

**CONTEXT-AWARE ASSOCIATION RULE MINING FOR
EFFECTIVE KNOWLEDGE EXTRACTION IN
CARDIOVASCULAR CLINICAL DECISION SUPPORT**

**Sura Mahmood Abdullah¹, Prof. Dr. Abbas Mohsin Al-
Bakry², Prof. Dr. Alaa K. Farhan³**

¹Iraqi Commission for Computers and Informatics/ University of Information
Technology and Communication

Iraq- Baghdad

phd202220702@iips.edu.iq

²University of Information Technology and Communication (UoITC)

Iraq- Baghdad

abbasm.albakry@uoitc.edu.iq

³Collage of Computer Sciences/ University of Technology-Iraq

Iraq- Baghdad

Alaa.K.Farhan@uotechnology.edu.iq

Abstract

Coronary artery disease (CAD) is a disorder that gradually develops due to the accumulation of fatty deposits within the arteries, which limits the blood supply to the heart. It is often not detected until serious complications occur, such as blockage or heart attacks. This paper proposes a framework for extracting medical association rules from patient data (facts file) using three advanced techniques—FP-Growth, Eclat, and HUIM—after cleaning, filtering, and converting the JSON fact file into a structured binary representation. FP-Growth provides efficient mining of frequent itemsets through the FP tree, while Eclat benefits from the intersections of transaction ID lists (TID-lists) within a depth-first search. Meanwhile, HUIM prioritizes "clinical utility" by focusing on rare high-impact features (such as significant biomarkers). The framework takes into account the contextual variability associated with regional factors and lifestyle in Iraq (for example, Basra vs. Baghdad). The resulting rules are filtered according to criteria: support, confidence, and lift to retain the strongest rules supporting the diagnosis. Eclat excels more than FP-Growth and HUIM; the average confidence is ~ 1.0 with near-zero variance, passing 100% of all confidence thresholds ($0.5 \rightarrow 1.0$), with an average support of ~ 0.39 and a moderate lift (~ 1.3). The impact of the differences from FP-Growth is significant (Cohen's $d \approx 1.6$), and the

ANOVA/Tukey results confirm its statistical superiority. These characteristics make it suitable as a verification/validation layer for reliable rules.

1. Introduction

The proliferation of medical data, especially in heart diseases such as coronary artery disease, has created a vital knowledge resource for healthcare institutions. Data mining allows for the transformation of this volume into practical knowledge by extracting patterns that support medical decision-making [1]. "Association rules" emerge as a fundamental tool for discovering relationships between clinical and demographic features and test results, such as the correlation of risk factors with coronary artery outcomes, with the ability to predict new links when subsequent data arrives [2]. Association rule mining is a data mining technique that aims to mine patterns and relationships between data sets in a large database. It usually consists of two main phases: generating frequent itemsets by identifying the sets that meet a minimum support threshold and generating association rules from the derived frequent itemsets that meet a minimum confidence threshold [3, 4].

This paper employed three algorithms for mining itemsets: the FP-Growth algorithm, which builds a frequent pattern tree (FP-tree) to derive those itemsets [5]; the Eclat algorithm, which uses a depth-first search approach to reach frequent itemsets [6]; and finally, the HUIM algorithm, which can be employed to discover combinations of items with high profitability [7]. Thus, the association rules are generated based on the extracted mining itemsets. Facts are usually extracted in the form of triples (entity, features, value) from unstructured medical texts, such as articles, reports, etc. These triples are considered the foundation for organizing data as mining rules and creating interconnected knowledge bases, which helps intelligent systems understand the relationships between entities more accurately and improve data analysis [8].

Relationships in the knowledge graph represent the connections between entities, and semantic descriptions clarify their characteristics and meanings. Features are assigned to nodes and edges in the "property graph." The knowledge graph is considered an organized representation of facts through the interaction of entities and relationships, resembling the term "knowledge base." Structurally, it consists of a network of nodes and links, and semantically, it enables the interpretation and inference of facts [9].

