Towards a Multi-dimensional Health Data Analysis Framework

Fazle Rabbi[1], Bahareh Fatemi[1], Suresh Kumar Mukhiya[2], and Yngve Lamo[3]

[1] Information Science and Media Studies, University of Bergen, Norway
fazle.rabbi@uib.no, bahareh.fatemi@uib.no
[2] Tryg Norge
suresh.mukhiya@tryg.no
[3] Department of Computer science, Electrical engineering and Mathematical sciences, Western Norway University of Applied Sciences, Norway
yngve.lamo@hvl.no

Abstract. Healthcare processes need to be streamlined to offer better healthcare services. Data analysis can be crucial in reducing costs, optimizing processes, and analyzing treatment effectiveness. However, data analysis in healthcare is complex due to the variety and complexity of patient data. This paper proposes a multi-dimensional comparative analysis method that offers healthcare professionals a lens to delve into healthcare datasets from various perspectives. The paper discusses the importance of comparative analysis in healthcare illustrated by two examples on how we can understand the pattern of comorbidity and how we can analyze the effectiveness of internet delivered psychological interventions. The paper presents a multi-dimensional comparative analysis framework covering various use cases in analysing healthcare data. The framework allows healthcare professionals to compare and contrast healthcare data across multiple dimensions, including clinical dimensions such as diagnosis, outcome measures, time dimension, patient dimensions (engagement, involvement), cost dimension, and other relevant factors. This approach offers a more insightful understanding of healthcare data and facilitates informed decision-making in healthcare practices.

Keywords: Healthcare data · Comorbidity · Abstraction· Category theory · Knowledge graph · Context.

1 Introduction

Healthcare expenses are increasing due to various factors such as rising costs of medicines and healthcare equipment, complexity of managing chronic diseases and comorbidity conditions, and growing number of mental illness. Identifying recurring patterns within healthcare processes is crucial for streamlining healthcare procedures and ultimately improving patient outcomes. Data analysis provides an opportunity to reduce the cost in healthcare by optimizing processes and maximizing resource utilization. Data analysis can also be used to analyze the effectiveness of various treatment options. However, data analysis in the healthcare domain is complex due to the variety of patients' complex healthcare-related issues. Healthcare professionals need to analyse data in the healthcare domain, which comes from multiple sources, for example, admission records, diagnosis reports, laboratory test results, and treatment procedures from various departments and clinics. Although there exist several healthcare ontologies, for example, ICD-10 [13] and SNOMED-CT [1], there needs to be a uniform framework for data analysis in the healthcare domain.

Researchers in the healthcare domain also carry out research activities that may require the use of new data formats. Researchers often use randomized controlled trials (RCTs) to assess the effectiveness of a particular intervention, treatment, or medical approach. RCTs serve as a gold standard for assessing treatment interventions. The primary purposes of conducting randomized controlled trials include determining causation, evaluating effectiveness. Analyzing the data from RCTs requires thorough investigation from various perspectives to identify potential variables that influence the intervention and treatment method. To provide a cost-effective solution to treat mental illness, several RCTs have been conducted. Internet-Delivered Psychological Treatment (IDPT) systems have the potential to provide evidence-based mental health treatments for a far reaching population at a lower cost [10,11]. In [9], the author presented a framework to develop an adaptive IDPT system that can adapt psychological interventions according to the users need, context, and preferences.

In this paper, we focus on the analysis of identifying recurring patterns within healthcare processes. This is a crucial need for streamlining healthcare procedures and ultimately improving patient outcomes. We propose to use a multi-dimensional comparative analysis method which offers a lens through which healthcare professionals can delve into the healthcare dataset from various perspective and allows them to explore diverse use-cases that span from evaluating treatment interventions to understanding the patterns of patients with comorbidities. The paper discusses the importance of comparative analysis in understanding the effectiveness of interventions such as dropout rates in RCTs. A multi-dimensional comparative analysis framework is proposed as a means to explore patient patterns and tailor interventions to enhance engagement.

The paper is organized as follows. In Section 2, we present several needs for comparative analysis in the healthcare domain. In Section 3 we present a multi-dimensional comparative analysis framework and show the applicability of the framework in analyzing healthcare dataset. In Section 4 we present closely related work and in Section 5 we conclude the paper.

2 Needs for comparative analysis

In this section, we present the need for comparative analysis in the healthcare domain. Comparative analysis can be a valuable tool for making informed decisions about healthcare. Below is a list of use-cases of the application of comparative analysis in the healthcare domain:

- Comparing the cost-effectiveness of two different treatments for a chronic disease.

- Comparing the long-term outcomes of different surgical procedures.
- Comparing the effectiveness of different public health interventions.
- Discovering the pattern of patients with comorbidity.

In the following subsections we present two specific cases that can benefit from comparative analysis of healthcare data.

