KMANPUB | AI and Tech in Behavioral and Social Sciences Vol. 4 No. 4 (2026): 1-11

Article history:
Received 10 March 2026
Revised 01 July 2026
Accepted 08 July 2026
Initial Publication 12 July 2026
Final Publication 01 October 2026

Enhancing BPMN Repository Consistency Using Hybrid Similarity Metrics and Scalable Retrieval

Amir Hossein. Kabiri Nameghi[1], Pouya. Sohofi[1], Hassan. Naderi[2*]

1. M.Sc. Student, Department of Software Engineering, Iran University of Science and Technology, Tehran, Iran
2. Faculty Member, Department of Software Engineering, Iran University of Science and Technology, Tehran, Iran

*** Corresponding author email address:** naderi@iust.ac.ir

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Kabiri Nameghi, A. H., Sohofi, P., & Naderi, H. (2026). Enhancing BPMN Repository Consistency Using Hybrid Similarity Metrics and Scalable Retrieval. *AI and Tech in Behavioral and Social Sciences, 4*(4), 1-11.
https://doi.org/10.61838/kman.aitech.5792

A B S T R A C T

Business Process Model and Notation (BPMN) has become a dominant standard for representing, communicating, and automating organizational workflows. Yet the growth of large process repositories has intensified inconsistencies caused by heterogeneous modeling conventions, duplicated fragments, inconsistent activity labels, and structurally similar models represented in divergent ways. These problems reduce process reuse, weaken repository governance, complicate integration, and increase maintenance costs. This study proposes an integrated framework for enhancing consistency in BPMN repositories by combining hybrid similarity assessment with repository-level recommendation mechanisms. The framework converts BPMN models into directed graphs and computes multilevel structural similarity using dynamic vector signatures inspired by BPMN-Sim. Semantic similarity is calculated from Sentence-BERT embeddings and cosine similarity, allowing conceptually equivalent labels to be detected even when lexical forms differ. The resulting hybrid similarity score is used by recommendation modules that propose standardized labels and reusable process elements during modeling. HDBSCAN clustering and HNSW indexing are incorporated to support scalable candidate retrieval. The approach was evaluated using public BPMN repositories and benchmark tasks. Results show that the proposed method outperforms structural-only, lexical, and semantic-only baselines. After label standardization, the method achieved Precision = 0.91, Recall = 0.90, and standard F1 = 0.90; after element standardization, it achieved Precision = 0.90, Recall = 0.88, and standard F1 = 0.89. The findings indicate that similarity assessment should be treated not only as a retrieval technique but also as a mechanism for repository standardization, reuse, and intelligent process modeling.

Keywords: *Business Process Management; BPMN; process similarity; repository consistency; Sentence-BERT; HDBSCAN; HNSW; process model retrieval; process reuse*

1. Introduction

Business Process Management (BPM) has become a central component of digital transformation because it enables organizations to describe, analyze, automate, and continuously improve their operational routines. Business process models are no longer merely documentation artifacts; they are now used to support enterprise architecture, workflow automation, process mining, compliance analysis, robotic process automation, and organizational knowledge management (Dumas et al., 2018; van der Aalst, 2016). As process-aware information systems become more deeply embedded in public and private organizations, the quality and consistency of

process models directly affect the reliability of downstream analysis and system implementation.

Among available process modeling languages, Business Process Model and Notation (BPMN) has become one of the most widely used standards because it offers a graphical notation that is understandable to business stakeholders while retaining sufficient formal structure for technical implementation (Mendling et al., 2010; Object Management Group, 2013). Empirical investigations of real-world repositories show that BPMN is used across diverse organizational domains and that large collections of models are increasingly created outside controlled laboratory environments (Türker et al., 2022). This broad adoption, however, has produced repositories in which models are created by different analysts, departments, and external contractors over long periods. As a result, repositories often contain duplicated fragments, inconsistent naming conventions, alternative structures for similar process behavior, and models with different levels of abstraction.

The problem is not limited to visual inconsistency. When similar activities are labeled differently, repository search becomes unreliable and model reuse declines. When structurally similar subprocesses are modeled independently, organizations lose opportunities for consolidation and standardization. When process repositories grow without governance, integration projects and process mining initiatives are forced to handle unnecessary heterogeneity, which increases cost and reduces interpretability (Dijkman et al., 2011, 2012). Therefore, consistency management in BPMN repositories is a practical requirement for organizations that seek to transform collections of models into reusable digital assets.

