, mapping between these two frameworks relies largely on individual nurses' expertise, making comprehensive problem extraction and subsequent diagnoses challenging, especially for novice nurses.

The high cognitive load imposed on nurses during the diagnostic process can adversely affect the accuracy of their clinical judgment [5]. To address this challenge, we proposed a foundational BERT-based nursing diagnosis support system in a previous study [6]. However, this initial approach, which directly mapped diverse patient observations to NANDA-I diagnoses [7], demonstrated limitations in its predictive accuracy.

Improving this accuracy requires a model that more faithfully represents the nurses' cognitive workflow. Specifically, we identified that systematically mapping FHP to NANDA-I diagnoses is an essential step. This mapping establishes an intermediary layer—the FHPs, which are aligned with nursing practice—between concrete yet diverse "observation items" and standardized yet abstract "NANDA-I diagnoses." This layer allows the system to capture contextual relationships, thereby enabling more precise diagnostic support.

Therefore, this study proposes and evaluates a data-driven approach to construct this intermediary layer, aiming to assess its practical effectiveness. Such an approach would support nurses in objectively identifying appropriate diagnoses, reducing over-reliance on individual experience. The conceptual framework of our study is illustrated in Figure 1.

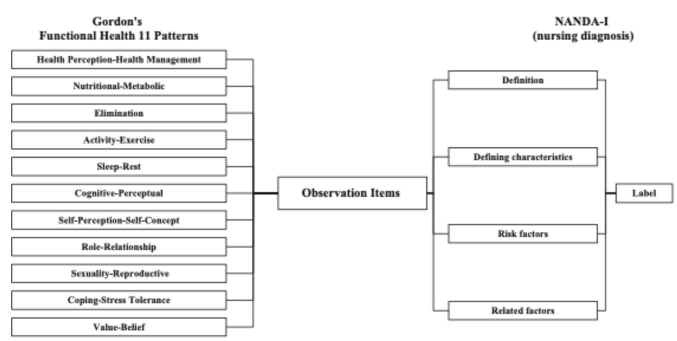

Fig. 1. Conceptual framework of the proposed method.

2025 International Electronics Symposium (IES) | 979-8-3315-5415-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/IES67184.2025.11161475

The paper is structured as follows. Section II reviews related work. Section III describes the methodology implemented to achieve acceptable mapping accuracy. Section IV outlines the experimental configurations that contribute to variations in the results. In Section Ⅴ, we discuss the limitations of the proposed methods and suggest directions for future improvement. Section Ⅵ presents the conclusion.

II. RELATED WORK

Recent advances in natural language processing (NLP) have established distributed representations, word and sentence embeddings, as the standard for representing clinical text, facilitating automated analysis of free-text nursing and medical records. A recent integrative review by Mitha et al. [9] further confirmed that NLP applied to nursing notes is a key approach for identifying care patterns and supporting clinical judgment. Transformer-based models such as BERT, XLNet, and RoBERTa now achieve Pearson correlations up to 0.91 on clinical semantic textual similarity benchmarks [10].. Traditionally, semantic relatedness between UMLS concepts has been measured using path-based or information content–based methods. In contrast, Mao and Fung (2020) demonstrated that combining text-derived embeddings with graph-based embeddings from UMLS hierarchies yields higher semantic relatedness scores compared to such conventional approaches [11].

In nursing informatics, Karacan et al. applied BERT-based phrase embeddings to align EPIC care-plan entries with standardized nursing terminologies (NANDA-I, NIC, NOC). They found that both general-domain models (e.g., Sentence-BERT) and clinical-specific models (e.g., BioClinicalBERT) performed well, with no consistent advantage of specialized models over general ones [12]. Reflecting the rapid evolution of this field, the latest research trends even explore generative AI for recommending nursing diagnoses directly from electronic records [13]. For languages without explicit word boundaries, Mutinda et al. demonstrated that a general-domain Japanese BERT slightly outperforms a clinical-specific model in capturing sentence-level similarity on Japanese clinical texts [14].

Finally, Beam et al. developed concept embeddings (cui2vec) by integrating massive, disparate health data sources—claims, clinical notes, and biomedical literature—into a common semantic space. Their work demonstrated that these embeddings effectively capture semantic relationships between medical concepts, thereby facilitating the identification of related concepts across these diverse data types [15].