Healthcare systems face significant challenges in dealing with the complexity and variability of clinical data, which negatively impacts the early and accurate diagnosis of cardiovascular diseases. The gap lies in the reliance on static knowledge bases in expert systems that do not adapt to evolving medical evidence, in addition to machine learning techniques and large chatbots that may provide unreliable recommendations when lacking organized clinical knowledge. These systems also do not take into account the environmental and demographic differences between patients.

The main contribution of this paper is the creation of a framework for extracting context-aware association rules, aimed at supporting clinical decisions related to heart diseases. This framework includes the following: First, the application of three mining algorithms (FP-Growth, Eclat, and HUIIM) to build association rules based on the extraction of frequent and high-utility patterns, with a focus on environmental and demographic contexts and lifestyle factors that can affect clinical outcomes. Secondly, a dynamic knowledge base is explicitly constructed to predict and diagnose heart diseases based on association rules derived from a set of facts. Finally, the explanation of how association rules are used to improve diagnostic decision support, reduce uncertainty, and provide a foundation for integration with conversational AI systems such as Gemini-based medical assistants.

This paper is systematized as follows: Section 2 discusses earlier research; Section 3 presents background information on association rule mining and its techniques; Section 4 explains the methodology of the proposed framework; Section 5 clarifies the experimental results and discusses them in Section 6; finally, the conclusions are presented in Section 7.

2. Related Works

Numerous studies have enhanced association rule mining (ARM) and knowledge graphs through various models and architectures:

Thurachon W. and Kreesuradej W. 2021 proposed an incremental technique for association rule mining based on the FIUFP-Growth algorithm and the ICP-tree structure with an embedded subtree; it retrieves the frequent itemsets and their support counts from the original database and uses them to mine efficiently on the updated database and ICP-tree, while improving the FP-tree structure and reducing the need to rescan the database. The evaluation was conducted on the synthetic dataset I10T15D100K (with 100,000 original transactions). The results at a support threshold of 3% showed that the average time for mining the increasing patterns was about 46% faster compared to FP-Growth, FUFPP-tree, pre-FUFPP, and FCFPIM. The time for tree updating and mining was about 88% of the time for FP-Growth, around 11% for FUFPP-tree, 9% for pre-FUFPP, and 14% for FCFPIM, with fewer subtrees generated and maintaining efficiency even with very low support or a significant difference in the increasing pattern from the original [10].

Patel P. et al. 2022 proposed an applied framework for mining association rules in liver disease data based on a systematic comparison between Apriori and FP-Growth on the Indian Liver Patient Dataset (ILPD) available on Kaggle, which was used as the benchmark dataset, with performance evaluation using metrics of confidence, lift, leverage, and conviction. Experimentally, FP-Growth outperformed by requiring only two passes over the data to generate frequent itemsets and rules, compared to more passes in Apriori. The best FP-Growth rules achieved confidence between 0.99 and 1.00 and a lift between 2.10 and 3.06. In contrast, the best Apriori rules had a confidence of

0.98 and a lift of 1.00, indicating stronger and more distinctive associations favouring FP-Growth [11].

Cheng Z. et al. 2023 presented the HUIM-SU algorithm for extracting High Utility Itemsets (HUIM), which relies on a simplified utility list to improve efficiency and reduce memory consumption during deep search. The algorithm uses repeated pruning based on the transaction-weighted utilization (TWU) to reduce the number of items and integrate a compressed construction tree to reduce search space and memory consumption. The algorithm also provides an extension tool to set upper bounds that reduce the computational burden. The BMS, Foodmart, Retail, Kosarak, Chess, Connect, Pumsb, and Accident datasets were used to assess the performance of the HUIM-SU algorithm; it demonstrated efficiency by reducing the number of candidates by about two to three orders of magnitude, achieving better execution time on five out of eight datasets, and maintaining a maximum memory usage of only 760 MB across all datasets [12].

