Leveraging Natural Language Processing for mHealth Development: A Component-Based Approach Using Nursing Taxonomies

William Niemiec
PPGC - Informatics Institute
Federal University of Rio Grande do Sul
Porto Alegre, Brazil
william.niemiec@inf.ufrgs.br

Anderson Rocha Tavares
PPGC - Informatics Institute
Federal University of Rio Grande do Sul
Porto Alegre, Brazil
artavares@inf.ufrgs.br

Érika Fernandes Cota
PPGC - Informatics Institute
Federal University of Rio Grande do Sul
Porto Alegre, Brazil
erika@inf.ufrgs.br

***Abstract*—This paper proposes the definition of core components of the healthcare domain that can be used as basic building blocks for the development of mHealth applications. Using Natural Language Processing techniques, we systematically analyze all health interventions defined in the Nursing Interventions Classification taxonomy to define a comprehensive set of basic components. Our results show that it is feasible to define a finite set of core components from the nursing domain language that can be further composed into health care plans thus establishing a foundation for developing mHealth solutions with reduced technical effort.**

***Index Terms*—mHealth, care plan, component-based software engineering, no-code, reusable components.**

I. Introduction

Mobile health (mHealth) applications have emerged as powerful tools for improving patient engagement, remote monitoring, and chronic disease management [1]. These applications offer numerous advantages, including enhanced communication between patients and healthcare professionals (HCPs) and improved access to care. Developing an effective mHealth app is both complex and costly, demanding a solid technical foundation in both medical and programming domains. Close collaboration between HCPs and developers is crucial to ensuring that these applications comply with clinical standards, regulatory requirements, and user needs. Moreover, this interdisciplinary partnership must be sustained over time to support the continuous maintenance necessary for delivering a reliable solution [2].

No-code (NC) platforms apply Component-Based Software Engineering (CBSE) principles to allow users without programming expertise to develop applications using graphical interfaces and predefined functional modules instead of writing code [3]. These platforms have become increasingly popular across the education, finance, and administration sectors by making software development more accessible and minimizing dependence on technical skills.

In theory, HCPs could leverage NC platforms to develop mHealth apps tailored to specific treatments, reducing dependence on IT teams and lowering costs [4]. In practice, however, NC platforms are constrained by basic functionalities and often fail to meet the specialized demands and more complex execution flows of healthcare applications. Moreover, a significant gap exists between the language and reasoning used by HCPs and the technical components they must navigate within an NC environment [5].

A possible solution to bridge this gap is to represent the reasoning and language of health care plans as high-level components that can be later translated into computational elements. By focusing on healthcare-specific components, one can reduce the gap between clinical and programming domains and assist in developing customizable digital solutions without extensive technical expertise. To this end, one needs elements that HCPs use in their daily practice as the basic construction components for the logic of the digital solution. Furthermore, one must define the relationship between those basic components and the computational elements executing that logic.

Previous work [6] proposed the use of nursing taxonomy as the basis for the definition of health component models that can be combined into a care plan and used, for example, on NC platforms for the healthcare domain. Five Nursing Basic Component (NBC) models were defined to represent basic actions that can be combined to define a care plan. For each NBC, the structure, behavior, and interface were specified, enabling them to be developed, deployed, and reused.

One important limitation of that work is the fact that only a subset of the nursing taxonomy is considered for the definition of the core components. Thus, the resulting NBCs, although valid, may not be enough to subsume all possible healthcare interventions and care plans.

In this work, we leverage current Natural Language Processing (NLP) techniques (e.g., topic modeling and in-context learning) to systematically analyze all interventions currently described in the Nursing Interventions Classification (NIC) and define a comprehensive set of NBCs. The proposed approach generates an NBC for each NIC through in-context learning [7], and merges similar ones through topic modeling techniques [8]. These NBCs can be further used to build more precise and concise healthcare solutions. We evaluate our approach using a combination of qualitative analysis and computational metrics, including the coherence, pairwise Jaccard distance, and proportion of unique words to measure topic quality. Our results show that a set of NBCs can be systematically extracted from the nursing taxonomy to represent precise and complete care plans. This improvement represents one additional step towards the implementation of a practical NC platform for the development of mHealth applications.

The paper is organized as follows: Section II reviews the concepts of nursing taxonomy and component models used in this work. Section III presents the proposed approach while Section IV details the implementation and results. Section V discusses related works and Section VI resumes the key findings, limitations, and potential implications of our approach. Section VII concludes the paper.