Collectively, these studies underscore the power of embedding-driven semantic similarity for mapping clinical concepts. Our work extends this foundation by introducing an indirect mapping between NANDA-I diagnoses and Gordon's Functional Health Patterns via an observation-item intermediary—a perspective not previously explored.

III. METHODOLOGY

A. Datasets

This study utilizes two primary datasets: a list of NANDA-I nursing diagnoses and an Observation Item Table based on Gordon's functional health patterns.

NANDA-I Nursing Diagnoses Dataset: This dataset was constructed from the official NANDA-I Definitions and Classification guide (2021-2023). It contains 56 distinct nursing diagnoses relevant to this study. Each row represents a single diagnosis and includes columns for its label, definition, defining characteristics, related factors, and risk factors.

Observation Item Table: This table was developed based on the assessment guidelines from Marjory Gordon's "Manual of Nursing Diagnosis," which details 11 functional health patterns (FHP). The dataset consists of 268 specific observation items (e.g., "How have your general health been recently?") categorized under these 11 patterns. Each row corresponds to a single observational question or item used by nurses during the assessment phase.

For the purpose of this study's core mapping experiment, the primary inputs are the canonical texts of the NANDA-I diagnostic components and the standardized observation items. The target output is a validated mapping, represented by similarity scores, that quantifies the semantic relationship between these two frameworks.

B. Proposed Method and Experimental Setup

Our research was conducted in a cloud-based Jupyter Notebook environment on Google Colaboratory, which provides scalable computational resources and access to powerful GPUs. The primary programming language used was Python (version 3.11), and the analysis relied on several key open-source libraries, including:

- Pandas: For data loading, manipulation, and management of our datasets.
- Sentence-Transformers: To utilize the pre-trained SBERT model for generating high-quality semantic embeddings from Japanese text.
- spaCy (with GiNZA): For advanced Japanese natural language processing (NLP), specifically for tokenization and lemmatization (extracting the base form of words).
- NLTK: To access the Japanese WordNet corpus for semantic text expansion.
- Scikit-learn: For calculating cosine similarity and performing other machine learning-related tasks.
- Matplotlib & Seaborn: For data visualization and plotting the results of our analyses.

C. Experimental Setup

1) Data Loading and Preprocessing:

The process begins by loading the NANDA-I diagnosis data and the Observation Item Table from CSV files into pandas DataFrames. All textual data, including nursing records and diagnostic definitions, undergo a rigorous preprocessing pipeline. This involves applying our preprocess_text function, which leverages the Japanese NLP model ja_ginza_electra via spaCy to perform lemmatization, while also removing stop words and punctuation to normalize the text for analysis.

2) Semantic Text Expansion:

To enrich the semantic content of the normalized text, we implemented a text expansion step using the expand_text_with_wordnet function. This function queries the Japanese WordNet corpus to find relevant synonyms and hypernyms for nouns and verbs within the text. This expansion helps to capture a broader range of related concepts that may not be explicitly mentioned in the original text.

3) Text Vectorization with SBERT:

The preprocessed and expanded texts are then converted into high-dimensional numerical vectors using a pre-trained Japanese Sentence-BERT (SBERT) model, specifically sonoisa/sentence-bert-base-ja-mean-tokens-v2. This model is adept at capturing the contextual meaning of sentences, transforming them into dense embeddings that are suitable for semantic comparison. This vectorization is applied to all canonical texts used in the mapping: the texts for each of the NANDA-I components (Characteristics, Related Factors, etc.) and the guideline texts for each observation item in the FHP table.

4) Similarity Calculation and Threshold-based Association:

The final stage involves calculating the semantic similarity between the vector of each FHP observation item and the vector of each NANDA-I diagnostic component (e.g., Definition, Characteristics) using the cosine_similarity function. This yields a matrix of similarity scores, quantifying the relatedness between every observation item and every diagnostic component..

To identify the most relevant associations, a thresholding mechanism is applied to these scores. Our study experimented with both fixed-value thresholds (e.g., 0.5, 0.7, 0.9) and data-driven percentile-based thresholds (e.g., 70th, 80th, 90th percentile) to determine the optimal method for association.

5) Association Strategy

To systematically associate nursing diagnoses with functional health patterns, we designed and compared two distinct association methods: the Individual Thresholding Method and the Weighted Integration Method.