Similarity measurement has been one of the principal approaches for addressing this problem. Early studies measured similarity primarily through structural comparison, graph edit distance, common nodes, control-flow relations, and behavioral profiles (Dijkman et al., 2009; Van Dongen et al., 2008; Weidlich et al., 2011). These techniques are useful because BPMN models are inherently graph-structured artifacts. However, purely structural methods may treat two semantically equivalent processes as dissimilar when designers use different labels or introduce minor modeling variations. Conversely, purely lexical or semantic methods may overestimate similarity between models that share labels but differ substantially in execution logic. This limitation has motivated hybrid methods that combine structural and semantic evidence (Ehrig et al., 2007; Zhou et al., 2019).

Recent work has improved the precision of structural BPMN comparison. BPMN-Sim, for example, provides a multilevel structural similarity technique designed for BPMN 2.0.2 elements and subprocess hierarchies (Garcia et al., 2023). At the same time, advances in natural language processing have made it possible to compute more accurate semantic similarity between process labels. Sentence-BERT transforms text into dense sentence embeddings, allowing activity labels to be compared beyond surface-level lexical overlap (Reimers & Gurevych, 2019). These developments suggest that effective repository consistency mechanisms should jointly consider control-flow structure and label semantics.

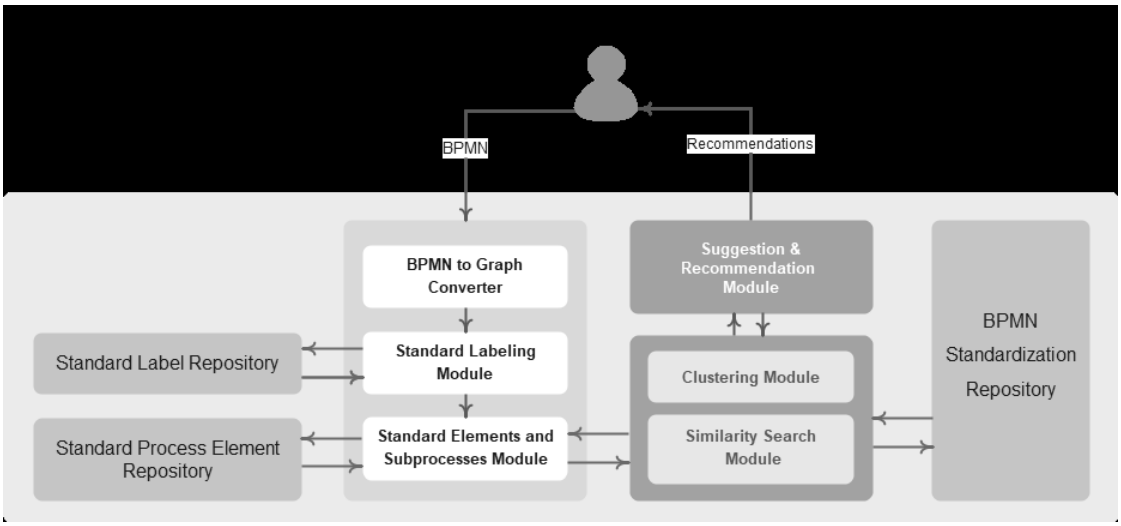
A second challenge concerns scalability. Many similarity algorithms rely on pairwise comparison and become computationally expensive when repositories contain thousands of models. Large-scale process retrieval research has therefore emphasized indexing, feature-based search, and efficient query mechanisms (Kunze & Weske, 2010; Yan et al., 2012; Zhu et al., 2024). In high-dimensional retrieval tasks, approximate nearest-neighbor methods such as HNSW have shown strong performance in balancing search speed and recall (Aumüller et al., 2020; Malkov & Yashunin, 2020). Density-based clustering, especially HDBSCAN, offers complementary benefits because it can group models without requiring the number of clusters to be fixed in advance and can isolate outliers (Campello et al., 2013; McInnes et al., 2017; Stewart & Al-Khassaweneh, 2022).

Despite these advances, a gap remains between similarity computation and practical consistency enhancement. Existing methods often identify similar models but do not actively guide designers toward standardized labels or reusable process fragments during modeling. In real organizational settings, a useful system must do more than retrieve similar models; it must provide design-time recommendations that reduce duplication, encourage reuse, and support repository governance. This study addresses this gap by proposing an integrated framework that combines structural and semantic similarity, label standardization, reusable-element recommendation, HDBSCAN clustering, and HNSW retrieval. Figure 1 presents the overall architecture of the proposed approach.

AITBSS AI and Tech in Behavioral and Social Sciences E-ISSN: 3041-9433

Figure 1

Architecture of the proposed similarity-based BPMN standardization framework.



