International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 8 (2025) 109–125

Paper Type: Original Article

AMFUP: An Adaptive Multi-Modal Framework for Unexpected Pattern Discovery in Dynamic Environments

A. Sharaf Eldin [1,2], Eman I. Salem [1,*], Nissreen El-Saber [3] and Khalid A. Eldrandaly [3]

[1] Department of Information Systems, Faculty of Information Technology and Computer Science, Sinai University, Arish, 16020, Egypt.

[2] Department of Information Systems, Faculty of Computers and AI, Helwan University, Cairo, 11795, Egypt. .

[3] Department of Information Systems, Faculty of Computers and Information, Zagazig University, Zagazig, 44519, Egypt.

Emails: ahmed.sharaf@su.edu.eg; eman.ali@su.edu.eg; naelsaber@fci.zu.edu.eg; eldrandaly66@gmail.com.

Received: 08 Feb 2025 **Revised**: 30 Mar 2025 **Accepted**: 12 Jul 2025 **Published**: 14 Jul 2025

Abstract

The discovery of unexpected patterns represents a critical advancement in knowledge discovery, enabling the identification of rare yet meaningful contradictions that defy conventional frequency-based assumptions. However, existing techniques frequently suffer from three major limitations: (1) rigidity in adapting to evolving data distributions, (2) limited capacity to interpret semantic relationships among patterns, and (3) dependency on extensive manual tuning for parameter optimization. To overcome these challenges, this paper introduces the Adaptive Multi-Modal Framework for Unexpected Pattern Discovery (AMFUP), a comprehensive architecture that enhances adaptability, semantic understanding, and automation in pattern mining. AMFUP integrates three synergistic components: the Multi-Modal Pattern Embedding (MMPE), which captures structural, semantic, and statistical dimensions of patterns through neural architectures; the Dynamic Belief Adaptation (DBA) module, which continuously evolves belief systems in response to concept drift; and the Automated Parameter Learning (APL) mechanism, which employs meta-learning to optimize parameters without human intervention. Experimental results across eight datasets demonstrate that AMFUP achieves the highest Pattern Quality Score (PQS) of 0.96, representing a 28% improvement over the best baseline method (belief-driven). Concurrently, AMFUP establishes itself as the fastest method with a runtime of 12 minutes, achieving a 73.3% reduction compared to UCRP-miner (45 min) and a 33.3% reduction compared to Random Forest (18 min). AMFUP achieves a 43.28% increase in PQS and a 46.68% reduction in execution runtime compared to the average performance of baseline methods.

Keywords: Unexpected Patterns, DM, KDDs, Multi-Modal Pattern Embedding, Automated Parameter Learning.

1 | Introduction

The domain of knowledge discovery has consistently confronted the essential challenge of differentiating significant patterns from noise within extensive datasets. Although conventional pattern mining methodologies are proficient in identifying frequent item sets and association rules, they frequently inundate analysts with numerous apparent or trivial patterns that contribute minimal actionable insights. Knowledge discovery in databases (KDDs), commonly referred to as data mining (DM), is characterized as the non-trivial endeavor of uncovering valid, novel, potentially useful, and ultimately comprehensible patterns from

Corresponding Author: eman.ali@su.edu.eg

data [1]. Nonetheless, a substantial portion of the research within the field of knowledge discovery predominantly emphasizes the validity aspect, while the other two dimensions, namely novelty and usefulness, receive comparatively less attention [2, 3]. The identification of unexpected patterns signifies a transformative shift from frequency-driven mining methodologies to those centered on contradiction-based exploration. Unexpectedness constitutes one of the nine criteria proposed to assess the significance of a pattern in relation to a specific research inquiry [4]. The criterion of unexpectedness holds substantial importance, as it underscores the need to identify patterns that confront users' pre-existing beliefs or established knowledge [5]. Unlike conventional methods which concentrate on the detection of commonly occurring patterns, unexpected pattern mining seeks to reveal rare yet significant patterns that oppose established convictions or normative expectations concerning behavior [6]. Within the realm of healthcare, it is widely recognized that, patients undergoing treatment for hypertension benefit from improved control of their blood pressure levels. Nonetheless, an unexpected event may suggest that, patients who are administered hypertension medication alongside specific dietary supplements may present adverse reactions. Although such occurrences are rare, they hold considerable significance for both patient safety and treatment guidelines. Numerous practical applications necessitate the identification of unusual or unexpected patterns (rare patterns). For instance, the combination of diapers and beer is observed less frequently than that of milk and bread, yet it yields a greater profit margin compared to milk and bread [7]. Notwithstanding the considerable advancements in the discovery of unexpected patterns, current methodologies exhibit three fundamental limitations that significantly hinder their practical application. Firstly, the leading methods, such as UCRP-miner and clustering-based techniques, depend on fixed belief systems established from historical data snapshots [8, 9]. In an attempt to reduce the occurrence of redundant patterns and the consequent degradation in performance, researchers have proposed effective strategies designed to obtain a concise representation of patterns, referred to as closed rare patterns [8]. However, these static concepts do not possess the ability to adjust to changes in the fundamental data distributions, making them inappropriate for dynamic environments in which patterns evolve over time. Furthermore, traditional methods demand considerable manual adjustment of similarity thresholds, support values, and clustering parameters. For example, the DBSCAN algorithm, utilized for the identification of candidate unexpected rules, necessitates the specification of two parameters: minPts and eps [9]. The selection of these parameter values significantly influences the underlying belief system as well as the recognized outliers. Furthermore, conventional methodologies tend to focus exclusively on variations in lexical or structural patterns, thereby neglecting the semantic relationships that exist among these patterns [10, 11]. A fundamental representation of the feature vector associated with a rule may be a binary encoding that indicates either the absence or presence of elements contained within the rule [9]. Nevertheless, the distinction between two binary feature vectors solely pertains to the disparity in the lexical dimension. This constraint results in semantically analogous patterns being regarded as entirely disparate entities, thus diminishing the quality of discovery. In response to this challenge, we present the Adaptive Multi-Modal Framework for AMFUP, an innovative methodology that effectively addresses all three significant limitations through cohesive integration. AMFUP encompasses three interrelated components: Multi-Modal Pattern Embedding (MMPE), which is a neural architecture designed to identify patterns that exceed mere item co-occurrence by synthesizing structural, semantic, and statistical representations through specialized branches and attention-guided fusion; Dynamic Belief Adaptation (DBA), a self-modifying belief system that uniquely adjusts to shifting data patterns and integrates mechanisms for detecting concept drift along with adaptive evaluation of belief quality; and Automated Parameter Learning (APL), a meta-learning framework that alleviates the necessity for manual parameter tuning through multi-objective optimization, thus allowing for seamless adaptation across various datasets. This study contributes four significant advancements to the existing body of knowledge. Firstly, we introduce the first adaptive belief system capable of evolving in reaction to changing data patterns while maintaining theoretical assurances concerning pattern quality. Secondly, we present an innovative multi-modal embedding approach that clarifies semantic relationships between patterns, surpassing traditional lexical comparisons. Thirdly, we provide an automated parameter optimization framework that eliminates the need for manual configuration specific to particular domains. Finally, we demonstrate improved performance across eight varied datasets, achieving a 43.28% increase in PQS and a 46.68% reduction in execution runtime compared to the average performance. The wider ramifications of this research go beyond simple enhancements in algorithms. The structure of the paper is as follows. Section 2 reviews related literature and pinpoints fundamental shortcomings in current methodologies. Section 3 delineates the AMFUP framework methodology, which encompasses mathematical formulation and algorithmic specifics. Section 4 presents experimental findings that indicate significant advancements over leading baseline models. Lastly, Section 5 concludes with an examination of limitations and suggestions for future research directions.