2.1 Discovering the pattern of comorbidity patients

Comorbidity introduces a higher risk of complications in many cases. For instance, diabetes and cardiovascular disease are often comorbid [5], and their coexistence can increase the risk of heart-related complications in patients. The treatment of comorbidity is not straightforward due to many reasons, such as delayed diagnosis, medication interactions, and side effects. The combined effects of different conditions can result in greater physical and mental health challenges that can have a cumulative impact on the well-being of the patients and can lead to a reduced quality of life. Therefore, understanding the patterns of comorbidities is essential as it will allow healthcare professionals to identify potential risks and take steps to prevent the worsening of conditions. However there are many challenges to discover the pattern of comorbidity patients and their progression of diseases. We need an efficient technique to extract the necessary and relevant information from healthcare data. Eliminating noise and irrelevant information from healthcare data is essential for data analysis. However, this task is not easy due to the huge amount of data captured in the healthcare system from multiple sources. A comparative analysis method which can filter healthcare data across various dimensions and abstraction levels would be useful for analyzing the pattern of comorbidities, as it would allow healthcare professionals to effectively test out their hypothesis about comorbidity patterns.

2.2 Comparing the effectiveness of psychological interventions

Randomized control trials are often used to measure the effectiveness of different treatment interventions, such as IDPT, for mental illness. In randomized control trials, populations are separated into two groups: (i) the experimental group that receives the intervention that is being tested and (ii) the control group that receives an alternative (conventional) treatment.

The outcome of treatment in these groups is then followed up to determine the effectiveness of the interventions. The multi dimensional analysis could be means for understanding causal effect in IDPT. In current approaches, the focus is on measuring the effect change but too little focus was provided on the causal relationships. While measuring the effectiveness of treatment interventions, it is essential to understand the subjects, determine the treatments' parameters, and address potential challenges such as dropout rates, particularly in IDPT. Various factors can influence dropouts in IDPT systems. Significant causes include a) lack of participation, b) technical issues, c) privacy concerns, d) perceived ineffectiveness of online therapies, e) time constraints, f) lack of support or proper guidance during online therapy, g) unexpected life events, h) limited motivation, i) complexity of the program, and j) poor user experience [11].

Understanding these causes can guide developing and implementing strategies to reduce dropout rates, such as improving user engagement, promptly addressing technical issues, enhancing privacy measures, and providing adequate support throughout treatment. To effectively address and mitigate the identified causes of dropout in IDPT programs, a multi-dimensional analysis framework is essential to adapt the IDPT program based on the user's needs and preferences. A multi-dimensional comparative analysis framework proposed in this paper allows us to discover patient patterns. Using such patterns, IDPT systems can proactively tailor interventions. Equipped with insight into patient behaviors, the IDPT system can dynamically adapt its content, pacing, and interactive elements to enable real-time adjustments to address issues such as lack of engagement, perceived ineffectiveness, or technical challenges.

3 Proposed framework

We present a conceptual framework for the analysis of health data in Figure 1 that supports comparative analysis from a variety of perspectives. Data from various sources are enriched with healthcare ontologies e.g., ICD-10, SNOMED-CT, and stored in a knowledge graph. The framework allows us to search for healthcare data across various dimensions and abstraction levels using graph patterns.

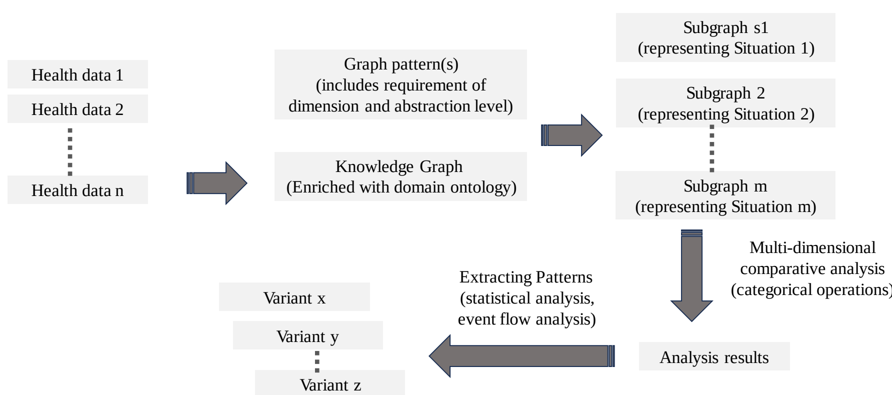


Fig. 1: Conceptual framework for healthcare data analysis

The proposed framework incorporates a multi-dimensional comparative analysis method which is based on categorical operations. The results of this comparative analysis help identify patterns among patients who share similar issues or disease trajectories. In this framework, we emphasize on augmenting healthcare data with domain ontology which is allowing us to do comparison and extract patterns and variants from various perspectives and abstraction levels. While we demonstrate a proof of concept using a graph database in this paper, the framework is also adaptable to SQL databases, employing set and relational operations as an alternative implementation approach.

Figure 2 illustrates an overview of the categorical operations for the comparative analysis. A knowledge graph is represented in the figure as I; Subgraphs O_1^*, O_2^* indicate two objects that are subject to comparison. These subgraphs may include the diagnosis or the symptoms of some patients. We identify the commonality of O_1^*, O_2^* by a pullback operation which is represented as object C in the figure. From O_1^* and C we can compute the object D_1 that represents the nodes and relationships present in O_1^* but not in C. Similarly we can compute D_2. With these objects, we can perform statistical analysis to measure the similarity and dissimilarity of different aspects of the healthcare data.