As shown in Figure 1, the framework links a BPMN-to-graph converter, standard label and process-element repositories, a similarity search module, a clustering module, and a suggestion module. The architecture is designed to make similarity analysis operational during model construction rather than using it only as an offline retrieval tool.

Accordingly, the study addresses four research questions: whether hybrid structural-semantic similarity improves BPMN model matching; whether label standardization improves semantic consistency; whether reusable-element recommendation reduces repository redundancy; and whether HDBSCAN combined with HNSW improves scalable BPMN retrieval compared with conventional repository-search approaches.

The main contributions of the study are: (1) a dynamic-vector structural representation for BPMN path comparison; (2) integration of structural similarity with Sentence-BERT-based label semantics; (3) design-time recommendation mechanisms for standardized labels and reusable fragments; and (4) a repository-level architecture that combines density-based clustering and approximate nearest-neighbor indexing for scalable retrieval.

2. Methods and Materials

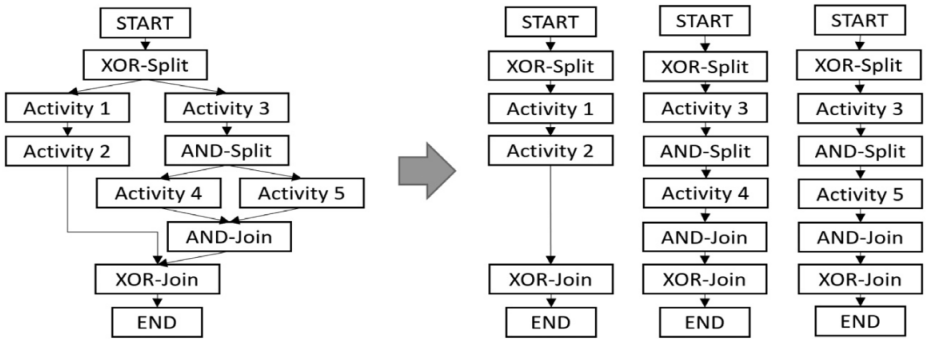
This study developed a hybrid framework for improving consistency in BPMN repositories. The methodological design follows a pipeline in which BPMN models are converted into graph representations, structural and semantic features are extracted, hybrid similarity is calculated, and the resulting similarity information is used for recommendation, clustering, and retrieval. The method was developed to support both repository analysis and design-time assistance, meaning that a new or updated model can be compared with existing repository content while the designer is still constructing the process.

Each BPMN model is represented as a directed graph composed of start and end events, activities, gateways, intermediate events, and sequence flows. To compute structural similarity, each model is decomposed into execution paths extending from start nodes to end nodes. This path-based decomposition follows the logic used in BPMN-Sim and related graph-based approaches, because execution paths preserve important control-flow information while allowing complex models to be represented as comparable units (Garcia et al., 2023; Van Dongen et al., 2008). Figure 2 illustrates the decomposition of a BPMN model into multiple execution paths.

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

Figure 2

Decomposition of a BPMN model into execution paths.



After decomposition, vector signatures are generated for each path. The original Sim-KT logic uses fixed signatures that indicate the presence and frequency of node and relation types. Figure 3 shows the basic vector-signature structure and Figure 4 shows how paths are mapped to vector signatures. These representations allow pairwise path comparison without directly applying costly graph edit operations to entire models.

Figure 3

Static vector-signature structure used in path representation.

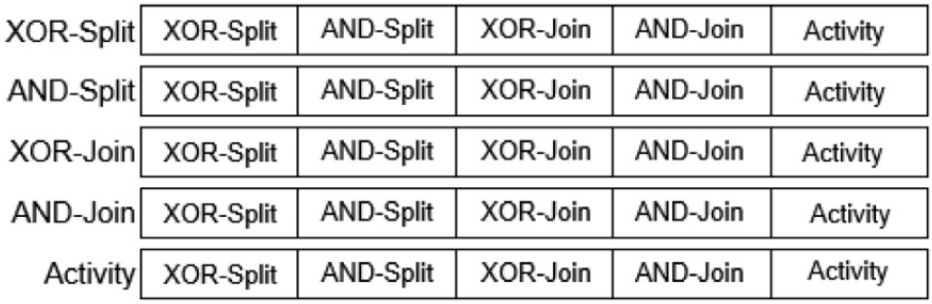
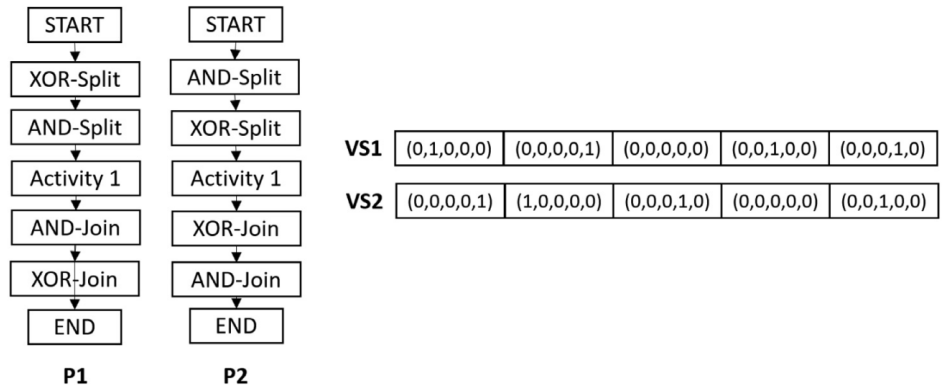