Individual Thresholding Method: This approach independently assesses the relevance of each NANDA-I component (i.e., Definition, Characteristics, Related Factors, and Risk Factors). Cosine similarity scores are calculated for each component against the observation items. An observation item is considered relevant if its similarity score exceeds a predefined threshold. We experimented with both fixed-value thresholds (0.5, 0.7, 0.9) and data-driven percentile-based thresholds (70th, 80th, 90th) to evaluate the optimal setting.

Weighted Integration Method: In contrast, this method combines the similarity scores from all four NANDA-I components into a single, integrated score for each observation item. To achieve this, the cosine similarity score of each component was multiplied by a specific weight, and these values were summed. The weights were determined based on the relative clinical importance of each component in the nursing diagnostic process. For instance, one set of weights we evaluated was: Definition: 0.2, Characteristics: 0.3, Related Factors: 0.3 Risk Factors: 0.2

A final association was made by applying a fixed threshold (e.g., 0.5, 0.6) to this final integrated score. This dual-method approach, combined with experiments comparing vectorization techniques (SBERT vs. TF-IDF) and the use of semantic expansion (with vs. without WordNet), allowed for a comprehensive, multi-faceted investigation into the most effective association strategy.

The system we designed to link the above nursing records with NANDA-I nursing diagnoses is shown in Figure 2.

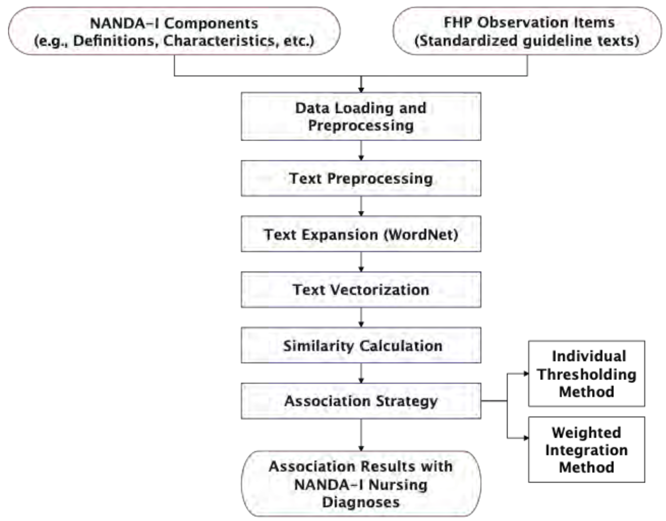

Fig. 2. Overview of the Text Processing Flow

D. Evaluation Method

Evaluation utilized case study data quantitatively. The top 5 predicted FHP items were compared against expert-annotated ground truth data. Evaluation metrics calculated included Precision (ratio of correct predictions within top 5), Recall (ratio of actual relevant FHP identified within top 5 predictions), and F1 score (harmonic mean balancing precision and recall). These metrics were averaged across all experiments to identify optimal methods and parameters.

IV. RESULTS

To evaluate the performance of the proposed method, 56 instances of NANDA-I diagnoses extracted from 11 nursing case studies were used, with Gordon's Functional Health Patterns (FHP) mapped by experts serving as the ground truth. For each experimental setting, the top five predicted FHP items were used as the predicted set, and Precision, Recall, and F1 score were calculated.

A. Main Comparison Results

Table 1 shows the average Precision, Recall, and F1 scores for the top five and bottom five experimental settings.

TABLE I. PERFORMANCE COMPARISON (TOP 5 AND BOTTOM 5)

Rank	Setting	Vectorizer	WordNet	Method	Thr_Type	Thr_Value	Precision	Recall	F1
1	SBERT_WN_Ind_Fix07	SBERT	Yes	Individual	Fixed	0.7	0.3033	0.6844	0.4072
2	SBERT_WN_Wgt1_Fix05	SBERT	Yes	Weighted(v1)	Fixed	0.5	0.2820	0.6626	0.3805
3	SBERT_NoWN_Ind_Pct70	SBERT	No	Individual	Percentile	70%	0.2754	0.6325	0.3702
4	SBERT_WN_Ind_Fix05	SBERT	Yes	Individual	Fixed	0.5	0.2557	0.5984	0.3481
5	SBERT_WN_Ind_Pct70	SBERT	Yes	Individual	Percentile	70%	0.2426	0.5615	0.3296
26	TFIDF_WN_Ind_Fix09	TFIDF	Yes	Individual	Fixed	0.9	0.0402	0.0601	0.0478
27	SBERT_NoWN_Wgt2_Fix06	SBERT	No	Weighted(v2)	Fixed	0.6	0.0738	0.0273	0.0388
28	TFIDF_WN_Wgt1_Fix05	TFIDF	Yes	Weighted(v1)	Fixed	0.5	0.0361	0.0328	0.0310
29	TFIDF_WN_Wgt2_Fix05	TFIDF	WN	Weighted(v2)	Fixed	0.5	0.0194	0.0273	0.0223
30	TFIDF_WN_Wgt2_Fix06	TFIDF	WN	Weighted(v2)	Fixed	0.6	0.0066	0.0109	0.0082