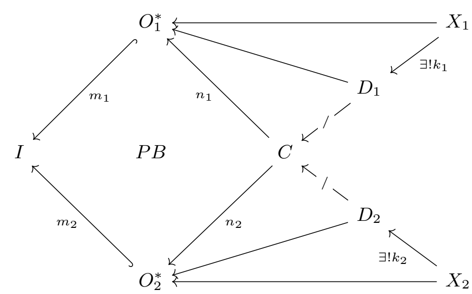


Fig. 2: Pullback object (C) computes the commonality between O_1^* and O_2^*; D_1 and D_2 objects are used to compute the dissimilarities between O_1^* and O_2^*

The idea of using comparative analysis can be further extended to study the progression of diseases among certain population of patient individuals. Figure 3 shows some categorical operations that can be performed to perform computational analysis to find the progression of diseases of some groups of patients. Subgraphs S_1, S_2, S_3.. shows the weekly progression of diseases of a patient group. This approach can also be used to analyze the progression of symptoms or the movement of patients in the hospital in various departments or clinics.

3.1 Comparative analysis for understanding comorbidity

In our previous work [6], we presented a validated learning approach for healthcare process analysis which incorporates two sequential processes. The first step is to identify groups of patients with various comorbidity issues and the second step is to analyze the progression of diseases in those groups of patients. The first step of the process relies on a community detection technique to identify patient groups with comorbidity diseases; and the second step involves human input regarding potential pattern of disease progression. In this step, the user makes different hypothesis about disease progression and the hypothesis is validated by extracting evidences from a healthcare dataset. We presented a variant

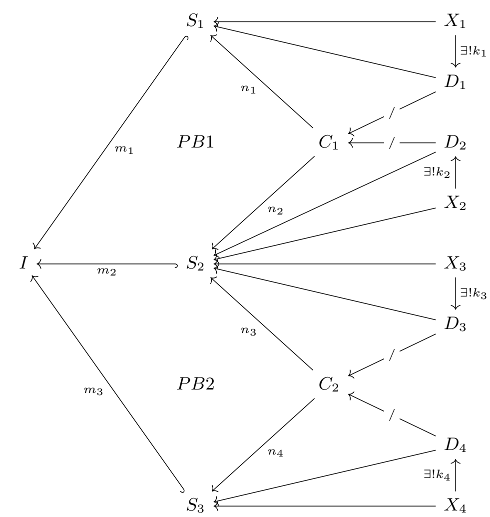


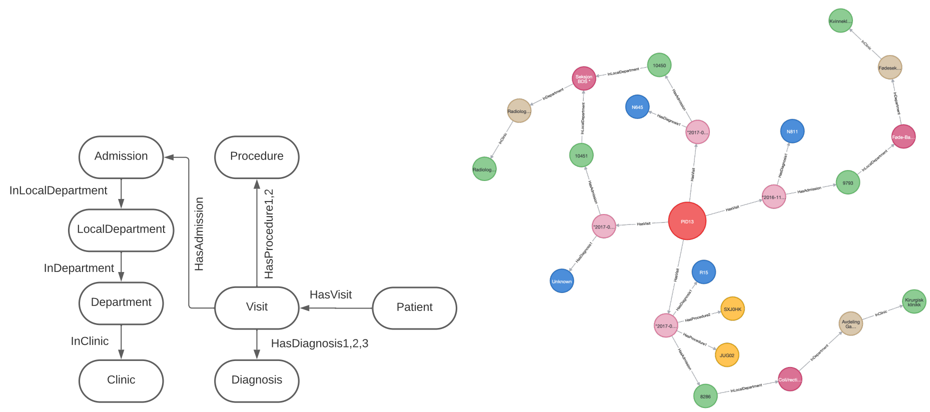
Fig. 3: Capturing the progression of diseases with pullback operation

of Linear Temporal Logic (LTL)[16] language to specify the following pattern of comorbidity issues:

- Diseases that appear one after another.
- Diseases that develop one after the other over short or extended periods of time.
- Diseases that occur after a continuous period of a condition such as high blood pressure.
- Diseases that appear continuously and at determinable time intervals.

While community detection proves to be beneficial in identifying certain groups of patients sharing common diseases, its application has limitations when analyzing comorbidity. This method offers limited insights about the factors that contribute to the comorbidity cases. If there is a group of patients who suffer from disease d_1 and d_2, it would be important to study the dissimilarity of these groups of patients with patients who are diagnosed with one of the diseases, but not both. For example, we may be interested in studying patients who have diabetes or kidney problems or both. A community detection algorithm may find a patient group who have both diseases but we also need to study the patients who have one of the diseases but not both. Hence, a framework for comparative analysis is required. A multi-dimensional comparative analysis framework would be appropriate to deal with the detailed analysis of the progression of diseases for patients with comorbidities.

Figure 4a presents a schema of a knowledge graph that allows us to instantiate a knowledge graph with patients' clinical information such as admission, diagnosis, procedure, etc. The knowledge graph is enriched with the ICD-10 on-



(a) (b)

Fig. 4: Model for patient profile and its relationship with ICD-10 and department ontologies [6] (4a), graph instance for patient with id=PID13 (4b).

tology. Figure 4b shows an instance of this graph schema that represents the information about the administration of a patient in our dataset.