Figure 4

Mapping of execution paths into vector signatures.

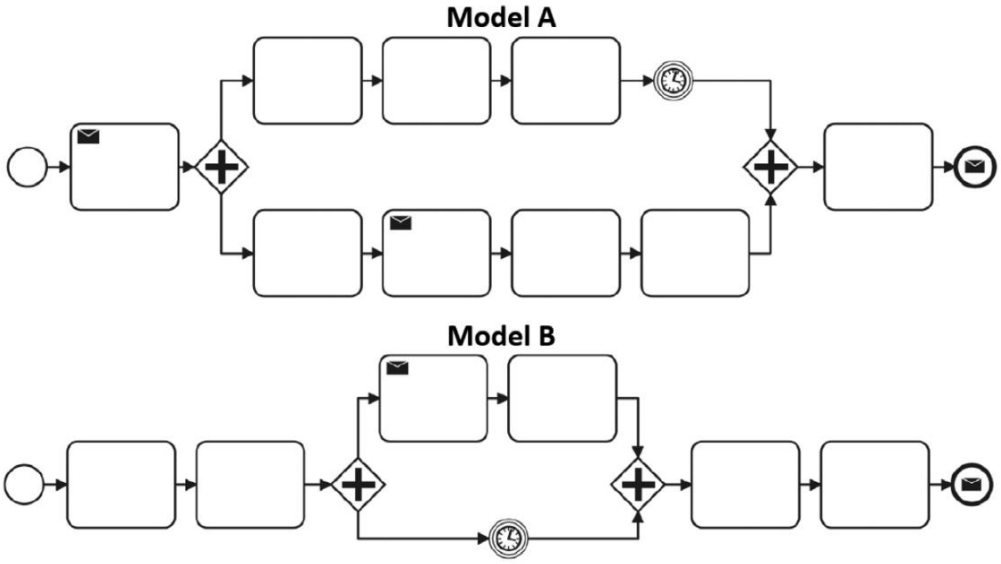


AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

Because BPMN 2.0.2 contains many element types, fixed vector signatures may become sparse and inefficient when models use only a small subset of possible constructs. Therefore, the proposed framework applies dynamic vector signatures that allocate dimensions only to elements and relations appearing in the compared models. This design improves memory efficiency and reduces unnecessary comparison cost while preserving structural discriminability. Figure 5 shows the mapping of paths into dynamic vector signatures, and Figure 6 provides an example of dynamic-vector calculation for two process models.

Figure 5

Mapping of BPMN paths into dynamic vector signatures.



DVS: V (vector) / C (component)

V	Event-Start-Simple	Event-End-Message	Event-Intermediate-Catching-Timer	Activity-Task	Activity-Task-Send	Gateway-Parallel-Split	Gateway-Parallel-Join
C	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,0)

Figure 6

Example of DVS calculation for two BPMN models.

Model A

	Event-Start-Simple	Event-End-Message	Event-Intermediate-Catching-Timer	Activity-Task	Activity-Task-Send	Gateway-Parallel-Split	Gateway-Parallel-Join
DVS 1	(0,0,0,0,1,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,1)	(0,1,1,2,0,0,0)	(0,0,0,0,0,1,0)	(0,0,0,1,0,0,0)	(0,0,0,1,0,0,0)
DVS 2	(0,0,0,0,1,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,0)	(0,1,0,1,1,0,1)	(0,0,0,1,0,1,0)	(0,0,0,1,0,0,0)	(0,0,0,1,0,0,0)