The highest performance (F1 score) across all evaluated settings was achieved by the SBERT vectorization method combined with WordNet expansion and the individual thresholding method with a fixed threshold of 0.7 (SBERT_WN_Ind_Fix07), resulting in an F1 score of 0.4072. This setting had a high Recall of 0.6844 but a moderate Precision of 0.3033.

As illustrated in Figure 3, SBERT-based methods (best F1 ≈ 0.41) outperformed TF-IDF-based methods (best F1 ≈ 0.30).

The effect of WordNet expansion depended on the vectorization method used. While WordNet expansion significantly improved performance in TF-IDF settings, its benefit was limited in SBERT settings. Nevertheless, the highest-performing setting utilized WordNet expansion, indicating its usefulness when combined with SBERT.

In comparing mapping methods, the highest performance was obtained using the individual thresholding method (SBERT_WN_Ind_Fix07, F1=0.4072). However, the weighted integration method (SBERT_WN_Wgt1_Fix05, F1=0.3805) also showed competitive performance.

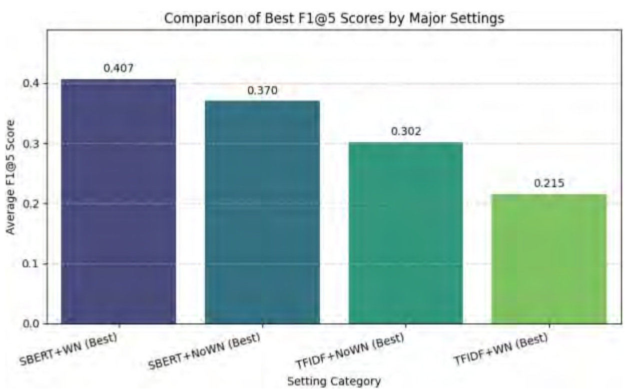

Fig. 3. Performance Comparison by Method

B. Impact of Threshold Settings

Threshold settings significantly influenced prediction performance. Figure 4 shows Precision, Recall, and F1 scores for different fixed thresholds in individual thresholding using SBERT with WordNet expansion.

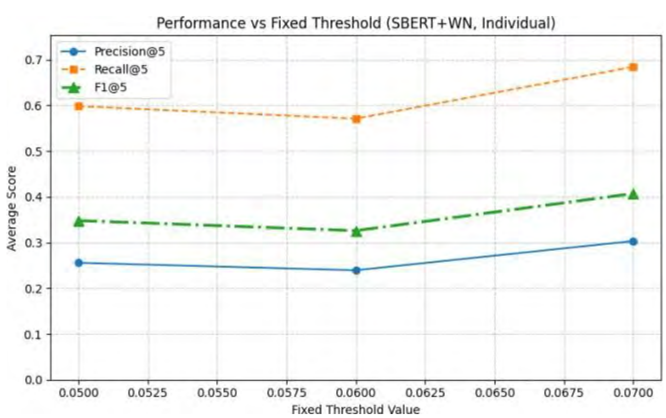

Fig. 4. Impact of Threshold Settings on F1 Score, Precision, and Recall

The F1 score peaked around a threshold of 0.06, indicating that optimal threshold selection is crucial for balancing Precision and Recall maximizing overall performance.

C. Distribution of Performance

Performance varied significantly among individual diagnosis instances, as shown in Table 1. Figure 5 illustrates the distribution of Precision and Recall scores across all experimental settings in the Precision-Recall (PR) space. Most settings were concentrated in the moderate to lower-left region of the PR space, indicating room for improvement. SBERT-based settings generally occupied a higher-performance region compared to TF-IDF-based settings.