In this section, we elaborate three comparative analysis tasks that can be performed using our categorical approach:

1. **Comorbidity analysis**: Extracting information about patients who have been diagnosed with multiple diseases.
2. **Progression of diseases**: Analysing the progression of diseases of some patient groups.
3. **Comparison of progression of diseases**: Analysing the progression of diseases of two different patient groups.

Comorbidity Analysis

Suppose we wish to study the patients who have been diagnosed with '*diseases of the digestive system (K00-K95)*' and '*diseases of the circulatory system (I00-I99)*'. In Figure 2, subgraph O_1^* and subgraph O_2^* represents patients with '*diseases of the digestive system*' and '*diseases of the circulatory system*' respectively. A pullback operation would give us object C representing patients who have been diagnosed with both diseases. We can extract the patients individuals from the knowledge graph using the Cypher query. Table 1 shows the Cypher query to extract this pullback object. From subgraph O_1^* and the pullback object C we can compute the object D_1 which includes patient individuals who have been diagnosed with '*diseases of the digestive system*' but have not been diagnosed with '*diseases of the circulatory system*'. While this comparative study allows us to identify the number of patients who have such comorbidity issues, we can use similar comparative analysis to check the progression of diseases or symptoms or the movement of patients flow in different departments or clinics.

Table 1: Cypher query for computation of pullback: patients diagnosed with '*diseases of the digestive*' and '*diseases of the circulatory systems*'

Cypher Query
`MATCH (p1:Patient)-[]→(v1:Visit)-[]→(d1:Diagnosis {Level1:'K00-K95'})`
`MATCH (p2:Patient)-[]→(v2:Visit)-[]→(d2:Diagnosis {Level1:'I00-I99'})`
`WHERE p1 = p2 RETURN DISTINCT p1;`

Progression of Diseases

Suppose we are interested to study the progression of disease in patients who have been diagnosed with '*diseases of the digestive system*'. To study this progression, we can perform the categorical operations in Figure 3 to carry out the computational analysis. Here in this situation, the subgraphs S_1, S_2, S_3.. indicate the monthly progression of diseases of a group of patients who have been diagnosed with '*diseases of the digestive system*'. The Cypher query used to extract this progression is shown in Table 2.

Table 2: Cypher query for extracting the progression of diseases in patients with '*diseases of the digestive system.*'

Cypher Query
`MATCH (p1:Patient)-[]→(v1:Visit)-[]→(d1:Diagnosis {Level1:'K00-K95'})`
`WHERE date(v1.visitDate) ≥ date("2015-01-01") AND`
`date(v1.visitDate) ≤ date("2015-01-31")`
`WITH collect(d1) AS` $d1_{collection}$
`MATCH (p2:Patient)-[]→(v2:Visit)-[]→(d2:Diagnosis)`
`WHERE date(v2.visitDate) ≥ date("2015-02-01") AND`
`date(v2.visitDate) ≤ date("2015-02-28") AND p1 = p2 AND`
`NOT d2 IN` $d1_{collection}$ `RETURN DISTINCT d2;`

Comparison of progression of diseases

In Figure 5 we present another comparative analysis using categorical operations where we compare the progression of diseases of two different patient groups e.g., the weekly progression of patients diseases who have been diagnosed with *'diseases of the digestive system'*. The subgraph $\alpha_1, \alpha_2, \alpha_3$ represents the diseases of patients with '*diseases of the circulatory system*' who have also been diagnosed with *'diseases of the digestive system'* and $\beta_1, \beta_2, \beta_3$ represent the diseases of patients who have been diagnosed with *'diseases of the digestive system'*, but NOT '*diseases of the circulatory system*'. This comparative analysis will give us an idea of the variety of diseases that patients with '*diseases of the circulatory system*' are prone to suffer while they are diagnosed with *'diseases of the digestive system'*. The Cypher query employed for comparing the progression of diseases of two patient groups is shown in Table 3. This comparative analysis can be performed at a variety of abstraction levels. Figure 6 illustrates a comparative analysis performed over two different abstraction level. The subgraphs