Model B

	Event-Start-Simple	Event-End-Message	Event-Intermediate-Catching-Timer	Activity-Task	Activity-Task-Send	Gateway-Parallel-Split	Gateway-Parallel-Join
DVS 1	(0,0,0,1,0,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,0)	(0,1,0,2,0,1,1)	(0,0,0,1,0,0,0)	(0,0,0,0,1,0,0)	(0,0,0,1,0,0,0)
DVS 2	(0,0,0,1,0,0,0)	(0,0,0,0,0,0,0)	(0,0,0,0,0,0,1)	(0,1,0,2,0,1,0)	(0,0,0,0,0,0,0)	(0,0,1,0,0,0,0)	(0,0,0,1,0,0,0)

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

Semantic similarity is computed from the textual labels of BPMN elements. Each activity label is transformed into a dense embedding using Sentence-BERT, and cosine similarity is used to quantify semantic proximity between labels (Reimers & Gurevych, 2019). Semantic similarity can be defined as SemSim(a,b) = cos(e_a,e_b), where e_a and e_b are the Sentence-BERT embeddings of two activity labels. The final path-level and model-level similarity scores combine structural and semantic similarity by a weighting coefficient: HybridSim = (1 - alpha) x StructSim + alpha x SemSim, where alpha in [0,1]. A value closer to zero emphasizes graph structure, whereas a value closer to one emphasizes label semantics. Precision, Recall, and F1-score were calculated consistently using F1 = 2 x Precision x Recall / (Precision + Recall). This adjustable parameter allows the method to be adapted to repositories in which structural variation or label variation is the dominant inconsistency source.

The framework includes two recommendation modules. The standard-label module compares each new activity label with the repository vocabulary. When semantic similarity exceeds a threshold, the system recommends the existing standardized label instead of storing another variant. The standard-element module compares new fragments and subprocesses with previously validated patterns. If a similar fragment exists, reuse is recommended; if not, the fragment can be considered as a candidate new standard after expert review. This logic follows the repository-refactoring perspective in which similarity analysis supports reuse and redundancy reduction rather than merely model retrieval (Dijkman et al., 2011; Khlif et al., 2017; Pérez-Castillo et al., 2019).

For scalability, process models are first clustered using HDBSCAN based on extracted structural-semantic feature vectors. HDBSCAN was selected because it does not require a predefined number of clusters and can detect outlier models that should not be forced into unsuitable groups (Campello et al., 2013; McInnes et al., 2017). Within each cluster, approximate nearest-neighbor retrieval is performed using HNSW indexing, which supports efficient search over high-dimensional embeddings and has been shown to provide strong speed-recall trade-offs (Aumüller et al., 2020; Malkov & Yashunin, 2020). The evaluation compared the proposed method with structural-only, semantic-only, lexical, and conventional repository-search baselines using Precision, Recall, F1-score, and comparative retrieval behavior.

Evaluation protocol. Precision, Recall, and F1-score were calculated from true-positive, false-positive, and false-negative decisions at the relevant evaluation level: model pairs for similarity assessment, activity labels for label standardization, and process fragments for reusable-element detection. The reported F1-scores were calculated using the standard harmonic-mean formula, F1 = 2 x Precision x Recall / (Precision + Recall).

3. Findings and Results

The evaluation was organized around four research questions: whether the hybrid method improves similarity assessment, whether label standardization improves semantic consistency, whether element standardization reduces redundancy, and whether clustering with approximate nearest-neighbor retrieval improves large-repository search. The results support the assumption that repository consistency improves when similarity analysis is treated as a combined structural, semantic, and retrieval problem rather than as a single isolated metric.

For the first research question, the Dispatch-of-goods dataset was used to compare Precision, Recall, and F1-score for the hybrid method. As shown in Figure 7, the proposed method achieved balanced performance across all three metrics. The values indicate that combining structural information with semantic label representations can detect similar process models more reliably than either dimension alone. The relatively close values of Precision, Recall, and F1-score also suggest that the method does not gain accuracy merely by becoming overly conservative or overly permissive; instead, it maintains a stable trade-off between detecting relevant similar models and avoiding false matches.

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

Figure 7

Evaluation of Precision, Recall, and F1-score in the Dispatch-of-goods dataset.

Table 1

Hybrid similarity performance in the Dispatch-of-goods dataset.

Dataset	Precision	Recall	F1-score
Dispatch-of-goods	0.90	0.88	0.89

Table 1 reports the same values numerically. The F1-score of 0.89 indicates that the integrated structural-semantic model achieved reliable overall classification of similar models. This result is consistent with prior findings that process similarity benefits from combining topology and semantics (Ehrig et al., 2007; Garcia et al., 2023; Zhou et al., 2019).