Wu H, Xiao Z. 2023 presented an analytical framework that integrates knowledge graph and machine learning techniques to measure the quantitative relationship between drugs and diagnostic methods in the context of the efficacy of dobutamine stress echocardiography for coronary artery disease. The study started with 26,451 research abstracts collected using keywords related to the disease and adenosine, and 7,277 abstracts were filtered to extract entities and diagnostic relationships and build the cognitive graph. The model was applied to three drugs (dipyridamole, dobutamine, and adenosine) and showed that adenosine achieved the greatest positive impact (0.87938) with the least negative impact (0.478387) compared to the other two drugs [13].

G. Vineela et al. 2024 proposed a comprehensive methodological framework for association rule mining based on the Eclat and DEclat algorithms. It uses a practical evaluation mechanism to classify the rules as "interesting" or "non-interesting" based on their applicability in market basket analysis. The evaluation was conducted on a dataset of transactions retrieved from Kaggle, comparing five algorithms (Apriori, FP-Growth, GSP, Eclat, and DEclat) using support, confidence, lift, and conviction metrics. Experimentally, Apriori generated 29,757 interesting rules compared to 2,779 uninteresting ones; FP-Growth generated 3,195 interesting and 582 uninteresting ones; Eclat generated 6,506 interesting and 3,292 uninteresting ones; DEclat generated 6,506 interesting and 3,292 uninteresting ones; and GSP generated 881 interesting and 491 uninteresting ones [14].

Sun Z. et al. 2024 presented a knowledge graph-based learning model to predict disease progression in an interpretable manner; the prediction is reformulated as a directed random walk on a knowledge graph linking diseases to risk factors, with a reinforcement learning policy controlling the transition of the "patient" from their current state to a potential disease node, resulting in an explanatory path that illustrates the progression of the condition. The evaluation was conducted on three electronic health record databases (PLAGH, MIMIC-III, and MIMIC-IV), achieving AUCs of 0.739, 0.639, and 0.643, respectively. It also recorded leading performance

among the top three results in PLAGH (2.008) and MIMIC-III (1.873), with similar performance in MIMIC-IV (1.788) [15].

3. Background

3.1 Association Rule Mining (ARM): is a rule-based approach for discovering relationships (associations/connections/dependencies) between variables in large datasets. In the clinical context, the rule is written in the form $X \rightarrow Y$, where X and Y represent two separate sets of symptoms or diseases such that $X \cap Y = \emptyset$. X is called the "antecedent" because it represents the elements already present in the data. In contrast, Y is called the "consequent" because it expresses the elements expected to appear conditioned on the occurrence of the antecedent [16].

3.2 FP-Growth: is an effective technique for mining frequent patterns based on the structure of the FP-tree (Frequent Pattern tree) [17]. It is characterized by extracting the complete set of patterns without generating candidate sets and using only two scans of the database: in the first scan, the frequencies of the items are calculated to arrange the items in descending order, and in the second scan, the items of each transaction are reordered according to this order and then inserted into the FP tree, where quantitative and structural information about the frequent itemsets is stored [18]. This structure allows for the extraction of patterns later with higher time and memory efficiency compared to methods based on generating candidates [19].

3.3 ECLAT: Equivalence Class Clustering and Lattice Traversal is an effective method for extracting frequent itemsets from transactional databases based on vertical data representation; each item is assigned a list of transaction identifiers in which it appears (TID-set) [20]. The support of any itemset is calculated by intersecting the TID lists of its elements, and the search space is explored using depth-first search (DFS), starting from single elements and then gradually expanding through additional intersections. Elements are grouped into equivalence classes with prefixes or identical TID-sets to reduce potential combinations [21]. Additionally, a prefix-tree/Trie structure represents subproblems and their links, allowing for efficient traversal and pruning by excluding any node that does not meet the minimum support threshold, along with all its branches [22]. In this mechanism, ECLAT quickly calculates support through intersections. It explores frequent itemsets without generating explicit candidate sets or repeatedly scanning the database, benefiting from significant search space compression provided by vertical representation [23].