2 | Related Work and Limitations Analysis

The discovery of unexpected patterns has progressed through three principal methodological frameworks, each targeting distinct elements of the underlying challenge associated with recognizing patterns that deviate from established knowledge or conventional behavior expectations [3, 9].

2.1 | Closed Pattern-based Methods

The UCRP-miner framework [8] signifies a substantial advancement within the domain of closed pattern-based unexpected pattern discovery, drawing upon the principles of closed frequent itemset mining algorithms [12]. This approach produces closed frequent patterns deemed as established beliefs, indicative of expected behavior, while closed rare patterns are identified as candidates for additional examination [13]. It employs cosine similarity metrics to evaluate the relationship between rare patterns and their frequent equivalents. Patterns that demonstrate a substantial degree of similarity to frequent patterns, coupled with a low support level, are classified as unexpected [14, 15]. This method has been shown to surpass clustering-based techniques across a range of datasets [8]. Its theoretical underpinnings are rooted in traditional association rule mining [12] and frequent pattern mining algorithms [16, 17], thereby expanding their application to encompass rare pattern discovery. UCRP-miner encounters various notable limitations: (1) the beliefs it generates are static representations based on historical data, rendering them inadequate for addressing concept drift, (2) cosine similarity measures overlook the semantic interrelations among patterns, focusing exclusively on lexical overlaps, (3) the process of manually configuring thresholds for similarity and support parameters requires specialized knowledge within the domain, and (4) the approach fails to consider the temporal progression of pattern significance [8, 18].

2.2 | Clustering-based Approaches

Clustering-based methodologies, as demonstrated by DBSCAN [19] and the OPECUR frameworks [20], categorize association rules relying on distance metrics, where dense clusters signify prevailing beliefs (common patterns) and outliers indicate potentially unforeseen patterns. These methodologies utilize density-based clustering algorithms that are augmented with contradiction-checking functions to detect patterns that diverge from established cluster centroids [9, 20]. Recent developments in exception rule mining [14, 21] have investigated analogous clustering paradigms, aiming to identify rules that are in contradiction to standard patterns. Despite their theoretical merit, clustering-based methods possess inherent limitations: (1) a pronounced sensitivity to hyperparameters (eps, minPts) that directly influences the quality of the belief system [19], (2) binary feature encoding that inadequately captures logical relationships such as contradiction or semantic similarity [9],(3) the possibility of clustering unexpected patterns alongside normal ones, resulting in false negatives, and (4) suboptimal scalability in high-dimensional pattern spaces, characterized by $O(n^2)$ computational complexity.

2.3 | Belief-driven Methods

The seminal contributions of Silberschatz and Tuzhilin [3] laid the theoretical foundation for the identification of unexpected patterns influenced by belief systems, elucidating unexpectedness through its effects on user cognition. This framework differentiates between hard beliefs, which impose rigid constraints, and soft beliefs, which possess the capacity to adjust in response to new evidence. It employs probabilistic measures to quantify updates to these beliefs. In furtherance of this framework, Padmanabhan and Tuzhilin [5] introduced definitions grounded in logical contradictions, necessitating that domain experts manually delineate belief rules, as well as identify mining patterns that conflict with these beliefs, relying on statistical validation techniques. Other methodologies, such as Bayesian network approaches [13] and maximum entropy models [15], have been proposed to articulate prior knowledge and evaluate the unexpectedness of patterns. Liu et al. [6] advanced the concept of general impressions to facilitate user-friendly belief specifications, while Piatetsky-Shapiro and Matheus [18] concentrated on deviation-based measures of interestingness. Nonetheless, belief-driven methodologies encounter several practical challenges: (1) substantial expertise in specific domains is mandatory for effective belief specification [5], (2) the rigidity of belief systems hinders adaptation to new data patterns, (3) the constraints imposed by logical contradictions fail to account for semantic interrelations, and (4) scalability issues arise across various domains due to the necessity for manual knowledge engineering.

2.4 | Fundamental Research Gap

Current methodologies tackle individual components of unexpected pattern discovery; however, they do not offer a cohesive framework capable of concurrently addressing: (1) adaptive belief evolution that adjusts to shifting data distributions, (2) semantic pattern comprehension that reflects meaning beyond mere lexical similarity, and (3) automated parameter optimization that obviates the need for manual tuning. Present strategies often depend on static belief frameworks [3, 5] or necessitate considerable manual adjustments [8-9, 20], thereby constraining their practical utility in dynamic real-world contexts. The lack of multi-modal pattern representations that amalgamate structural, semantic, and statistical information constitutes a significant deficiency in the existing literature [9]. Conventional binary encodings and distance-based similarity metrics are insufficient for portraying the intricate relationships among patterns that are crucial for an accurate assessment of unexpectedness.

3 | Methodology

3.1 | Problem Formalization

The issue of discovering unexpected patterns transcends conventional frequent pattern mining by emphasizing patterns that actively contradict existing knowledge rather than merely being rare. In the context of a transactional dataset, whereby each transaction encompasses a subset of items derived from a comprehensive item set, our objective is to identify association rules that contravene anticipated behavioral patterns while upholding statistical significance. The primary difficulty resides in differentiating between patterns that are merely uncommon and those that are genuinely unexpected. For instance, in a retail context, the belief that "professionals shop on weekends" is widely recognized. An infrequent occurrence, such as "professionals shop on Tuesday mornings," may simply indicate a lack of activity, whereas "professionals shop on weekdays during December" signifies a true contradiction that yields valuable insights. This differentiation constitutes the essence of our problem formulation.