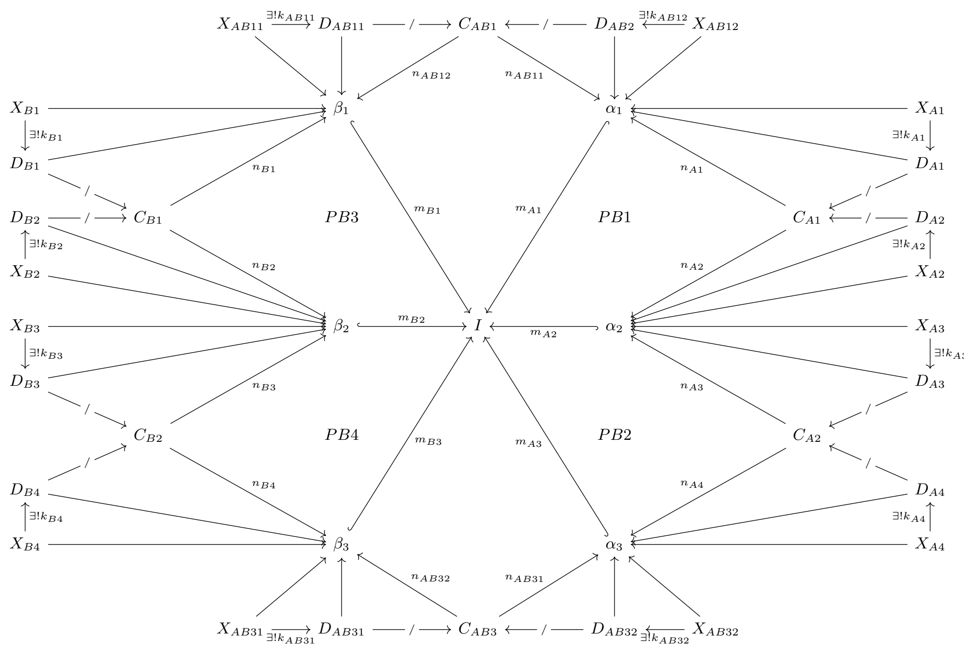


Fig. 5: Comparison of progression of diseases

$\alpha_1, \alpha_1^{'}$ represent a group of patients diagnosis with lower and higher abstraction level (respectively). The progression of diseases in this group of patients is shown in subgraphs $\alpha_2, \alpha_2^{'}$. These patient groups progression of diseases can be compared with another patient groups (i.e., subgraphs $\beta_1, \beta_1^{'}$ and $\beta_2, \beta_2^{'}$) by means of categorical operations as shown in Figure 6.

Table 3: Cypher query for comparative analysis of diseases

Cypher Query
MATCH (p1:Patient)-[]→(v1:Visit)-[]→(d1:Diagnosis {Level1:'K00-K95'})
WHERE date(v1.visitDate) $\geq$ date("2015-01-01") AND
date(v1.visitDate) $\leq$ date("2015-12-31")
WITH collect(d1) AS $d1_{collection}$
MATCH (p2:Patient)-[]→(v2:Visit)-[]→(d2:Diagnosis {Level1:'I00-I99'})
WHERE date(v2.visitDate) $\geq$ date("2015-01-01") AND
date(v2.visitDate) $\leq$ date("2015-12-31") AND NOT d2 IN $d1_{collection}$
RETURN DISTINCT d2;

Figure 7 illustrates a computational model for analyzing two sets of progressions. $\alpha_1, \alpha_2, \alpha_3$.. (resp. $\beta_1, \beta_2, \beta_3$..) represents the progression of situations (e.g., diseases, symptoms) at a certain abstraction level j; $\alpha'_1, \alpha'_2, \alpha'_3$.. (resp. $\beta'_1, \beta'_2, \beta'_3$..) represents the progression of situations specified at a higher level of abstraction k. $C_{\alpha 12}$ is the pullback object of $\alpha_1 \rightarrow I$ and $\alpha_2 \rightarrow I$. Similarly, $C_{\alpha 23}$, $C_{\beta 12}$ and $C_{\beta 23}$ are the other pullback objects computed from the situations represented at

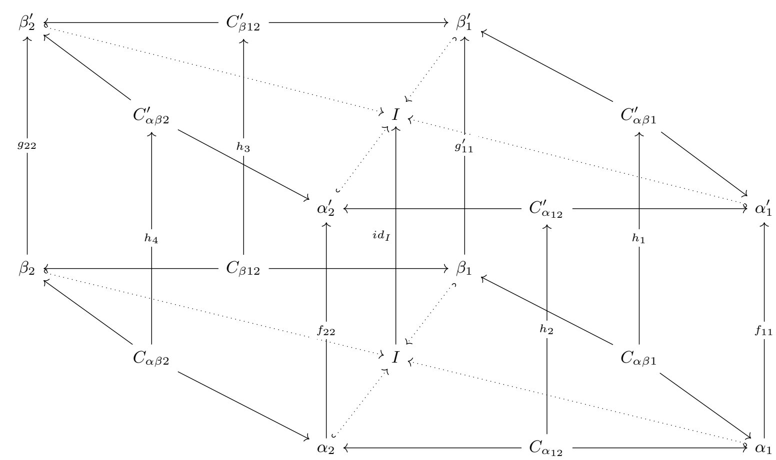


Fig. 6: Comparison of progression of diseases at a higher level of abstraction

level j. The evolution of progressions are represented at a higher level of abstraction at the top of the figure in $\alpha_1, \alpha_2, \alpha_3..$ (resp. $\beta_1, \beta_2, \beta_3..$). Figure 7 shows the co-limit objects CL_α, CL_β at level j which represents combined commonality of the progressions. The pullback of r, t (resp. r', t') is shown in the figure as Z (Z'). The pullback object Z represents the commonality in the progressions in $\alpha_1, \alpha_2, \alpha_3..$ and $\beta_1, \beta_2, \beta_3...$ An empty pullback object Z would indicate that the progressions in $\alpha_1, \alpha_2, \alpha_3..$ and $\beta_1, \beta_2, \beta_3..$ are considered to be different at abstraction level j.