For the second research question, label-standardization performance was evaluated in the Credit-scoring dataset. Table 2 shows that the proposed method outperformed Levenshtein, TF-IDF, BM25, FastText-average, and SBERT baselines. The improvement over lexical methods is expected because lexical methods are sensitive to surface-form differences. The improvement over SBERT alone indicates that semantic embeddings become more useful when they are embedded within a repository-standardization mechanism rather than used as a standalone similarity score.

Table 2

Evaluation after label standardization in the Credit-scoring dataset.

Method	Precision	Recall	F1-score
Levenshtein	0.75	0.73	0.74
TF-IDF	0.82	0.79	0.80
BM25	0.83	0.81	0.82
FastText-avg	0.85	0.82	0.83
SBERT	0.89	0.87	0.88
Proposed method	0.91	0.90	0.90

As shown in Table 2, the proposed method achieved Precision = 0.91, Recall = 0.90, and standard F1-score = 0.90. The improvement is practically meaningful because standardized labels increase the likelihood that future designers retrieve and reuse existing process fragments. From a repository-governance perspective, this result

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

shows that label harmonization is not only a cosmetic issue; it directly improves model comparability and search effectiveness.

For the third research question, the element-standardization module was evaluated on the Recourse dataset. Table 3 compares the proposed method with graph edit distance (GED), Behavioral Profile, and BPMN-sim baselines. The proposed method achieved the strongest reported performance, with Precision = 0.90, Recall = 0.88, and standard F1-score = 0.89. These results indicate that reusable process fragments can be detected more accurately when structural and semantic evidence are evaluated jointly.

Table 3

Evaluation after element standardization in the Recourse dataset.

Method	Precision	Recall	F1-score
GED	0.78	0.76	0.77
Behavioral Profile	0.80	0.79	0.79
BPMN-sim	0.84	0.80	0.82
Proposed method	0.90	0.88	0.89

The comparison in Table 3 is important because process-element reuse depends on more than local label similarity. A candidate fragment may use similar labels but implement different control-flow logic, or it may be structurally similar while using different terminology. The proposed framework reduces both types of error by applying hybrid comparison before recommending reuse. This result supports the use of similarity analysis as an active refactoring and standardization mechanism.

For the fourth research question, the scalability component was treated as a retrieval-oriented architectural layer. The HDBSCAN + HNSW configuration combines density-based grouping with approximate nearest-neighbor retrieval and is intended to reduce the search space compared with exhaustive pairwise comparison. Because standalone CSR, query-time, indexing-time, recall@k, precision@k, and memory-use values are not reported, the scalability component is interpreted as a design contribution rather than a fully benchmarked runtime result.

Overall, the reported results support improvements in model matching, label consistency, and process-element reuse. The central empirical contribution is strongest for hybrid similarity and standardization, while the scalability layer remains an architectural extension that requires further quantitative runtime validation.

4. Discussion and Conclusion

The findings of this study show that BPMN repository consistency can be improved by integrating hybrid similarity analysis with design-time recommendation and scalable retrieval. This is important because many organizations now treat process repositories as long-term digital assets. If the repository contains inconsistent labels, duplicated fragments, and incompatible modeling conventions, downstream activities such as process mining, compliance checking, automation, and enterprise integration become less reliable. The results suggest that similarity assessment should not be used only to answer the question of whether two models are similar; it should also guide how future models are designed and standardized.

The first major finding is that the hybrid similarity engine provides more balanced performance than single-perspective baselines. Structural similarity captures the topology of the process, including gateways, events, activities, and sequence flows. This is necessary because two business processes may use similar terminology while implementing different control-flow behavior. However, structural information alone is not sufficient because different designers often use different labels for equivalent activities. Semantic similarity based on Sentence-BERT embeddings compensates for this weakness by identifying conceptual equivalence between labels (Reimers & Gurevych, 2019). The results therefore confirm the theoretical expectation that structural and semantic information are complementary rather than interchangeable.

The improvement after label standardization is especially relevant for organizational repositories. A repository may contain activities such as “register customer,” “create client record,” and “open customer profile,” all referring to nearly the same function. If each variant is stored as an independent concept, model retrieval becomes fragmented and process reuse declines. The

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

proposed standard-label module reduces this fragmentation by recommending existing labels when semantic similarity is high. This aligns with the broader process-governance view that modeling quality depends not only on syntactic correctness but also on terminological consistency (Awadid & Nurcan, 2016; Mendling et al., 2010).