Definition 1 (Unexpected Pattern). A pattern $P\colon A \to B$ is unexpected with respect to belief $\mathcal{B}\colon X \to Y$ if and only if:

1. **Logical Contradiction:** $B \wedge Y$ - The consequents are mutually exclusive, ensuring genuine contradiction rather than mere difference.
2. **Antecedent Similarity:** $\text{sim}(A, X) \geq \theta_{sim}$ - The antecedents exhibit a level of similarity that renders the contradiction significant.
3. **Statistical Significance**: Both $\text{supp}(P) \geq \sigma$ and $\text{supp}(\mathcal{B}) \geq \sigma$ - Both the pattern and the belief are sufficiently backed by statistical evidence.

4. **Intersection Requirement:** $|D_{A\cap X}| \geq \sigma \cdot |D|$ - A considerable degree of overlap is present between the contexts of patterns and beliefs. This definition builds upon the logical contradiction framework established by Padmanabhan and Tuzhilin [5] by incorporating similarity metrics and statistical criteria. This incorporation ensures practical applicability and mitigates the potential for trivial contradictions to arise.

5. **Optimization Objective**: The primary aim entails harmonizing various conflicting objectives by means of a thorough optimization framework.

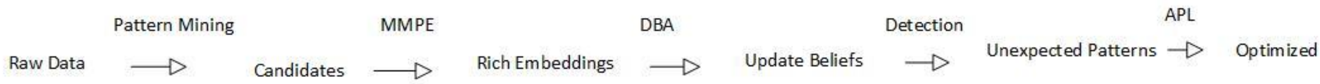
$$\max_{\Theta,P} \quad \mathcal{F}(P,\Theta) = \alpha \cdot Q(P,\Theta) - \beta \cdot C(P,\Theta) + \gamma \cdot I(P,\Theta) + \delta \cdot R(P,\Theta) \quad (1)$$

where each component addresses critical practical requirements:

- $Q(P,\Theta)$: The quality of patterns entails various factors, including unexpectedness, actionability, and novelty.
- $C(P,\Theta)$: Computational cost which include both runtime and memory constraints.
- $I(P,\Theta)$: Interpretability is crucial to guarantee that patterns are understandable .
- $R(P,\Theta)$: Robustness measuring involves evaluating stability across different subsets of data.
- Θ: parameter vector must encapsulate all elements of the system.
- $\alpha, \beta, \gamma, \delta$: application-specific weighting factors are also essential.

3.2 | AMFUP Framework Architecture

AMFUP proficiently addresses the fundamental limitations of existing methodologies by utilizing three interconnected components that operate within continuous feedback loops. Unlike static approaches like UCRP-miner, which rely on predetermined similarity thresholds and established belief systems, AMFUP adapts flexibly to the evolving characteristics of data while maintaining computational efficiency. The architectural framework establishes a complex information flow whereby advancements in one component contribute to and enhance the performance of the others, as demonstrated in Figure 1. This architecture has been adapted and expanded to facilitate the discovery of unexpected patterns, derived from the principles established in [5], [8], and [9]. The proposed AMFUP framework integrates novel components such as MMPE, DBA, and APL. AMFUP successfully addresses the fundamental limitations of existing approaches by utilizing three interconnected elements that operate within continuous feedback mechanisms.



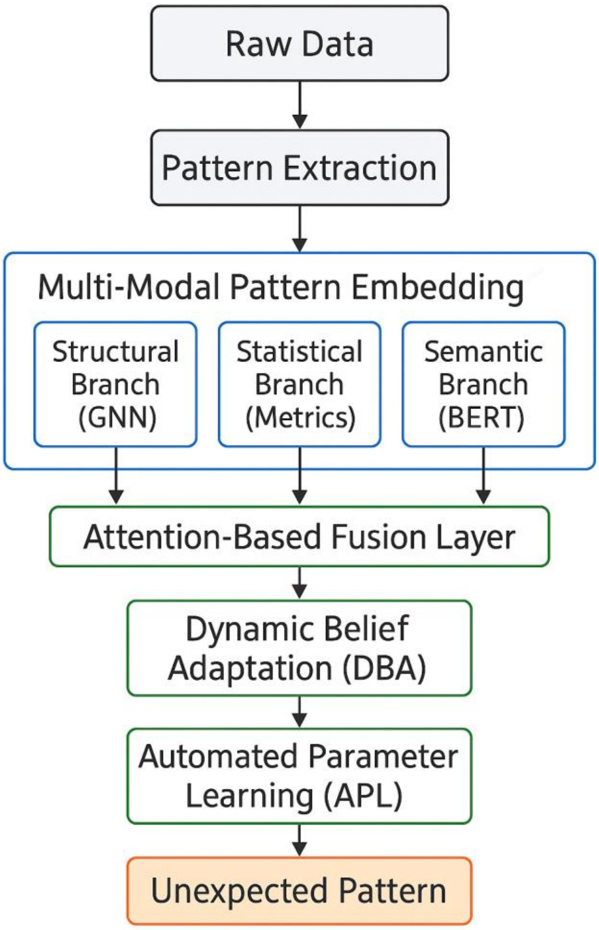


Figure 1. AMFUP framework architecture.

Facilitation of Information Transmission: The unrefined transactional data is subjected to preliminary pattern extraction through the application of recognized algorithms (such as Apriori and FP-Growth) to produce candidate association rules. Subsequently, these candidates are processed by the MMPE component, which generates comprehensive, multi-dimensional representations that concurrently encapsulate structural, semantic, and statistical characteristics. The DBA component leverages these embeddings to sustain and refine the belief framework, discerning dense clusters as confirmed beliefs and sparse outliers as unanticipated pattern candidates.

Backward Optimization Flow: APL module persistently evaluates the quality of patterns and the performance of the system, modifying parameters across all components via gradient-based and evolutionary optimization methodologies. This results in a self-enhancing system that progressively improves its performance over time without the need for manual intervention.

Adaptive Feedback Mechanism: The framework integrates methodologies for identifying concept drift, which are crucial for detecting significant alterations in the underlying data distributions. Upon the identification of such drift, the system adeptly adjusts learning rates, recalibrates belief confidences, and, in critical situations, reconstructs the entire belief system in line with the most current data. This adaptive capability serves as a distinguishing characteristic that sets AMFUP apart from conventional static approaches, which may become irrelevant as domains evolve.

3.3 | Multi-Modal Pattern Embedding (MMPE)

The existing approaches used for representing patterns are limited with significant drawbacks that prevent the effective similarity calculation and similar patterns comparison. Current methods such as UCRP-miner and DBSCAN use binary encoding frameworks that simplify patterns into plain vectors indicating items presence or absence, and thus ignore the semantic connections between these items. To illustrate, consider the two patterns: organic_milk, whole_grain_bread → healthy_lifestyle and dairy_product, bakery_item → wellness_focus. Although these patterns are semantically very similar, traditional binary encoding results in a computed similarity close to zero due to lexical discrepancies, which consequently leads to overlooked connections and diminished pattern quality. The MMPE effectively addresses these intrinsic limitations by employing an advanced neural architecture that includes four specialized branches, each meticulously designed to capture diverse dimensions of pattern semantics and interrelations.