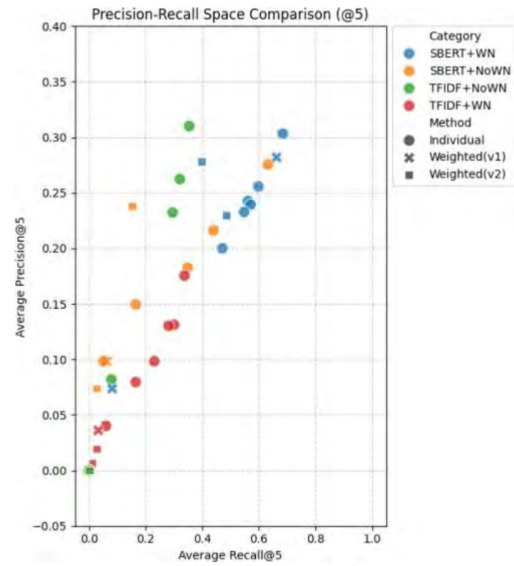

Fig. 5. Precision-Recall Distribution of Experimental Configurations

V. Discussion

This study proposed a data-driven approach to map the relationship between NANDA-I nursing diagnoses and Gordon's 11 Functional Health Patterns (FHP) by employing text similarity techniques and an observation item table, and evaluated its effectiveness. The key result is that the combination of Sentence-BERT (SBERT) for vectorization, WordNet-based semantic expansion, and the individual threshold method (fixed threshold of 0.7) achieved the highest performance among the evaluated settings.

A. Discussion of Main Findings

The experimental results clearly demonstrate that SBERT, which accounts for contextual information, is more effective for the task than the classical term frequency–based TF-IDF (Figure 1). This can be attributed to SBERT's ability to capture subtle nuances and contextual semantic relationships embedded in the definitions of nursing diagnoses, diagnostic indicators, and observation guidelines. This finding aligns with previous research using context-sensitive language models—including ClinicalBERT—that have shown superior performance in clinical note analysis [16] .Our results also support the broader conclusions of recent integrative reviews, such as that of Mitha et al. [9], which identified NLP as a key approach for analyzing nursing notes and supporting clinical judgment. Furthermore, our successful application of a general-domain SBERT model is consistent with the findings of Karacan et al., who also found that general-purpose models perform effectively for mapping nursing terminologies [12].

The impact of WordNet-based semantic expansion varied with the vectorization method. For TF-IDF, WordNet expansion contributed to performance improvement, whereas for SBERT, its effect was limited or, in some settings, slightly reduced performance. This is likely because SBERT, having acquired rich semantic information during its pre-training phase, renders additional lexical expansion redundant or even introduces noise in certain instances. Similar trends have been reported for domain-specific language models such as BioBERT [17] , which is consistent with our findings. Nonetheless, since the highest performance setting in this study employed WordNet expansion, it is possible that, under specific conditions, it complements SBERT's representational capabilities. Further investigation into the selection and integration of expanded vocabulary is needed to fully leverage the benefits of WordNet expansion.

While the individual threshold method achieved the highest performance in this study, the weighted integration method also demonstrated high performance under certain threshold settings (Table 1). This finding suggests that the components of NANDA-I (i.e., definitions, indicators, and factors) may hold varying degrees of importance in relation to FHP and that an integrated evaluation of these components is effective. There remains room for further optimization of the weighting parameters in the weighted integration method. Moreover, the threshold setting was found to have a profound impact on performance (Figure 2).

In this study, the final mapping was achieved by adopting a fixed threshold of 0.7 within the individual threshold method. This threshold is applied to normalized similarity scores. To avoid instance-specific scale effects, raw cosine similarities (sraw) were first min-max normalized to a 0-1 scale for each test case according to Eq. 1:

$$s_{norm} = \frac{s_{raw} - sim_{min}}{sim_{max} - sim_{min}} \quad (1)$$

On the other hand, the threshold–performance graph in Figure 2, which is plotted using raw similarity scores, shows the F1 score peaking at approximately $s_{raw} \approx 0.06$. This raw value corresponds to a normalized score of approximately $s_{norm} \approx 0.7$ for a typical data distribution in our dataset, justifying our use of 0.7 as the fixed cut-off. Notably, the fact that the threshold maximizing the F1 score differs from the threshold maximizing Recall suggests the need to adjust the threshold according to whether the prediction system prioritizes accuracy or comprehensiveness—a critical consideration for practical applications. Although a choice between fixed thresholds and percentile-based thresholds might be made according to the distribution characteristics of the data, no clear superiority could be determined within the scope of this study.