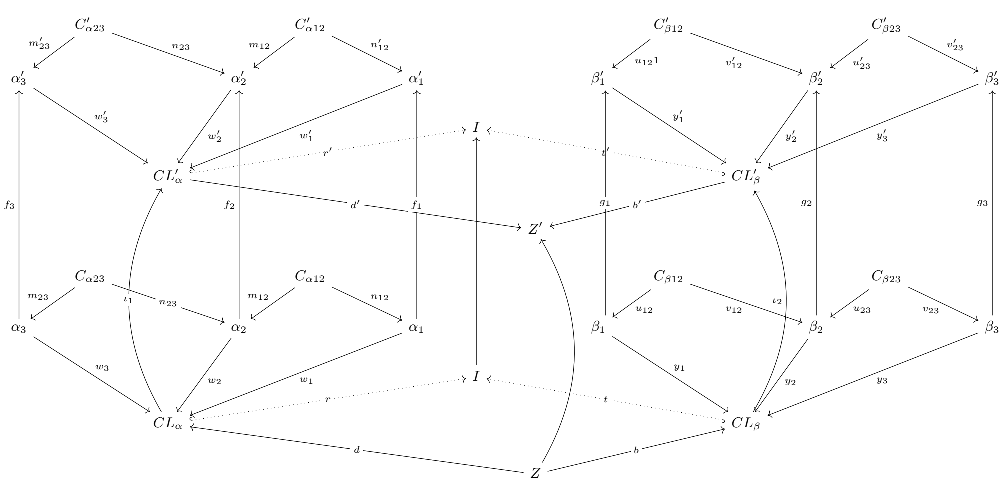


Fig. 7: Computational model for analyzing two sets of progressions.

Variant Identification

The result of the comparative analyses can be used to identify variants. In [6], we proposed a linear temporal logic language LTL_{EOT} to represent a variety of comorbidity patterns for determining patient variants. LTL_{EOT} allows us to formulate queries to find patients who exhibit similar patterns in their healthcare event log consisting of patients admission, diagnosis, procedure related information with timestamp. The syntax of LTL_{EOT} includes the incorporation of the following:

- ontologies so that the base case of the satisfaction relation refers to an instance of ontology concept ; and,
- specific time intervals.

With this formulation, we specify comorbidity patterns, specifying the diagnoses and time intervals between events. A few examples are shown below:

- Find all events of patients who are diagnosed with iron deficiency anemia (D50), right after being diagnosed with chronic kidney disease (N18):

$$\Diamond_{(\geq\ 0\ seconds)}(N18) \wedge \Box_{(\geq\ 0\ seconds)}\Big((N18) \rightarrow \bigcirc_{(\geq\ 0\ seconds)}(D50)\Big) \tag{1}$$

- Find all the events of patients who have been diagnosed with diabetes mellitus (E08-E13) and within 2 years have been diagnosed with chronic kidney disease (N18):

$$\Diamond_{(\geq\ 0\ seconds)}(E08 - E13) \wedge \Box_{(\geq\ 0\ seconds)}\Big((E08 - E13) \rightarrow \Diamond_{(\leq\ 2\ years)}(N18)\Big) \tag{2}$$

The formulas include temporal operators, $\Box$ (always), $\Diamond$ (eventually), $\bigcirc$ (next-time), which enables us to specify properties of events as they evolve over time.

3.2 Comparative analysis for IDPT:

In this section, we present an application of comparative analysis for analyzing the effectiveness of treatment interventions. Patients suffering from mental health morbidities can be assigned one or more therapies (IDPT). IDPT involves one or more modules. Each module has some prerequisites that each patient must fulfil to complete. Each module consists of different tasks. These tasks can be interactive or informative. In Figure 8, we present a knowledge graph schema which allows us to store information about patients involvement in IDPT (adapted from [9]). *Informative tasks* provide learning materials about mental health issues, therapy, symptoms, use cases, and several ways to manage them. The main objective of such educational materials is to provide psychoeducation so that:

- Patients and their associated families can learn about symptoms, causes, remedies and treatment concepts;

– Patients can understand self-help programs and steps required to manage their illness;
– Patients can correlate their situations with similar others, which helps to vent their frustrations.
– Such educational materials are in the form of reading tasks (text), listening (audio), and watching (video).

Interactive tasks differ from informative tasks in that they involve user interactions, often in the form of exercises. Such exercises can be physical activities or computerized tasks. Examples of physical activities include workouts and mindfulness exercises like breathing exercises, walking certain distances, stretching, or performing other activities. Examples of computerized exercises include filling in the blanks, answering questions (Q/A), multiple-choice questions (MCQ), and providing feedback. Feedback tasks involve using free text, rating systems, or multiple-choice questions.

Evaluation is essential in all the parts of IDPT. Each task and module has an evaluation method. Overall evaluation of module tasks gives the gross evaluation of the therapy.