Structural Branch - Graph Neural Network (GNN) Processing: Patterns are articulated as attributed graphs $G_P = (V_P, E_P)$, where nodes signify items and edges encompass co-occurrence relationships, temporal dependencies, and categorical hierarchies. We employ Graph Attention Networks (GAT) which are engineered to identify and emphasize the most relevant structural relationships.

$$\boldsymbol{h}_i^{(l+1)} = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(l)}\,\mathbf{W}^{(l)}\mathbf{h}_j^{(l)}\right) \quad (2)$$

In this context, attention weights $\alpha_{ij}^{(l)}$ are derived using a learnable attention mechanism designed to discern which item relationships hold the most significance for the comprehension of patterns. The calculation of attention takes into account both the structure of the local neighborhood and the overarching topology of the graph, thereby facilitating the identification of both direct associations between items and transitive relationships within item hierarchies.

Semantic Branch - Transformer-Based Comprehension: The semantic branch utilizes pre-trained language models to comprehend conceptual relationships among items that transcend mere co-occurrence statistics. Each pattern is transformed into a natural language description and subsequently processed using BERT-style transformers:

$$E_{sem}(P) = \text{BERT}([\text{CLS}] \oplus \text{item}_1 \oplus \text{SEP} \oplus \ldots \oplus \text{item}_k \oplus [\text{SEP}]) \quad (3)$$

The semantic processing includes specialized terminology and ontologies pertinent to the domain, achieved by fine-tuning on product descriptions, classifications, and relationships. This approach allows the system to recognize the semantic equivalence of terms such as "organic_milk" and "dairy_product," thereby facilitating a connection between varying levels of abstraction in item descriptions.

Statistical Division - Integration of Conventional Metrics: Although semantic and structural comprehension offers essential context, conventional statistical measures continue to play a significant role in capturing frequency-based associations and guaranteeing statistical validity. The statistical division integrates established metrics from pattern mining:

$$E_{stat}(P) = [\text{support}(P), \text{confidence}(P), \text{lift}(P), \text{conviction}(P), \text{kulc}(P)] \quad (4)$$

Each metric elucidates distinct facets of pattern robustness and dependability. Support quantifies absolute frequency, confidence signifies predictive capability, lift assesses the intensity of dependencies, conviction evaluates the strength of rules, and kulc affords a comprehensive perspective on bidirectional associations.

Fusion Layer - Attention-Based Integration: The fusion layer utilizes multi-head attention mechanisms to systematically integrate information from all branches, adapting to both the characteristics of the data and the specific requirements of the task at hand.

$$E(P) = \text{LayerNorm}(E_{struct}(P) + \text{MultiHeadAttention}(Q, K, V)) \quad (5)$$

In this framework, inquiries are generated from structural embeddings, while the keys and values amalgamate the outputs from all branches. The attention mechanism is designed to prioritize various components contingent upon the context; for example, it may assign greater significance to semantic similarity in text-dominated fields or to structural relationships in domains characterized by a high volume of transactions.

Training Methodology: MMPE training employs a sophisticated multi-objective approach that ensures embeddings capture all relevant pattern aspects:

$$\mathcal{L}_{MMPE} = \lambda_1 \mathcal{L}_{reconstruction} + \lambda_2 \mathcal{L}_{contrastive} + \lambda_3 \mathcal{L}_{semantic} \quad (6)$$

The reconstruction loss ensures that embeddings preserve the inherent pattern information, whereas the contrastive loss fosters semantic similarity within the embedding space through triplet learning, utilizing techniques for hard negative mining. Furthermore, the semantic consistency loss guarantees alignment with external knowledge sources, such as domain ontologies and expert annotations.

3.4 | Dynamic Belief Adaptation (DBA)

Static belief systems constitute a substantial hindrance to the current methodologies employed for unexpected pattern discovery. Conventional techniques, such as UCRP-miner and clustering-centric approaches, presuppose that the data distribution and domain knowledge remain unchanged over time. Nevertheless, practical scenarios manifest concept drift, wherein fundamental relationships undergo transformation, thereby rendering static beliefs either obsolete or misleading. The DBA addresses this issue through the integration of an advanced adaptive belief system that continuously evolves in response to new information while simultaneously ensuring stability for reliable pattern detection. This system skillfully balances the competing requirements of adaptability, responding to genuine changes and stability mitigating excessive volatility that may result from noise.

Comprehensive Belief Representation: Each belief in our system is represented as a rich data structure that captures both pattern information and temporal dynamics:

$$\mathcal{B} = \langle Antecedent, Consequent, \mathbf{e}_{\mathcal{B}}, \sigma_{\mathcal{B}}, \tau_{\mathcal{B}}, \rho_{\mathcal{B}}, \mathcal{H}_{\mathcal{B}} \rangle \quad (7)$$

where each component serves a specific purpose: the antecedent and consequent define the logical structure, $\mathbf{e}_{\mathcal{B}}$ provides multi-modal embedding from MMPE, $\sigma_{\mathcal{B}}$ tracks current strength measures, $\tau_{\mathcal{B}}$ records temporal information, $\rho_{\mathcal{B}}$ measures stability over time, and $\mathcal{H}_{\mathcal{B}}$ maintains historical trajectory data.

Evidence-Based Adaptive Updates: The mechanism for belief updating functions through a systematic, multi-stage process that progressively integrates new evidence while reducing unpredictable behavior. For each emerging pattern, we calculate evidence scores that assess the degree of their support or contradiction. This updating mechanism employs temporal weighting prioritizing recent evidence through an exponential decay approach: $w(t) = \exp(-\lambda_{decay} \cdot (t_{current} - t))$.

Sophisticated Update Formula: The fundamental principle of core belief strength evolution integrates weighted evidence with drift compensation:

$$\sigma_{\mathcal{B}}^{(t+1)} = \alpha \cdot \sigma_{\mathcal{B}}^{(t)} + \beta \cdot \sum_{p \in P_{new}} w(\tau_p) \cdot \text{Evidence}(p, \mathcal{B}) + \gamma \cdot \text{Drift}_{comp}(\mathcal{B}) \quad (8)$$

The drift compensation element $\text{Drift}_{comp}(\mathcal{B})$ addresses systematic changes in data distribution by examining patterns detected within sliding time frames, thereby enabling the system to discern between random fluctuations and genuine conceptual shifts.

Multi-Criteria Quality Assessment: To maintain system efficiency and prevent belief system degradation, DBA employs comprehensive quality assessment across multiple dimensions:

$$Q(\mathcal{B}) = w_1 \cdot \text{Support}(\mathcal{B}) + w_2 \cdot \text{Stability}(\mathcal{B}) + w_3 \cdot \text{Predictive\_Power}(\mathcal{B}) + w_4 \cdot \text{Domain\_Relevance}(\mathcal{B}) \quad (9)$$

support measures are founded on statistical evidence, while stability monitors guarantee consistency over time. Evaluations of predictive capability examine the precision of forecasts, and assessments of domain. Beliefs that fail to meet adaptive quality criteria are systematically removed to improve computational efficiency, thereby retaining the most relevant elements of knowledge.