B. Significance and Limitations of Performance Scores

The highest F1 score achieved in this study (approximately 0.41) may appear limited in absolute terms. However, this is likely due to the inherent difficulty of the mapping task between NANDA-I and FHP, differences in their conceptual levels of abstraction, and the variability in their application in clinical settings, compounded by the inherent ambiguity of natural language descriptions. Limitations of evaluation metrics are widely discussed in the field of clinical natural language processing [18], necessitating consideration of natural language complexity and variability in expert judgments. Moreover, the evaluation data used in this study was based on a single assessor's judgment, which may introduce some variability. Consequently, while our evaluation results indicate the initial utility of the proposed approach as an assessment support tool for nurses, they also suggest that further improvements are required.

Nevertheless, the significance of this study lies in demonstrating the feasibility of a data-driven approach for mapping NANDA-I and FHP—a task that has traditionally relied on the tacit knowledge of nurses. By quantitatively comparing and analyzing various technical components—including vectorization, semantic expansion, mapping methods, and threshold settings—this study provides a foundation for future system development. In particular, the superior performance of SBERT over the TF-IDF baseline may serve as a guideline for technology selection in subsequent implementations. Furthermore, even the limited predictive capability demonstrated here—for instance, by suggesting relevant FHP perspectives for nurse assessments or proposing likely diagnostic candidates—indicates the potential of the approach to serve as a decision support tool. Specifically, this system could function as a "cognitive safety net" to mitigate the high cognitive load inherent in the diagnostic process. By presenting a comprehensive list of potential diagnoses, it can help prevent omissions, especially for novice nurses, and allow practitioners to focus on the final clinical judgment. For nursing education, it could serve as a valuable tool for students to explore the complex relationships between patient observations and standardized diagnostic frameworks. However, significant improvements in accuracy are essential before this method can be used as a direct diagnostic tool.

C. Limitations of This Study

This study has several limitations. First, the evaluation dataset comprised only 56 instances, which raises concerns about the generalizability of the results. In addition, the selection of 54 NANDA-I diagnoses and 268 observation items may have influenced the outcomes. Second, the evaluation was confined to the Top-5 predictions, and metrics that assess the overall ranking were not employed. Third, the range of language models, mapping algorithms, and parameter settings examined was limited, and alternative methods might yield different results. Fourth, no qualitative evaluation was conducted; hence, the clinical validity and practical utility of the predictions remain unverified.

D. Future Directions

Based on the results and limitations of this study, several future directions are suggested. First, evaluations using larger and more diverse datasets, along with a refined ground truth established by multiple experts, are desirable. Next, the use of language models specialized for the nursing domain and the introduction of alternative mapping approaches—such as graph neural networks—are expected to further enhance performance. Incorporating clinical contextual information (e.g., patient background and detailed observation narratives) may also prove beneficial. Furthermore, exploring recent advancements like generative AI for suggesting diagnoses, as demonstrated by Lee et al. [11], could offer an alternative, albeit different, avenue for future research.

Finally, qualitative evaluation by nursing experts is necessary to verify the clinical validity and practicality of the proposed approach, ultimately leading to the development of a prototype nursing assessment support tool and a subsequent usability evaluation.

VI. Conclusion

This study proposed a data-driven approach to map NANDA-I nursing diagnoses with Gordon's 11 Functional Health Patterns (FHP) using text similarity techniques and an observation item table. The combination of Sentence-BERT (SBERT) for vectorization, WordNet for semantic expansion, and the individual threshold method (fixed at 0.7) achieved the highest performance among the evaluated configurations, with an average F1 score of 0.4072.

SBERT significantly outperformed TF-IDF, demonstrating its advantage in capturing nuanced semantic relationships within nursing diagnoses. While WordNet improved TF-IDF-based methods, its impact on SBERT was limited, likely due to redundancy or noise. The individual threshold method outperformed the weighted integration method, emphasizing the importance of selecting appropriate threshold values according to the system's objectives.