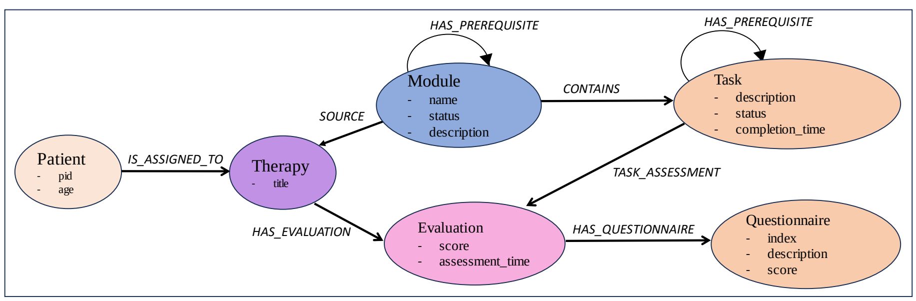


Fig. 8: Knowledge graph schema for Internet Delivered Psychological Treatments

Analyzing drop-out in IDPT: Dropout is a major problem in IDPT programs. There might be a variety of reasons for the dropout and it is essential to understand the cause of dropouts. Our proposed comparative analysis framework would allow researchers to study the dropout cases from a variety of perspectives. For example, in Attention Deficit/Hyperactivity Disorder (ADHD) interventions, we might be interested to study the role of concentration issues against completing modules and tasks. The comparative analysis method of two patients group as shown in Figure 2 can be adopted for analyzing dropout of patients. We may represent subgraphs O_1^*, O_2^* consisting of patients who have concentration issues and dropped out from the intervention program accordingly. The pullback object C would indicate the patients who have concentration issue and dropped out from the intervention program. Similarly we can analyze the dominance of other parameters that influences the drop-out from IDPT programs.

Analyzing the dominance of symptoms: The comparative analysis method of capturing the progression of issues as presented in Figure 3 may be used to analyze the significance of certain symptoms among patients. Suppose we wish to study the effectiveness of modules in a therapeutic program among patients who have high concentration issue. The subgraphs $S_1, S_2, S_3, ..$ in Figure 3 may represent patients who have started module $M_1, M_2, M_3, ..$ (respectively) and the patients have high concentration issue found at the time of starting each module. A categorical pullback of $S_1 \to I$ and $S_2 \to I$ would give us patients who have concentration issue after completing module M_1; The object D_1 represent the patients whose concentration problem have been resolved after the completion of module M_1.

Personalized treatment: In order to develop personalized treatment interventions it is required to identify variants of patient groups. In [10], Mukhiya et. al., presented a rule based approach for adaptive IDPT. Comparative analysis would provide us a mechanism to identify the variants of various patients group who have participated in a treatment program.

The comparative analysis as presented in Figure 5 may be used to analyze the significance of certain mental conditions of patient groups and the effectiveness of a therapeutic program. We may represent subgraphs $\alpha_1, \alpha_2, \alpha_3$ consisting of patients completion of modules over different weeks who have scored more than X in Adult ADHD Self-Report Scale (ASRS); and $\beta_1, \beta_2, \beta_3$ may represent subgraphs consisting of patients completion of modules over different weeks who have scored more than Y in Patient Health Questionnaire (PHQ-9). This comparative analysis would indicate the effectiveness of individual modules in reducing certain mental conditions in a therapeutic program. This knowledge can be useful identify variants of patient population and offer personalized treatments using adaptive treatment method.

4 Related works

In [14], Partington et. al., presented a comparative analysis of clinical processes of four Australian hospitals. They presented the use of process-mining technique to discover the control-flow and the performance of the processes at each hospital and the discovered process models were used for comparison. Through an exploratory approach they have identified four comparative points based on known drivers of costs and/or patient health outcomes: the proportion of patients admitted to an inpatient care setting; the throughput timing between ED presentation and movement to an inpatient setting (Admission); the frequency of procedures (diagnostic/treatment) provided; the total length of stay for patients. Partington et. al., did not present any comparative analysis technique for identifying patterns of comorbidity issues and effectiveness of treatment interventions. In this paper, we have presented the need for comparative analysis across various healthcare scenarios and presented a generic framework for comparative analysis.

Scientists from various fields of research have investigated comorbidity and have used different methodologies to deal with its complexities. To identify the prevalence of comorbidities of mental and behavioral disorders, Cha et al. [4]

propose an analysis based on association rules. Boytcheva et al. [3] propose a cascade data mining approach specifically tailored for frequent pattern mining in comorbidity studies. Several studies fall into the category of network analysis. Jones et al. [8] define four network statistics to identify symptoms that connect two mental disorders. Social network analysis and graph theory have also been used to understand the comorbidity of two chronic diseases [7]. Bottrighi et al. [2] proposed a knowledge-based approach to run-time comorbidity management to support physicians during the execution of the Clinical Practice Guidelines (CPGs) on a specific patient. Piovesan et al. [15] using Computer-Interpretable Guidelines, the history of the status of the patient, and the log of the clinical actions executed on them, propose an Answer Set programming-based method for the treatment of comorbid patients. However, these works do not generalize the need for comparative analysis for analyzing comorbidity patterns and do not promote the use of a multi-dimensional comparative analysis framework as we have proposed in this paper.