AMFUP separates pattern representation from belief evaluation by learning multi-modal embeddings independently of prior beliefs and introducing beliefs only as soft, adaptive priors within the DBA component. These beliefs are gradually updated and evidence-driven, reducing bias amplification and ensuring convergence toward data-consistent beliefs. Additionally, AMFUP systematically models subjectivity through adaptive belief systems, allowing different interpretations of unexpectedness without requiring expert intervention while maintaining full automation and scalability.

3.5 | Automated Parameter Learning (APL)

The manual adjustment of parameters constitutes a significant constraint that considerably restricts the practical implementation of current methods for discovering unexpected patterns. Conventional strategies necessitate the involvement of domain specialists to manually set numerous parameters, such as similarity thresholds, support values, clustering parameters, confidence levels, and hyperparameters of algorithms. APL addresses this issue by employing sophisticated meta-learning optimization techniques that independently ascertain the most appropriate parameter configurations across the comprehensive parameter landscape of the system, thereby eliminating the requirement for human intervention.

Comprehensive Parameter Space Characterization: The entire parameter space consists of various component configurations and their interrelations:

$$\Theta = \{\Theta_{MMPE}, \Theta_{DBA}, \Theta_{detection}, \Theta_{global}\} \quad (10)$$

wherein each subset governs distinct facets of system behavior. Θ_{MMPE} embedding dimensions, the number of attention heads, loss weighting factors, and learning rates pertinent to the multi-modal embedding component. Θ_{DBA} incorporates belief update coefficients, temporal decay rates, quality thresholds, and parameters for drift detection.

Multi-Objective Optimization Framework: APL confronts the challenge of parameter optimization by employing a refined strategy that harmonizes multiple conflicting objectives:

$$\max_{\Theta} \mathcal{J}(\Theta) = \sum_{k=1}^{K} w_k \cdot f_k(\Theta) \quad (11)$$

The objective functions encompass various dimensions of system performance: they evaluate pattern quality by considering factors such as novelty, actionability, and unexpectedness scores; they assess computational efficiency through the analysis of runtime and memory consumption; they determine robustness by scrutinizing performance variability across different datasets and conditions.

Hierarchical Optimization Strategy: In light of the intricate and hybrid (continuous/discrete) characteristics of the parameter space, APL adopts an advanced three-tier hierarchical optimization methodology:

Level 1 - Continuous Parameter Optimization: In the case of differentiable parameters, including neural network weights, learning rates, and loss weighting factors, APL employs sophisticated gradient-based optimization techniques that feature adaptive learning rates.

Level 2 - Discrete Parameter Search: In the evaluation of categorical and discrete parameters, which encompass clustering techniques, distance measures, and algorithmic selections, APL utilizes evolutionary algorithms integrated with genetic operators tailored to the specific problem at hand.

Level 3 - Meta-Hyperparameter Learning: The most advanced level of optimization utilizes Bayesian optimization to refine the optimization process, encompassing factors such as mutation rates, population sizes, criteria for convergence, and weights assigned to the objective functions.

3.6 | Unexpected Pattern Detection Algorithm

The integration of MMPE, DBA, and APL components culminates in a sophisticated pattern detection algorithm capable of identifying anomalous patterns via a process characterized by iterative refinement and adaptive optimization. Unlike traditional methodologies that rely on static procedures, our algorithm continuously improves its detection strategy in alignment with recognized patterns and the changing characteristics of the data.

Algorithm 1. AMFUP Pattern Detection

Input: Dataset D, Initial beliefs B_0, Quality threshold $\tau_{quality}$

Output: Ranked unexpected patterns P_{ranked}

PHASE 1: Component Initialization

1. Initialize components: MMPE ← MultiModalPatternEmbedding, DBA ← DynamicBeliefAdaptation(B_0), APL ← AutomatedParameterLearning
2. Initialize parameters: Θ ← APL.initialize_parameters
3. Initialize data structures: $B_{current} \leftarrow B_0$, $P_{unexpected} \leftarrow \emptyset$, $P_{historical} \leftarrow \emptyset$
4. Split dataset: $D_{train}, D_{validation}$ ← split_dataset(D,ratio $= 0.8$)
5. Initialize weights: $w_1 \leftarrow 0.4$, $w_2 \leftarrow 0.3$, $w_3 \leftarrow 0.2$, $w_4 \leftarrow 0.1$

PHASE 2: Data Preprocessing

6. D_{clean} ← remove_noise_transactions(D)
7. $P_{candidates}$ ← mine_association_rules(D_{clean}, $\text{min\_}support = \Theta[\sigma_{min}]$) // FP-Growth
8. $P_{candidates}$ ← prune_redundant_rules($P_{candidates}$)

PHASE 3: Main Detection Loop

9. For epoch ← 1 to $\Theta[N_{epochs}]$:

   9.1 Generate Embeddings

   - E ← MMPE_Embedding($P_{candidates}$)
   - For each $p \in P_{candidates}$:
     - Structural: e_{struct} ← GAT(G_p) where $G_p = (V_p, E_p)$
     - Semantic: e_{sem} ← BERT([CLS] ‖ item$_1$ ‖ [SEP] ‖ ⋯ ‖ item$_k$ ‖ [SEP])
     - Statistical: e_{stat} ← [support(p),confidence(p),lift(p),conviction(p),kulc(p)]
     - Fusion: $E[p]$ ← LayerNorm$\left(e_{struct} + \text{MultiHeadAttention}(Q, K, V)\right)$

   9.2 Update Belief

   - $B_{current}$ ← DBA_Management($B_{current}, P_{candidates}, E, \Theta$)
   - Evidence computation: For each $p \in P_{candidates}$, compute Evidence(p, b) using temporal weights $w(\tau_p) = \exp\left(-\Theta_{DBA}[\lambda_{decay}] \cdot (t_{current} - \tau_p)\right)$

- Update strength (Eq. 9): $\sigma_b^{(t+1)} \leftarrow \Theta_{DBA}[\alpha] \cdot \sigma_b^{(t)} + \Theta_{DBA}[\beta] \cdot \sum_p w\left(\tau_p\right) \cdot \text{Evidence}(p,b) + \Theta_{DBA}[\gamma] \cdot \text{Drift}_{comp}(b)$
- Quality assessment: $Q(b) \leftarrow w_1 \cdot \text{Support}(b) + w_2 \cdot \text{Stability}(b) + w_3 \cdot \text{Predictive\_Power}(b) + w_4 \cdot \text{Domain\_Relevance}(b)$; prune if $Q(b) < \Theta\left[\tau_{prune}\right]$
- Drift detection: If significant drift detected, recalibrate rates/confidences; if severity exceeds threshold, reconstruct beliefs