II. Background

This section presents the concepts of the nursing taxonomy (which represents the reasoning we want to model in the proposed CBSE approach) and the concept of the component model (which will be used during the modeling phase).

A. Nursing taxonomy

In nursing practice, care plans and interventions are created based on a systematic nursing process that includes assessment, diagnosis, planning, implementation, and evaluation. The care plan ensures continuity of care and provides a framework for coordinating actions among healthcare providers. Interventions are specific actions or strategies implemented to address a patient's needs and achieve desired health outcomes. Together, care plans and interventions ensure that patient care is both systematic and individualized.

NIC standardizes possible nursing interventions. They can be independent or collaborative, direct or indirect, and individual or group oriented [9]. NIC is composed of seven domains, 30 classes, more than 500 interventions, and 13,000 actions or activities (hereafter, called activities). In this work, we refer to activities defined in the NIC reference book [9]. Finally, activities can be classified into two groups: nurse-dependent and independent. The first includes activities that require nurse participation (for example, catheter replacement). The second group contains activities that patients can do on their own (for example, drinking water every 2 hours). In this work, we focus on the second group.

B. Component model

A component model defines the structure, composition, and interaction of software components within a system. It establishes the rules governing component reusability, intercommunication, and execution to ensure modularity and scalability [10]. On the other hand, components are an instantiation of component models. For example, a "Medication" component model could define a name, a dosage, and the time the patient should take a medicine. Assume that the user has to take two pills of a medicine called "Aspirex" at 8 AM. The instantiated component would have the "Aspirex" name, two pills as the dosage, and the time set to 8 AM.

The creation of new component models requires the definition of key characteristics, including the component's purpose, functionality, dependencies, and interfaces. By clearly establishing these elements, the model dictates how components interact and integrate, contributing to efficient system design and development. This structured approach facilitates the creation of reusable and interoperable components within a system. In this paper, we focus on developing healthcare component models based on NIC.

III. Proposed approach

We assume that mHealth applications are mostly useful to help patients autonomously follow the instructions and interventions proposed by the HCPs. In this context, the NIC taxonomy is the domain language used by HCPs to define the activities associated with each mapped intervention from the care plan. Thus, our ultimate goal is to model activities of the NIC taxonomy as NBCs that can be further used to generate computational elements to build specialized digital solutions, such as mHealth applications.

As explained in Section II, there are more than 13,000 activities from NIC, so a manual generation of NBCs is not feasible. *In this work, we want to investigate the hypothesis that an NLP-based approach will allow us to extract a more comprehensive set of health components from the NIC taxonomy*, thus providing a richer and more precise interface for HCPs to define digital care plans. To do this, we first need to explore the complete NIC taxonomy and extract the activities of each NIC. Then, we need to define a component model for each semantic group defined in the previous step. Figure 1 depicts the proposed approach that is detailed next.

A. Defining syntactic groups from NIC Taxonomy

We start by extracting all health interventions mentioned in the NIC's reference book [9] and filtering nurse-independent activities. We perform this filter because our goal is to help discharged patients. Then, we ask an LLM to analyze each nurse-independent activity associated with an intervention and to generate an NBC. To do this, we define a suitable prompt that provides some examples of NBCs generated from activities of NICs (in-context learning [7]). After all activities from interventions are analyzed, high similarity between the NBCs may exist. We use an NLP technique, called topic modeling [8], to identify these similar NBCs. It consists of automatically grouping words that frequently appear together, helping to identify patterns and structures in large text datasets.

We use two preprocessing strategies to remove possible noise from topic modeling. First, we remove stop words, special characters, and irrelevant words. Then, we merge similar words before the application of the topic modeling technique. Finally, we run a topic modeling technique (see Section IV-B) to generate a set of topics grouping NBCs by syntactic similarities, which we call "syntactic groups".