The element-standardization results extend this argument from labels to structures. Reusable fragments and subprocesses are often hidden in repositories because they appear under different names or with minor structural modifications. Identifying these fragments helps organizations reduce redundancy, consolidate similar processes, and preserve validated design patterns. This finding is consistent with earlier research on process model refactoring and consolidation, which emphasized that model repositories contain opportunities for systematic improvement if similar fragments can be detected reliably (Dijkman et al., 2011; La Rosa et al., 2013; Pérez-Castillo et al., 2019). The contribution of this study is to connect such reuse to an operational recommendation framework.

Scalability is an important architectural objective of the framework. Pairwise comparison is theoretically straightforward but practically inefficient when repositories grow. HDBSCAN + HNSW addresses this issue by narrowing the search space and then retrieving approximate nearest neighbors efficiently. In the present study, however, this component is discussed primarily as an architectural layer because repository size, indexing time, query latency, recall@k, precision@k, and memory consumption were not independently benchmarked. HDBSCAN avoids the limitation of K-means, which requires a predefined number of clusters and performs poorly when cluster shapes and densities vary (Campello et al., 2013; McInnes et al., 2017). HNSW then supports efficient retrieval in the resulting feature space, consistent with findings from approximate nearest-neighbor research (Aumüller et al., 2020; Malkov & Yashunin, 2020).

From a practical perspective, the framework can support several organizational scenarios. In digital transformation projects, process analysts often redesign processes that are already partially represented elsewhere in the organization. A design-time recommendation mechanism can reduce this duplication. In merger and integration projects, similar processes from different units can be identified and consolidated. In public-sector digital services, standard process labels and reusable service fragments can support cross-agency harmonization. In process mining projects, cleaner and more consistent model repositories can improve the interpretability of discovered or conformance-checked models (van der Aalst, 2016; Zhu et al., 2024).

The study also has implications for AI-assisted process modeling. Generative and recommendation-based modeling tools require reliable repository knowledge to produce useful suggestions. If the underlying repository is inconsistent, automated suggestions may reproduce existing disorder. The proposed approach offers a structured way to prepare repositories for intelligent modeling by establishing standardized labels, reusable fragments, and efficient retrieval pathways. In this sense, the framework can be viewed as an infrastructure layer for future process-modeling assistants.

Several limitations should be acknowledged. First, the evaluation relied on public BPMN repositories and benchmark datasets. These datasets provide transparency and reproducibility, but they may not fully represent the constraints of proprietary industrial repositories. Second, the structural-semantic weighting coefficient was treated as configurable rather than automatically learned. Future work should investigate adaptive weighting based on repository characteristics. Third, the current approach focuses on static BPMN models. Behavioral execution logs, event sequences, and runtime conformance information were not included, although such information may improve similarity assessment for executable processes. Fourth, the study does not provide full quantitative evidence for the scalability layer; therefore, future studies should report runtime and retrieval metrics on repositories of different sizes.

Future research should test the framework in real industrial and public-sector repositories, evaluate user acceptance among process analysts, and investigate whether design-time recommendations measurably reduce modeling time and inconsistency rates. Further work should also integrate multilingual embeddings, process-mining logs, knowledge graphs, and large language models. A reproducible implementation package, including code, parameter files, dataset links, and evaluation scripts, would substantially improve the credibility and reuse value of the study. Overall, the findings indicate that hybrid similarity analysis, when combined with standardization and scalable retrieval, provides a promising foundation for improving the quality, sustainability, and practical usability of BPMN repositories.

Authors' Contributions

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

Amir Hossein Kabiri Nameghi contributed to conceptualization, methodology, software development, data collection, formal analysis, validation, visualization, and manuscript preparation. Pouya Sohofi contributed to investigation, data curation, implementation, experimental evaluation, and manuscript revision. Hassan Naderi contributed to supervision, methodology refinement, validation, critical review, project administration, and final approval of the manuscript. All authors read and approved the final manuscript.

Declaration

Generative artificial intelligence was used only as a language-editing and writing-support tool during manuscript preparation. All scientific ideas, methodological design, experimental procedures, data interpretation, and final manuscript content were developed, verified, and approved by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the manuscript.

Transparency Statement

The datasets analyzed in this study were obtained from publicly available BPMN repositories and benchmark datasets cited in the manuscript. Additional implementation details and configuration information are available from the corresponding author upon reasonable request.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethics Considerations

This study did not involve human participants, animals, clinical data, private organizational data, or personally identifiable information. The evaluation was based on publicly available BPMN repositories and benchmark datasets intended for research use. Therefore, formal ethical approval and informed consent were not required.