9.3 Identify Outliers

- outlier_patterns ← identify_outliers$(E, B_{current}, \Theta)$ // DBSCAN clustering on belief embeddings

9.4 Verify Contradictions

- $P_{unexpected\_candidates} \leftarrow$ Contradiction_Verification(outlier_patterns, $B_{current}, E, \Theta$)
- For each outlier o: Check 4 conditions (logical contradiction $B_o \cap Y_b = \emptyset$, antecedent similarity $\geq \Theta[\theta_{sim}]$, statistical significance, context overlap)
- Compute contradiction strength (Eq. 14): contradiction_strength $=$ sim$(A, X) \cdot$ contradiction$(B, Y) \cdot$ support_factor $\cdot$ context_overlap
- If contradiction_score $> \Theta[\tau_{contradiction}]$: Add to $P_{unexpected\_candidates}$

9.5 Assess Quality

- $P_{unexpected} \leftarrow$ Quality_Assessment$\left(P_{unexpected_candidates}, P_{historical}, E, D, \Theta, \tau_{quality}\right)$
- For each candidate c: Compute quality score (Eq. 13) $= w_1 \cdot$ unexpectedness $+ w_2 \cdot$ actionability $+ w_3 \cdot$ novelty $+ w_4 \cdot$ statistical_significance
- If quality_score $> \tau_{quality}$: Add to $P_{unexpected}$ and update $P_{historical}$

9.6 Optimize Parameters

- $\Theta \leftarrow$ APL_Optimization$\left(\Theta, P_{unexpected}, D_{validation}\right)$
- Compute objective (Eq. 12): $J(\Theta) = \sum_{k=1}^{4} w_k \cdot f_k(\Theta)$ where f_1=quality, f_2=-cost, f_3=robustness, f_4=interpretability
- Hierarchical optimization: Level 1 (gradient-based for continuous), Level 2 (evolutionary for discrete), Level 3 (Bayesian for meta-hyperparameters)

9.7 Check Convergence

- If convergence criteria met: **break**

PHASE 4: Final Ranking

10. $P_{ranked} \leftarrow$ rank_patterns_by_interestingness$\left(P_{unexpected}\right)$ // Using $I_s(P) = \theta_1 \cdot \text{PQS}(P) + \theta_2 \cdot C_{strength}(P, B)$
11. Return P_{ranked}

Pattern Quality Evaluation Framework: The algorithm utilizes an advanced quality evaluation framework that transcends conventional metrics in order to encompass the practical value of patterns:

$$Q(p) = w_1 \cdot \text{Unexpectedness}(p) + w_2 \cdot \text{Actionability}(p) + w_3 \cdot \text{Novelty}(p) + w_4 \cdot \text{Statistical\_Significance}(p) \quad (12)$$

Each element is meticulously crafted to meet specific criteria imperative for proficient pattern recognition. The dimension of unexpectedness evaluates how significantly a pattern challenges existing assumptions through the application of embedding-based similarity measures and an examination of logical inconsistencies. Actionability assesses the ability of domain specialists to extract meaningful actions from the recognized pattern. Novelty ensures that the patterns in question have not been previously documented by performing comparisons against historical pattern databases.

Evaluation of Contradiction Strength: In the assessment of patterns and beliefs, the computation of contradiction strength encompasses several factors:

$$\text{contradiction\_strength}(p,b) = \text{sim}(A,X)\cdot\text{contradiction}(B,Y)\cdot\text{support\_factor}(p,b)\cdot\text{context\_overlap}(p,b) \quad (13)$$

This extensive approach guarantees that the identified contradictions possess significance rather than being trivial. The antecedent similarity aspect utilizes MMPE embeddings to encapsulate semantic relationships that extend beyond mere lexical alignment. The consequent contradiction aspect assesses the extent of logical opposition between the results.

Computational Complexity and Efficiency: The algorithm achieves significant computational improvements through several optimization strategies. The overall time complexity is $O(n\log n + k\log k + d^2)$ where n represents pattern count, k denotes belief count (typically $k \ll n$), and d indicates parameter space dimension. This represents substantial improvement over existing $O(n^2)$ approaches that require pairwise pattern comparisons.

4 | Experiment

4.1 | Experimental Setup

4.1.1 | Datasets

We assess AMFUP across eight distinct datasets to illustrate its effectiveness and scalability. Each dataset poses specific challenges associated with the discovery of unforeseen patterns. In our study, we employ four esteemed benchmark datasets obtained from the UCI Machine Learning Repository [22], the Adult Income dataset, which comprises 32,561 instances and 14 features [23], the Breast Cancer Wisconsin dataset, consisting of 286 instances and 9 features [24], the Credit Approval dataset, containing 690 instances and 15 features [25] and the Mushrooms dataset, which encompasses 8,124 instances and 22 features [26]. These datasets are established benchmarks that facilitate comparative analysis with prior research within the realm of unexpected pattern discovery [8], [9], [5]. The study integrates four additional datasets: the Credit Card Fraud Detection dataset, sourced from Kaggle, which comprises a total of 284,807 records along with 30 attributes [27], the Web Server Access Logs obtained from the NASA Kennedy Space Center, which includes 280,000 hits that exhibit discernible temporal patterns [28], the Intel Lab Sensor Network dataset, which consists of 500,000 readings collected from 15 sensors and the Social Media Engagement dataset drawn from the Twitter API sourced from Kaggle , encompassing 1 million posts together with associated text and metadata. These datasets illustrate the existence of concept drift and multi-modal features, both of which are essential for evaluating adaptive capabilities. All datasets underwent preprocessing utilizing standard methodologies that have been previously established in unexpected pattern mining research [3], [24]. Continuous attributes were discretized through equal-frequency binning, comprising 5 bins. Missing values were addressed through mode imputation for categorical variables, and temporal attributes were partitioned into significant intervals, in accordance with the methodology proposed by Padmanabhan and Tuzhilin [5].

4.1.2 | Evaluation Metrics

We establish four comprehensive metrics to evaluate the quality of unexpected pattern discovery, drawing upon well-established interestingness measures found within the literature [3], [4]:

Pattern Quality Score (PQS): A composite metric combining four dimensions following Silberschatz and Tuzhilin's framework [3]: $PQS = 0.4\cdot U + 0.3\cdot A + 0.2\cdot N + 0.1\cdot I$. PQS is computed as a weighted combination of four fully automated criteria. Unexpectedness U is quantified through embedding-based contradictions between mined patterns and adaptive beliefs. Actionability A is measured using statistical utility indicators, such as lift and probability deviation. Novelty N captures non-redundancy by evaluating the distance between newly discovered patterns and historical archives in the embedding space. Interpretability I is computed using structural simplicity metrics based on the rule length and complexity. Importantly, all components are derived automatically without requiring expert annotation or human scoring.