B. Defining NBCs from syntactic groups

After grouping the nurse-independent activities into similar syntactic groups, we have to convert them back into the final

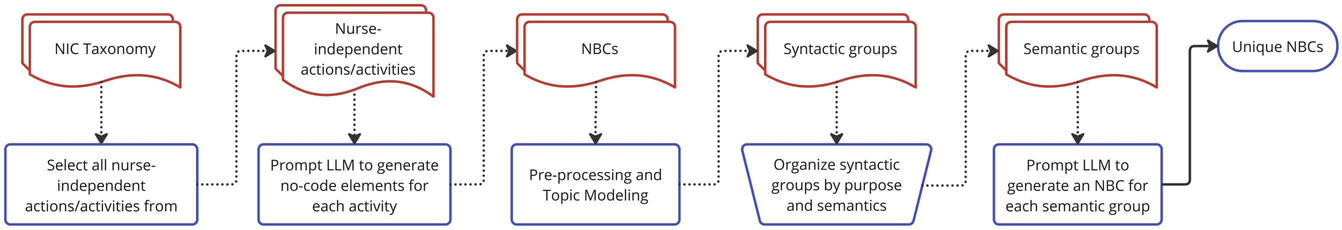

Fig. 1. Methodology overview

NBCs. Thus, we organize the syntactic groups previously obtained by manually analyzing their semantics and purpose. For each syntactic group, if there already exists a semantic group that can represent it, the element is mapped to the existing group. Otherwise, a new semantic group is created following the structure shown in Table I, representing the new semantics. Lastly, for each semantic group created, we used the LLM model to generate an NBC, providing a very similar prompt of Figure 2. Each semantic group is composed of a list of the 8 most representative words of the topics of its syntactic group members. At the end, we have the final NBC that represents the group.

TABLE I
NBC INTERFACE (ADAPTED FROM [6])

Name	Description
Name	Component identification
Description	Summarize what the component does.
Type	Indicates if it is a periodic or non-periodic behavior.
Input	List of parameters that the component receives.
Output	Information obtained at the end of the component execution.
Preconditions	Set of necessary conditions for component execution
Postconditions	Set of what should hold true after the execution of a component
Dependencies	Describe the relationships between components and external factors a component relies on

IV. Implementation and Results

In this section, we describe how we applied the proposed methodology. For better understanding, we have divided it into three subsections: IV-A the extraction of NBCs from the nursing interventions book [9], IV-B the generation of syntactic groups, and IV-C the generation of the semantic groups and the final set of NBCs.

For our experiments, we selected Meta Llama 3 [1] with 70 billion parameters and BERTopic [8] for NBC definition and NBC grouping, respectively. Llama 3 is open source, trained for general purposes, and has been largely evaluated in the last two years.

A. Extracting NBCs from nursing interventions

We extracted every activity from the NICs reference book [9] and stored them in a JSON file. After that, we used Meta Llama 3 to generate NBCs for every activity that patients can execute without the help of a nurse (nurse-independent activity).

We use in-context learning and provide a prompt with six examples along with the generated NBC as part of the prompt. The mapping examples were built by selecting NICs with a previously known mapping, for which a basic component had been manually created. Figure 2 shows the final prompt defined after a few experiments. Lines 1 and 2 introduce the LLM to the task, while lines 4 and 5 describe the NBC's structure. Lines 7 and 8 provide the mapping examples, and lines 10 and 11 specify the output format. Finally, lines 13 and 14 provide the input, i.e., the activity of a NIC. We note that the term "no-code element" is used in the prompt because it generates better results than the word "NBC".

```
1  You are an AI model expert in healthcare informatics
       and software engineering. Your task is to identify
        what no-code elements are necessary for
       implementing actions/activities from nursing
       interventions (NICs) into mHealth applications for
        patients. These elements are nodes that can be
       connected, representing a semantic flow.
2  Note that some interventions may not have a no-code
       element associated, as they may require human
       judgment or observation that cannot be replicated
       by a no-code element.
3
4  ### No-code element structure
5  {{ Table I }}
6
7  ### Examples
8  {{ examples }}
9
10 ### Output format
11 {{ Table I }}
12
13 ## Action/activity
14 {{ NIC action/activity }}
```

Fig. 2. Prompt developed in this work. Note that texts between curly braces ("{" and "}") are inputs embedded in the prompt

From this experiment, 13.067 NBCs were generated, but many of them present similar elements, as expected. Thus, we proceed to the generation of syntactic groups from the initial set of NBCs as detailed below.

B. Generating syntactic groups

To enhance the analysis, it was necessary to merge related NBCs into broader, semantically meaningful groups. Given the large number of NBCs generated, a manual approach would

[1] https://github.com/meta-llama/llama3

have been impractical and subjective. Therefore, we have chosen topic modeling techniques to identify similar content across the syntactic groups. From these similar contents, we can generate NBCs with high coherence.