References

Aumüller, M., Bernhardsson, E., & Faithfull, A. (2020). ANN-Benchmarks: A benchmarking tool for approximate nearest neighbor algorithms. *Information Systems*, *87*, 101374. https://doi.org/10.1016/j.is.2019.02.006

Awadid, A., & Nurcan, S. (2016). A systematic literature review of consistency among business process models. Proceedings of BPMDS/EMMSAD,

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. Advances in Knowledge Discovery and Data Mining,

Dijkman, R., Dumas, M., García-Bañuelos, L., & Käärik, R. (2009). Aligning business process models. Proceedings of the IEEE International Enterprise Distributed Object Computing Conference,

Dijkman, R., La Rosa, M., & Reijers, H. A. (2011). Identifying refactoring opportunities in process model repositories. *Information and Software Technology*, *53*(9), 937-948. https://doi.org/10.1016/j.infsof.2011.04.001

Dijkman, R., La Rosa, M., & Reijers, H. A. (2012). Managing large collections of business process models: Current techniques and challenges. *Computers in Industry*, *63*(2), 91-97. https://doi.org/10.1016/j.compind.2011.12.003

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). *Fundamentals of Business Process Management* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-56509-4

Ehrig, M., Koschmider, A., & Oberweis, A. (2007). Measuring similarity between semantic business process models. Proceedings of the Fourth Asia-Pacific Conference on Conceptual Modelling,

Garcia, M. T., Nunes, M. M., Fantinato, M., & Peres, S. M. (2023). BPMN-Sim: A multilevel structural similarity technique for BPMN process models. *Information Systems*, *116*, 102211. https://doi.org/10.1016/j.is.2023.102211

Khlif, W., Ben-Abdallah, H., & Ben Ayed, N. E. (2017). A methodology for the semantic and structural restructuring of BPMN models. *Business Process Management Journal*, *23*(1), 16-46. https://doi.org/10.1108/BPMJ-12-2015-0186

Kunze, M., & Weske, M. (2010). Metric trees for efficient similarity search in large process model repositories. Business Process Management Workshops,

La Rosa, M., Dumas, M., Uba, R., & Dijkman, R. (2013). Business process model merging: An approach to business process consolidation. *ACM Transactions on Software Engineering and Methodology*, *22*(2), 1-42. https://doi.org/10.1145/2430545.2430547

Malkov, Y. A., & Yashunin, D. A. (2020). Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(4), 824-836. https://doi.org/10.1109/TPAMI.2018.2889473

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

Mendling, J., Reijers, H. A., & van der Aalst, W. M. P. (2010). Seven process modeling guidelines (7PMG). *Information and Software Technology*, *52*(2), 127-136. https://doi.org/10.1016/j.infsof.2009.08.004

Object Management Group. (2013). *Business Process Model and Notation (BPMN), Version 2.0.2*. Object Management

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

Group. https://www.ce.gov.br/setur/wp-content/uploads/sites/92/2018/04/formal-13-12-09.pdf

Pérez-Castillo, R., Fernández-Ropero, M., & Piattini, M. (2019). Business process model refactoring applying IBUPROFEN: An industrial evaluation. *Journal of Systems and Software*, *147*, 86-103. https://doi.org/10.1016/j.jss.2018.10.012

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing,

Stewart, G., & Al-Khassaweneh, M. (2022). An implementation of the HDBSCAN* clustering algorithm. *Applied Sciences*, *12*(5), 2405. https://doi.org/10.3390/app12052405

Türker, J., Völske, M., & Heinze, T. S. (2022). BPMN in the Wild: A Reprise. Proceedings of ZEUS 2022,

van der Aalst, W. M. P. (2016). *Process Mining: Data Science in Action* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-49851-4

Van Dongen, B. F., Dijkman, R. M., & Mendling, J. (2008). Measuring similarity between business process models. Advanced Information Systems Engineering,

Weidlich, M., Dijkman, R., & Mendling, J. (2011). The ICoP framework: Identification of correspondences between process models. Advanced Information Systems Engineering,

Yan, Z., Dijkman, R., & Grefen, P. (2012). Fast business process similarity search. *Distributed and Parallel Databases*, *30*(2), 105-144. https://doi.org/10.1007/s10619-012-7089-z

Zhou, C., Liu, C., Zeng, Q., Lu, F., & Duan, H. (2019). A comprehensive process similarity measure based on models and logs. *IEEE Access*, *7*, 69257-69273. https://doi.org/10.1109/ACCESS.2018.2885819

Zhu, R., Huang, Y., Liu, L., Zhou, W., Zhang, X., Chen, Y., & Cai, L. (2024). Business process retrieval from large model repositories for Industry 4.0. *IEEE Transactions on Services Computing*, *17*(1), 306-321. https://doi.org/10.1109/TSC.2023.3348294

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433