3.4 HUIM: High Utility Itemset Mining is a utility-oriented approach in data mining aimed at discovering itemsets whose utility exceeds a user-defined threshold [24]. The utility of the group is defined as the sum of the utilities of its elements, where the utility of an element represents a combination of quantity (count/frequency) and importance or value (such as profitability or clinical impact) [25]. Unlike frequent pattern mining, which relies solely on the presence/absence of elements and assigns them equal weight, HUIM allows for the integration of additional information and weighting of elements according to their relevance, enabling the retrieval of high-

value patterns even if their frequency is limited [26]. High-utility itemsets are used to build association rules that support decision-making, and their usefulness has been proven in various applications such as discovering high-profit patterns in retail transaction databases or extracting co-expression patterns in gene expression data [27].

3.5 Knowledge Base/Graph: is defined as a directed graphical representation that gathers and conveys real-world knowledge according to formal semantics that enable computers to process it efficiently, where nodes represent the entities of interest, and edges represent the semantic relationships between them. A knowledge base is a set of data that models facts and relationships in the form of triples (subject, predicate, object) [28, 29]. When these triples are represented as a graph, the knowledge base itself becomes a knowledge graph where nodes are entities and edges are relationships—hence the two concepts are often used interchangeably. And since relationships are not necessarily symmetrical, the direction of the edge is essential, making the knowledge graph a directed graph [30, 31].

2.

4. Methodology

The proposed framework converts patient records in JSON format into a structured binary features table ready for mining by one-hot encoding categorical variables, binning continuous values (such as age and BMI) into categories, cleaning columns, and removing those with no variance. Based on this binary table, three techniques are applied: FP-Growth, which selects about 40 of the most frequent binary features to extract frequent itemsets and generate rules, ranking the best ones by lift; Eclat selects moderately frequent features and converts each row of the table into a "transaction" (a list of items with a value of 1). It then explores the itemsets through TID intersections, generates rules, and sorts them based on confidence and lift. HUI-Miner constructs a utility matrix that prioritizes clinically significant elements (critical/severe/risk_/dx_/symptom_) to extract high-utility itemsets and ranks the rules accordingly. The rules extracted from each technique are evaluated using the bootstrap method through 30 resampling iterations to measure the stability of confidence, support, and lift (means and standard deviations). The rules are directly classified as strong if they exceed the initial thresholds; otherwise, they are considered temporary and stored in a buffer. There, the average of the three metrics is calculated for each rule and compared to a minimum threshold defined as the average of the composite indices for all temporary rules. Rules that surpass this threshold are upgraded to the category of strong rules. Subsequently, a final review is conducted to reduce the confidence in rules with confidence = 1 and very low support to avoid non-generalizable rules. Finally, after recalculating the bootstrap statistics, only the rules that meet the specified criteria are included in the final set.