Topic modeling is an unsupervised machine learning method that discovers underlying themes in textual data by grouping documents based on shared patterns of word usage. In this study, we used BERTopic [8] and HDBSCAN [11] with cosine distance for performing the experiment. The first one was selected because it is a state-of-the-art topic modeling framework that leverages modern language models to provide meaningful clusters and topics, while HDBSCAN is the standard clustering algorithm used by BERTopic. We also performed two preprocessing steps to remove noise as described in Table II. Through a heatmap (Figure 3), it is visible that there are several components with similarity higher than 60%, and the algorithm identified 179 topics out of the 13.067 provided NBCs.

TABLE II
PREPROCESSING STEPS

Step	Preprocessing
1	Remove stop words
2	Remove special characters, such as '#' and '@'
3	Remove the irrelevant words: "no-code", "code", "element", "name", "parameter", "parameters", "specification", "specifications", "specifications", "coding", "boolean", "none", "null"
4	Replace the following similar words with "multimedia": "media ", "http", "https", "videos", "video", "url", "image", "images", "html", "site", "sites", "link", "links"
5	Replace the following similar words with "registry": "log", "logger", "logging"
6	Replace the following similar words with "boolean": "no", "yes", "not"

We selected three metrics to evaluate our model: coherence, pairwise Jaccard distance, and proportion of unique words. The first metric measures how semantically consistent or meaningful the words in a topic are, where higher coherence indicates that the words within a topic tend to co-occur together in similar contexts. The second one measures the dissimilarity between two topics by comparing the similarity of words. It ranges from 0 (identical topics) to 1 (completely disjoint topics). Lastly, the proportion of unique words evaluates the uniqueness of words within a topic compared to other topics, where a higher proportion indicates that the topic contains more words that do not appear in other topics, signifying distinctiveness.

The evaluation results are presented in Table III, showing that the topics have high cohesion and are very different from each other. This can be visualized in Figure 4, where each topic is composed of similar words and one topic is very different from the others. In the context of this paper, each topic is a grouped NBC.

C. Generating semantic groups and NBCs

Besides the syntactic similarity, semantic analysis can also be applied to further reduce the number of basic components without compromising expressiveness. This is done using the LLM model used in Section IV-A. From this step, we reached 67 NBCs and classified them by function, as depicted in Figure IV. Finally, we defined each NBC using the interface of Table I. The complete set of health basic components obtained is available in a public repository. [2]

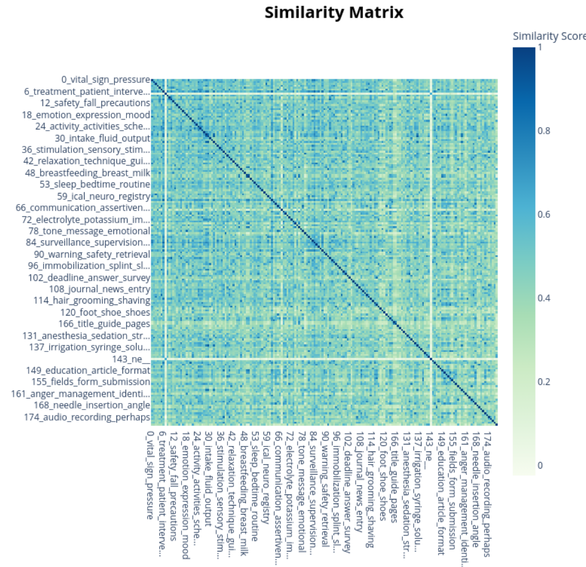

Fig. 3. Similarity heatmap of the NBCs. Note that dark blue means high similarity.

TABLE III
EVALUATION OF THE MODEL

Metric	Value
Coherence	0.6496
Proportion of Unique Words	0.6722
Pairwise Jaccard Distance	0.9957

V. RELATED WORK

Research in mHealth has experienced significant growth in recent years, driven by advancements in mobile technology, artificial intelligence, and the increasing demand for remote healthcare solutions. Rivera-Romero et al. (2023) [12] explore the current state of personalization in mHealth solutions, particularly focusing on the methods and strategies used to tailor mobile health interventions. Their study highlights the increasing interest in personalized mHealth since 2020 and identifies key areas where personalization is applied. They emphasize the use of behavioral change theories and motivational strategies, including gamification and personalized messaging, to enhance user engagement and adherence.