The study [12] examined the impact of weekly SMS reminders on adherence to an IDPT for adults with ADHD. The results indicated that the overall module completion, logins, and coping strategy practice slightly improved. Although SMS reminders can influence engagement, they do not uniformly improve overall adherence. To enhance the effectiveness of interventions such as the self-guided IDPT for ADHD, the proposed multi-dimensional framework for analyzing user patterns can be employed, especially in the context of SMS reminders. The framework considers user engagement patterns, allowing for the identification of personalized strategies to improve adherence. By tailoring reminders based on a comprehensive analysis of user patterns, the self-guided IDPT can be adapted to better suit the diverse needs and preferences of adults with ADHD, ultimately fostering more sustained engagement and positive outcomes.

5 Conclusion

In this paper, we explored the significance of comparative analysis in the healthcare domain and proposed a multi-dimensional comparative analysis framework to address two challenges of data analysis in healthcare: (i) Comparing the effectiveness of different public health interventions; and (ii) Discovering the pattern of patients with comorbidity. We presented a formal approach to address the common need for comparative analysis in healthcare. The proposed framework facilitates comparative analysis of healthcare datasets from various perspectives, essential for comprehending the patterns of patients with comorbidities. The paper also highlighted the importance of comparative analysis in understanding the effectiveness of interventions and presented computational methods based on formal method techniques. The presented approach is generic enough to be applicable in various healthcare scenarios.

References

1. Bodenreider, O., Cornet, R., Vreeman, D.J.: Recent developments in clinical terminologies - snomed ct, loinc, and rxnorm. Yearbook of medical informatics **27**,

129–139 (Aug 2018)
2. Bottrighi, A., Piovesan, L., Terenziani, P.: Run-time support to comorbidities in glare-sscpm. (2019)
3. Boytcheva, S., Angelova, G., Angelov, Z., Tcharaktchiev, D.: Mining comorbidity patterns using retrospective analysis of big collection of outpatient records. Health information science and systems **5**(1), 1–9 (2017)
4. Cha, S., Kim, S.S.: Discovery of association rules patterns and prevalence of comorbidities in adult patients hospitalized with mental and behavioral disorders. In: Healthcare. vol. 9, p. 636. Multidisciplinary Digital Publishing Institute (2021)
5. Davis, J., Chung, R., Juarez, D.: Prevalence of comorbid conditions with aging among patients with diabetes and cardiovascular disease. Hawaii medical journal **70**, 209–13 (10 2011)
6. Fatemi, B., Rabbi, F., MacCaull, W.: A Validated Learning Approach to Healthcare Process Analysis Through Contextual and Temporal Filtering, pp. 108–137. Springer Berlin Heidelberg, Berlin, Heidelberg (2024). https://doi.org/10.1007/978-3-662-68191-6_5
7. Hossain, M.E., Khan, A., Uddin, S.: Understanding the comorbidity of multiple chronic diseases using a network approach. In: Proceedings of the Australasian Computer Science Week Multiconference. pp. 1–7 (2019)
8. Jones, P.J., Ma, R., McNally, R.J.: Bridge centrality: A network approach to understanding comorbidity. Multivariate behavioral research **56**(2), 353–367 (2021)
9. Mukhiya, S.K.: A software framework for adaptive and interoperable internet-delivered psychological treatments. Ph.D. thesis, Høgskulen på Vestlandet (2021), https://hvlopen.brage.unit.no/hvlopen-xmlui/handle/11250/2778982
10. Mukhiya, S.K., Ahmed, U., Rabbi, F., Pun, K.I., Lamo, Y.: Adaptation of idpt system based on patient-authored text data using nlp. 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) pp. 226–232 (2020). https://doi.org/10.1109/CBMS49503.2020.00050, https://api.semanticscholar.org/CorpusID:221474223
11. Mukhiya, S.K., Wake, J.D., Inal, Y., Lamo, Y.: Adaptive systems for internet-delivered psychological treatments. IEEE Access **8**, 112220–112236 (2020). https://doi.org/10.1109/ACCESS.2020.3002793
12. Nordby, E.S., Gjestad, R., Kenter, R.M., Guribye, F., Mukhiya, S.K., Lundervold, A.J., Nordgreen, T.: The effect of sms reminders on adherence in a self-guided internet-delivered intervention for adults with adhd. Frontiers in Digital Health **4**, 821031 (2022)
13. Organization., W.H.: ICD-10 : international statistical classification of diseases and related health problems / World Health Organization. World Health Organization Geneva, 10th revision, 2nd ed. edn. (2004)
14. Partington, A., Wynn, M., Suriadi, S., Ouyang, C., Karnon, J.: Process mining for clinical processes: A comparative analysis of four australian hospitals. ACM Trans. Manage. Inf. Syst. **5**(4) (jan 2015). https://doi.org/10.1145/2629446, https://doi.org/10.1145/2629446
15. Piovesan, L., Terenziani, P., Dupré, D.T.: Conformance analysis for comorbid patients in answer set programming. Journal of Biomedical Informatics **103**, 103377 (2020)
16. Pnueli, A.: The temporal logic of programs. In: 18th Annual Symposium on Foundations of Computer Science (sfcs 1977). pp. 46–57 (1977). https://doi.org/10.1109/SFCS.1977.32