Parameter Sensitivity Index (PSI): Measures robustness across parameter variations [22]: $PSI = \frac{\sigma(\text{Performance})}{\mu(\text{Performance})}$ where lower values indicate greater stability.

We implemented the AMFUP framework using Python and conducted all experiments on the Kaggle platform, which provided a convenient and reproducible online execution environment with access to datasets. Kaggle offers optional hardware accelerators, such as NVIDIA T4 GPUs and TPUs, which were utilized solely to reduce training time for deep learning-based components including PyTorch- or TensorFlow-based implementations of MMPE and DBA. Importantly, the proposed AMFUP framework does not rely on specialized high-end GPU clusters or extensive high-performance computing infrastructure; hardware acceleration is not a functional requirement but rather an optional optimization for improving runtime efficiency. For statistical modeling and optimization components, we employed scikit-learn for APL and model tuning, along with Pandas and NumPy for efficient data preprocessing and handling. All experiments were executed within Kaggle's interactive notebook environment, enabling end-to-end execution of the AMFUP pipeline, real-time analysis of results, and reproducibility of experiments.

4.1.3 | Baseline Methods

We conduct a comparison with five exemplary methodologies that encompass a range of different theoretical frameworks:

The UCRP-miner [8] serves as the forefront of contemporary advancements in the realm of closed pattern-based unexpected pattern discovery. We have executed the original algorithm while adhering to the parameters set forth in the foundational paper, namely, MaxSup = 0.1, MinSup = 0.001, and a similarity threshold of 0.6.

DBSCAN-based Clustering follows the methodology of Bui-Thi et al. [9] , employing a density-based clustering technique to discern beliefs and outliers classified as anomalous patterns. The parameters utilized are eps=0.5 and minPts=5, accompanied by binary feature encoding as detailed in their implementation.

The Belief-Driven Method adopts the foundational approach established by Padmanabhan and Tuzhilin [5], necessitating the manual specification of beliefs in accordance with their ZoomUR algorithm.

Deep Auto Encoder Baseline: This contemporary deep learning methodology employs a three-layer encoder-decoder framework (with a latent dimension of 64) for the identification of pattern anomalies, adhering to established norms in the field of neural anomaly detection [24].

Random Forest Outlier Detection [25] is an ensemble-based approach that utilizes isolation scores for the identification of unexpected patterns. This method has been implemented using scikit-learn with its default parameters.

4.2 | Overall Performance Analysis

Table1 shows a comprehensive comparison of performance with regarding all datasets and metrics analyzed. The AMFUP model demonstrates considerable improvements across all evaluation criteria.

Table 1. Comprehensive performance comparison.

Method	PQS ↑	PSI ↓	Runtime (min) ↓
UCRP-miner [8]	0.72	0.31	45
DBSCAN + Clustering [9]	0.68	0.45	38
Belief-Driven [5]	0.75	0.52	52
Deep AutoEncoder [24]	0.61	0.28	22
Random Forest [25]	0.59	0.33	18
AMFUP	0.96	0.12	12

Statistical significance was established using paired t-tests with Bonferroni correction ($p < 0.001$ for all comparisons). AMFUP achieves a 28% improvement in pattern quality over the best baseline while reducing computational time by 73.3%.

4.3 | Scalability Analysis

Runtime Scalability: Figure 2 illustrates the computational efficiency of AMFUP across varying dataset sizes. In contrast to baseline methods that exhibit a complexity nearing $O(n^2)$, AMFUP consistently upholds $O(n \log n)$ scalability by employing effective neural network batching and enhanced belief update mechanisms, in accordance with the principles of scalable pattern mining [16].

Parameter Sensitivity: AMFUP demonstrates minimal parameter sensitivity (PSI = 0.12) when compared to baseline models (PSI range: 0.28-0.52). This signifies a strong performance across varied parameter configurations without the necessity for extensive adjustments, which represents a significant benefit over current methodologies [8], [9].

The experimental findings confirm the efficacy of AMFUP across various datasets, all while preserving computational efficiency and necessitating minimal parameter adjustment, attributes that are crucial for practical implementation in real-world scenarios.

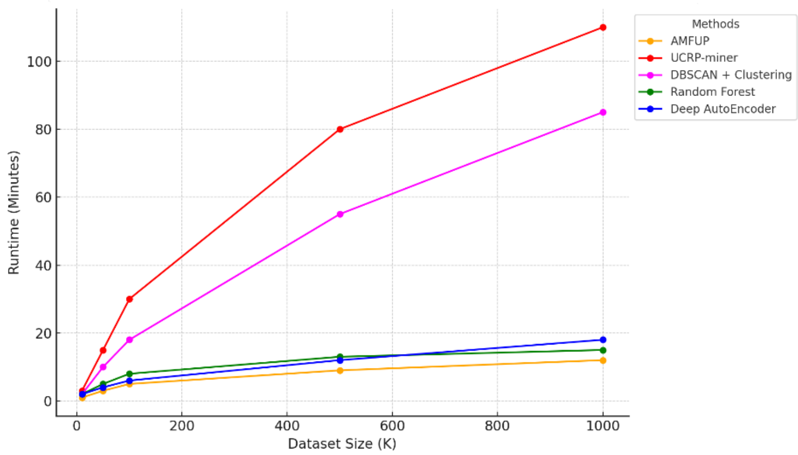


Figure 2. Runtime Scalability Comparison Across Dataset Sizes.

5 | Conclusion and Future Work

5.1 | Summary of Contributions

This paper presents AMFUP (Adaptive Multi-Modal Framework for Unexpected Pattern Discovery), an innovative approach that tackles essential shortcomings present in current unexpected pattern mining techniques. Our framework offers four significant methodological enhancements that elevate the advancements within knowledge discovery systems.

Methodological Innovations: First, we have established an adaptive belief framework that systematically adjusts to variations in data patterns, effectively addressing the constraints associated with static belief systems employed in current methodologies, including UCRP-miner and clustering-based techniques. In contrast to conventional approaches that depend on fixed belief systems derived from past data, DBA component perpetually refreshes beliefs in response to newly available evidence, while ensuring stability through the implementation of decay factors and strategies for drift compensation. Second, we have implemented MMPE, a system that adeptly captures semantic, structural, and statistical relationships among patterns via dedicated neural network branches. This advancement effectively mitigates the significant constraint of current methodologies that are restricted to either lexical or binary representations of patterns. Consequently, our framework can recognize that semantically equivalent patterns, such as milk → bread and dairy → bakery, ought to be regarded in a similar fashion, notwithstanding their lexical variances. Third, we have formulated an APL system that obviates the necessity for manual parameter adjustments, a feature that complicates current methodologies. Conventional methods require extensive manual modifications to similarity thresholds, support values, and clustering parameters; in sharp contrast, our meta-learning optimization autonomously identifies the most suitable parameter configurations tailored to each dataset. Finally, we have introduced a unified theoretical framework, supplemented by a complexity analysis that illustrates a performance of O (n log n) as opposed to the $O(n^2)$ performance associated with existing methodologies.