[2]https://github.com/williamniemiec/williamniemiec/tree/main/publications/2025/Leveraging%20Natural%20Language%20Processing%20for%20mHealth%20Development%20-%20A%20Component-Based%20Approach%20Using%20Nursing%20Taxonomies

Fig. 4. Top 10 topics with the top 8 most frequent words

TABLE IV
TOTAL ITEMS IN EACH NBCS CATEGORY

Category	**Total NBCs**
Clinical Care and Procedures	11
Medication and Treatment Management	9
Mental Health and Well-being	6
Vital Signs Monitoring and Control	5
Planning and Routine Management	5
Social and Behavioral Well-being	5
Entertainment and Motivation	4
Hygiene and Physical Health	4
Communication and Shared Care	4
Nutrition and Eating Habits	4
Education and Health Guides	3
Physical Activities and Mobility	3
Symptoms and Diagnosis	2
Rest and Sleep	2

Giunti et al. (2018) [13] present the design and evaluation of ”More Stamina,” a gamified mHealth solution aimed at helping individuals with multiple sclerosis manage their fatigue. The study emphasizes the importance of user-centered design in developing effective health interventions. The researchers incorporated behavioral change models and gamification elements through iterative prototyping and heuristic evaluation to create an engaging and personalized task management tool.

Our work aims to further bridge the gap between the healthcare and technology domains by using the principles of component-based software engineering [10]. This approach improves software maintainability, scalability, and overall development efficiency by breaking down functionality into modular units. We investigate the definition of core component models in the language of the HCP so that they can be leveraged to address the complexity of implementing mHealth solutions.

Another foundation of our approach is the use of currently available natural language processing techniques. Liu et al. (2023) [14] examine the potential of an LLM model in clinical practice, focusing on areas such as clinical decision support, medical documentation, and intelligent question-answering. The results show that the LLM model can generate accurate differential diagnosis lists and provide valuable insights for clinical decision-making, cancer screening, and optimizing decision-support systems.

Bhukya et al. [15] evaluate the application of NLP and topic modeling techniques to analyze patient reviews from online healthcare platforms. By identifying key themes, the research aims to extract insights that can help improve the quality of care and access to healthcare services. The results show that topic modeling can successfully identify common patient concerns and satisfaction areas across primary care, specialist care, and hospital services.

We use LLMs to perform an exhaustive analysis of a document with a comprehensive set of health interventions described in natural language. The result of this analysis will guide us toward the definition of meaningful high-level components that can be combined into a care plan logic and further translated into computational elements of a corresponding mHealth application.

VI. DISCUSSION

In this chapter, we delve into the key findings, limitations, and potential implications of our proposed approach.

A. *Key Findings*

Our approach leverages CBSE principles to bridge the gap between the health domain and computational models. This enables HCPs to design and customize mHealth applications without the need for extensive informatics expertise. By converting NBCs into reusable software components, we enhance the flexibility and adaptability of mHealth solutions. Also, our approach generated 67 NBCs, a significant improvement over the original study, which identified only 5 NBCs.

B. *Limitations and Threats to validity*

We identified three threats to the validity of our findings, which are discussed next.

- Population validity: We interviewed 14 nurses from different specializations and identified NANDA-NIC-NOC - a standard for care plan creation - and used it (only NIC) in our approach. However, it may be possible that there are other standards or other elements we do not address in this work. This risk is mitigated because we described a broader methodology, defined in Section III, that can be used in other standards beyond NIC.
- Internal validation: Another key threat pertains to the semantic grouping process. As this step is performed manually, it introduces a degree of subjectivity, potentially influencing the categorization of data points. To mitigate this risk, we intend to automate this process in future work.
- Construct validation: Given the reliance on LLM models to identify and define NBCs, there is a risk of misinterpretation or incomplete representation of the interventions, or even new elements can show up depending on the

chosen model. To mitigate this risk, we ensured that the prompts and data preprocessing steps were designed to be model-agnostic, allowing future studies to replicate our methodology with alternative LLMs for comparative analysis.

- Conclusion validity: The choice of clustering algorithm for topic modeling impacts the structure and interpretation of extracted topics. In our study, we employed HDBSCAN, a density-based clustering approach. Future work should explore additional clustering techniques to further validate our findings.

C. Practical Implications

By enabling HCPs to develop and customize mHealth applications, we can improve patient engagement and adherence, leading to better health outcomes. The dependency reduction on informatics teams also means that healthcare providers can respond more quickly to changes in patient needs and treatment protocols. Moreover, the cost savings associated with possible reduced hospitalizations and consultations can be substantial, making this approach economically beneficial for healthcare systems.