Despite the promising performance, the inherent difficulty of mapping NANDA-I and FHP remains, due to conceptual differences and linguistic variability. Nonetheless, this study establishes a foundation for data-driven nursing diagnosis support, highlighting the importance of optimal vectorization and threshold selection.

References

[1] K. B. Leary and M. Lee. Patient Safety Amid Nursing Workforce Challenges. Apr.2024.

[2] M. Lunney, “Critical need to address accuracy of nurses’ diagnoses,” *Online J. Issues Nurs.*, vol. 13, no. 1, Jan. 2008.

[3] W. P. Sudarmaji, S. Sholihin, R. A. Permana, A. Soares, and Y. A. Nugraha, “A Clinical Decision Support System as a tool to improve the accuracy of nursing diagnoses,” *J. NERS*, vol. 14, no. 3, pp. 388–392, Jan. 2020.

[4] M. Kaizu and Y. Mranaka, “Problems in Utillizing Electronic Nursing Documentation Related to the Status of Data Input Recording of Post-assessment Nursing Diagnosis and Nursing Care Plan,” Med.Nus. Res. , vol. 13, no. 2, pp. 42–51, 2017.

[5] J. Sweller, “Cognitive load during problem solving: Effects on learning,” *Cogn. Sci.*, vol. 12, no. 2, pp. 257–285, Apr. 1988.

[6] T. Kubo, V. Sornlertlamvanich, and T. Charoenporn, “Assessment pattern mapping in NANDA-I nursing diagnoses framework by BERT approach,” *ECTI Trans. Comput. Inf. Technol.*, vol. 19, no. 1, pp. 65–74, Nov. 2024.

[7] Herdman, T. Heather, S. Kamitsuru, and C. T. Lopes, *Nanda International Nursing Diagnoses 12th ed*. Thieme Medical Publishers, 2021.

[8] M. Gordon, Manual of Nursing Diagnosis, 13th ed. Burlington, MA, USA: Jones & Bartlett Learning, 2014.

[9] Shazia Mitha, Jessica Schwartz, Mollie Hobensack, Kenrick Cato, Kyungmi, Arlene Smaldone, Maxim Topaz, “Natural Language Processing of Nursing Notes: An Integrative Review,” *Comput. Inform. Nurs., Nursing*, 41(6),pp. 1–8.2023.

[10] X. Yang, X. He, H. Zhang, Y. Ma, J. Bian, and Y. Wu, “Measurement of semantic textual similarity in clinical texts: Comparison of transformer-based models,” *JMIR Med. Inform.*, vol. 8, no. 11, p. e19735, Nov. 2020.

[11] Y. Mao and K. W. Fung, “Use of word and graph embedding to measure semantic relatedness between Unified Medical Language System concepts,” *J. Am. Med. Inform. Assoc.*, vol. 27, no. 10, pp. 1538–1546, Oct. 2020.

[12] B. Karacan et al., “Probing the EHR for standardized nursing data,” in Proc. 8th ACM/IEEE Int. Conf. Connected Health: Appl., Syst. Eng. Technol. (CHASE), Orlando, FL, USA, 2023, pp. 209–210.

[13] H. Ju *et al.*, “Generative AI-based nursing diagnosis and documentation recommendation using virtual patient electronic nursing record data,” *Healthc. Inform. Res.*, vol. 31, no. 2, pp. 156–165, Apr. 2025.

[14] F. W. Mutinda, S. Yada, S. Wakamiya, and E. Aramaki, “Semantic textual similarity in Japanese clinical domain texts using BERT,” *Methods Inf. Med.*, vol. 60, no. S 01, pp. e56–e64, Jun. 2021.

[15] A. L. Beam *et al.*, “Clinical concept embeddings learned from massive sources of multimodal medical data,” in *Biocomputing 2020*, WORLD SCIENTIFIC, 2019, pp. 295–306.

[16] K. Huang, J. Altosaar, and R. Ranganath, “ClinicalBERT: Modeling clinical notes and predicting hospital readmission,” *arXiv :1904.05342*, Apr.2019.

[17] J. Lee *et al.*, “BioBERT: a pre-trained biomedical language representation model for biomedical text mining,” *arXiv : 1901.08746* , Jan.2019.

[18] K. Kreimeyer *et al.*, “Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review,” *J Biomed Inform*, vol.73. pp. 14–29, 2017. https://doi.org/10.1016/j.jbi.2017.07.012