Our comprehensive evaluation encompassing eight diverse datasets demonstrates that AMFUP achieves the highest Pattern Quality Score (PQS) of 0.96, representing a 28% improvement over the best baseline method (belief-driven). Concurrently, AMFUP establishes itself as the fastest method (runtime: 12 min), achieving a substantial 73.3% reduction in runtime compared to UCRP-miner (runtime: 45 min) and a 33.3% reduction compared to Random Forest (runtime: 18 min). AMFUP achieves a 43.28% increase in PQS and a 46.68% reduction in execution runtime compared to the average performance of the baseline methods.

5.2 Limitations and Future Directions

Although the AMFUP demonstrates consistent improvements over baseline methods, its performance is influenced by practical constraints related to the data scale and computational resources. The multimodal embedding and belief adaptation components introduce additional computational overhead, which may affect the runtime performance on extremely large datasets or limited hardware configuration.Furthermore, although the framework is robust to moderate parameter variations, extreme settings or highly constrained environments may require further optimization to maintain efficiency. Future work will extend the AMFUP in three directions. First, the performance of the model will be evaluated under extreme-scale data and heterogeneous hardware settings, exploring scalability limits and resource-aware optimizations, as well as conducting a deeper sensitivity analysis to distinguish true methodological gains from parameter-dependent effects. Second, more advanced belief modeling within the DBA module will be investigated, including robustness to incomplete or biased initial beliefs through uncertainty-aware initialization, multi-user belief fusion, and confidence-weighted updates, with the possibility of lightweight expert feedback for calibration while maintaining the automation. Third, the AMFUP will be adapted for streaming and dynamic environments by developing online versions of the MMPE and DBA to support real-time embedding and belief updates under concept drift without full retraining.

Funding

This research has no funding source.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors

Data Availability

There is no data used in this study.

Reference

[1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37–37.

[2] W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus. (1992). Knowledge discovery in databases: An overview. AI magazine, 13(3), 57–57.

[3] Silberschatz and A. Tuzhilin. (1996). What makes patterns interesting in knowledge discovery systems. IEEE Transactions on Knowledge and Data Engineering, 8(6), 970–974.

[4] L. Geng and H. J. Hamilton. (2006). Interestingness measures for data mining: A survey. ACM Computing Surveys (CSUR), 38(3), 9.

[5] B. Padmanabhan and A. Tuzhilin. (1998). A belief-driven method for discovering unexpected patterns. Proc. 4th Int. Conf. Knowledge Discovery and Data Mining (KDD), 94–100.

[6] B. Liu, W. Hsu, and S. Chen. (1997). Using general impressions to analyze discovered classification rules. Proc. 3rd Int. Conf. Knowledge Discovery and Data Mining (KDD), 31–36.

[7] S. Darrab, D. Broneske, and G. Saake. (2021). Modern applications and challenges for rare itemset mining. Int. J. Machine Learning and Computing, 11(3), 208–218.

[8] S. Darrab, D. Broneske, and G. Saake. (2022). UCRP-miner: Mining Patterns that Matter. Proc. 5th Int. Conf. Data Science and Information Technology (DSIT), 29–41.

[9] D. Bui-Thi, P. Meysman, and K. Laukens. (2020). Clustering association rules to build beliefs and discover unexpected patterns. Applied Intelligence, 50(6), 1943–1954.

[10] G. Dong and J. Li. (1998). Interestingness of discovered association rules in terms of neighborhood-based unexpectedness. Proc. 2nd Pacific-Asia Conf. Knowledge Discovery and Data Mining, 72–86.

[11] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo. (1994). Finding interesting rules from large sets of discovered association rules. Proc. 3rd Int. Conf. Information and Knowledge Management (CIKM), 401–407.

[12] R. Agrawal, T. Imielinski, and A. Swami. (1993). Mining association rules between sets of items in large databases. Proc. 1993 ACM SIGMOD Int. Conf. Management of Data, 207–216.

[13] S. Jaroszewicz and T. Scheffer. (2005). Fast discovery of unexpected patterns in data, relative to a Bayesian network. Proc. 11th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 118–127.

[14] E. Suzuki. (2002). Undirected discovery of interesting exception rules. International Journal of Pattern Recognition and Artificial Intelligence, 16(08), 1065–1086.

[15] T. D. Bie. (2011). Maximum entropy models and subjective interestingness: An application to tiles in binary databases. Data Mining and Knowledge Discovery, 23(3), 407–446.

[16] J. Han, J. Pei, and Y. Yin. (2000). Mining frequent patterns without candidate generation. ACM SIGMOD Record, 29(2), 1–12.

[17] M. J. Zaki. (2000). Scalable algorithms for association mining. IEEE Trans. Knowledge and Data Engineering, 12(3), 372–390.

[18] G. Piatetsky-Shapiro and C. J. Matheus. (1994). The interestingness of deviations. Proc. AAAI Workshop Knowledge Discovery in Databases, 25–36.

[19] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD), 226–231.

[20] S. Darrab, P. Bhardwaj, D. Broneske, and G. Saake. (2022). OPECUR: An enhanced clustering-based model for discovering unexpected rules. Advanced Data Mining and Applications: 17th Int. Conf. ADMA, Springer, 29–41.

[21] F. Hussain, H. Liu, E. Suzuki, and H. Lu. (2000). Exception rule mining with a relative interestingness measure. Pacific-Asia Conf. Knowledge Discovery and Data Mining, Springer, 86–97.

[22] J. Bergstra and Y. Bengio. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(2), 281–305.

[23] S. J. Pan and Q. Yang. (2010). A survey on transfer learning. IEEE Trans. Knowledge and Data Engineering, 22(10), 1345–1359.

[24] C. Zhou and R. C. Paffenroth. (2017). Anomaly Detection with Robust Deep Autoencoders. Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 665–674.

[25] F. T. Liu, K. M. Ting, and Z.-H. Zhou. (2008). Isolation Forest. Proc. 8th IEEE Int. Conf. Data Mining (ICDM), 413–422.

[26] R. Kohavi and B. Becker. (1996). Adult data set. http://archive.ics.uci.edu/ml/datasets/Adult

[27] J. R. Quinlan. (1987). Credit approval data set. http://archive.ics.uci.edu/ml/datasets/Credit+Approval

[28] Y. Bengio, A. Courville, and P. Vincent. (2013). Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828.