Our approach also has the potential to significantly impact healthcare delivery in underserved and rural areas. By enabling HCPs in these regions to develop mHealth applications, we can improve access to quality care for populations that traditionally face barriers to healthcare services.

VII. Conclusion

This work proposes to systematically extract and define nurse-independent activities from the nursing taxonomy aiming at finding a finite set of basic actions that can be later instantiated and combined into a care plan to help discharged patients. These components can, for example, be connected to create a flowchart, representing an object from the nursing domain, which may be parsed to generate computational elements. In that way, these basic components can act as a bridge between the nursing and computing domains, reducing the learning curve for HCPs and helping the development of more precise tools for this domain through CBSE.

In future work, we plan on defining a computational model for representing care plans using the defined NBCs. Also, we intend to conduct workshops with users, including both patients and nurses, to validate the usability and effectiveness of our approach. These efforts will further refine our approach and enhance its applicability in real-world clinical environments.

Acknowledgements

Special thanks to Professor Viviane Pereira Moreira and student Marco Antônio Athayde de Aguiar Vieira, who greatly contributed to the improvement of this work. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This work is also partially supported by Propesq-UFRGS. Also, some experiments in this work used the PCAD infrastructure, http://gppd-hpc.inf.ufrgs.br, at INF/UFRGS.

References

[1] M. S. Marcolino, J. A. Q. Oliveira, M. D'Agostino, A. L. Ribeiro, M. B. M. Alkmim, and D. Novillo-Ortiz, "The impact of mhealth interventions: systematic review of systematic reviews," *JMIR mHealth and uHealth*, vol. 6, no. 1, p. e8873, 2018.

[2] A. Subramaniam, E. Hensley, R. Stojancic, J. Vaughn, and N. Shah, "Careful considerations for mhealth app development: lessons learned from questexplore," *mHealth*, vol. 8, no. 0, 2022. [Online]. Available: https://mhealth.amegroups.org/article/view/96403

[3] T. Beranic, P. Rek, and M. Heričko, "Adoption and usability of low-code/no-code development tools," in *Central European Conference on Information and Intelligent Systems*. Faculty of Organization and Informatics Varazdin, 2020, pp. 97–103.

[4] J. Wong, M. Driver, and P. Vincent, "Low-code development technologies evaluation guide," 2019.

[5] A. Haque, B. Arifuzzaman, S. A. N. Siddik, A. Kalam, T. S. Shahjahan, T. Saleena, M. Alam, M. R. Islam, F. Ahmmed, M. J. Hossain *et al.*, "Semantic web in healthcare: a systematic literature review of application, research gap, and future research avenues," *International Journal of Clinical Practice*, vol. 2022, 2022.

[6] W. Niemiec and E. Cota, "Towards a component-based framework for mhealth apps: Bridging the gap between the nursing domain language and the computation domain," *Journal of Systems and Software*, 2025. [Online]. Available: http://dx.doi.org/10.2139/ssrn.5087361

[7] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," in *Advances in neural information processing systems*, vol. 33, 2020, pp. 1877–1901.

[8] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[9] H. K. Butcher, G. M. Bulechek, J. M. M. Dochterman, and C. M. Wagner, *Nursing interventions classification (NIC)-E-Book*. Elsevier Health Sciences, 2018.

[10] U. K. Tiwari and S. Kumar, *Component-based software engineering: Methods and metrics*. Chapman and Hall/CRC, 2020.

[11] L. McInnes and J. Healy, "Accelerated hierarchical density based clustering," in *2017 IEEE international conference on data mining workshops (ICDMW)*. IEEE, 2017, pp. 33–42.

[12] O. Rivera-Romero, E. Gabarron, J. Ropero, and K. Denecke, "Designing personalised mhealth solutions: An overview," *Journal of Biomedical Informatics*, p. 104500, 2023.

[13] G. Giunti, V. Mylonopoulou, O. R. Romero *et al.*, "More stamina, a gamified mhealth solution for persons with multiple sclerosis: research through design," *JMIR mHealth and uHealth*, vol. 6, no. 3, p. e9437, 2018.

[14] J. Liu, C. Wang, and S. Liu, "Utility of chatgpt in clinical practice," *Journal of medical Internet research*, vol. 25, p. e48568, 2023.

[15] B. Bhukya and M. Sheshikala, "Nlp based topic modeling for healthcare: Analyzing patient reviews to improve quality of care and access to services," in *2023 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*. IEEE, 2023, pp. 7–12.