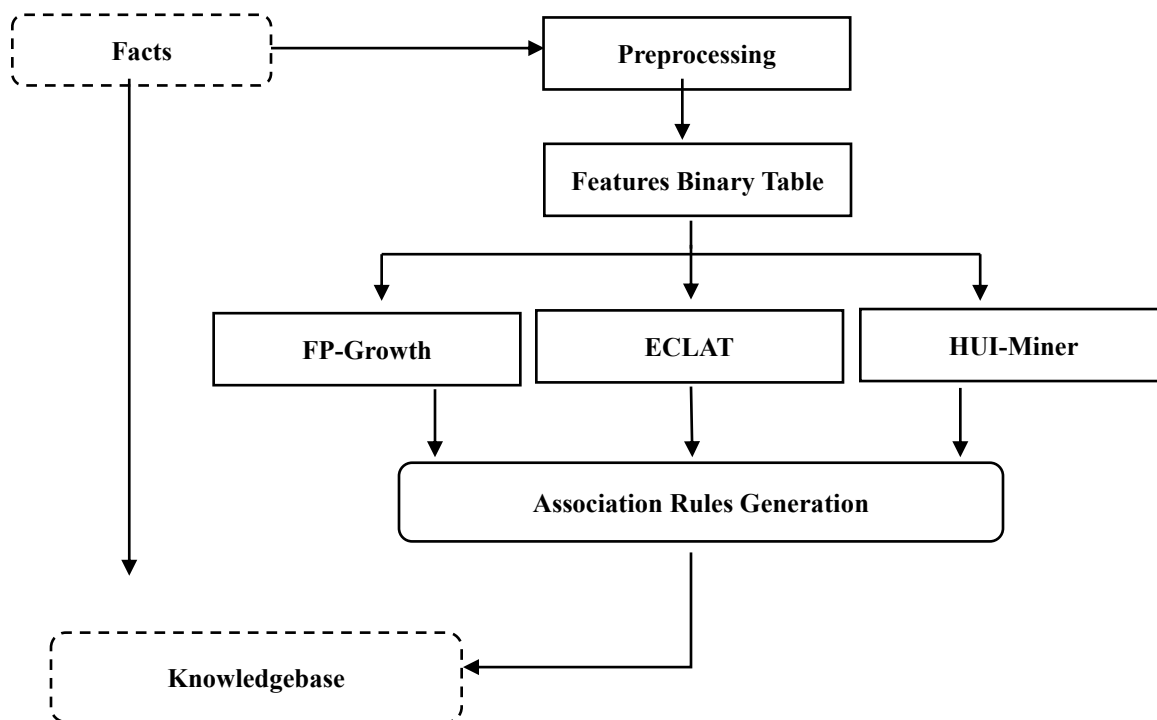


Figure 1: Workflow for the Generation of the Knowledgebase

4.1 *Input Data*

The input data is a JSON file that includes patient records organized as facts extracted from clinical texts on coronary artery disease, each representing a single patient. Each record contains a set of features and their values, including demographic features, symptoms, family history, smoking history, work environment, lab results, heart tests, diagnoses, medications, and risk factors. These facts are constructed by identifying entities from texts using the Named Entity Recognition (NER) technique and then classifying them into predefined categories using the Conditional Random Field (CRF) classifier. Based on the context of each entity, its features and values are extracted to form facts (entity–feature–value).

4.2 *Preprocessing*

This process begins by reading patient records from the fact file (.JSON) and converting each record into a set of organized features integrated into a single data frame that includes demographics (age with classifications, gender, and BMI categories), vital signs (blood pressure levels and pulse indicators), symptoms, risk factors, diagnoses (after purifying and standardizing the name with binary flags and counters), laboratory results (by comparing values to a reference to produce low/normal/high/critical indicators), and cardiac features such as ejection fraction and

ECG results. Missing values are handled consistently (0 for binary, median for numerical, and "unknown" for categorical), and several secondary features are derived (such as age/BMI categories, elderly flag, and symptom/risk/diagnosis counters). This results in an organized clinical feature table (not fully binary yet) that is later used to adjust support and confidence thresholds according to the sample size. Thereafter—and before mining—the table is converted into an enhanced binary representation by defining and selecting the ready binary columns (such as male, elderly, and anything starting with symptom_, risk_, dx_, and ecg_), encoding categorical variables in a one-hot manner for fields like severity, age categories, gender, and BMI categories, binning continuous variables (age, BMI, and symptom/risk/diagnosis counts) into containers that are then converted into binary columns, and finally performing a final cleanup by filling in gaps, converting values to 0/1, and deleting columns with no variance. With these steps, an optimized binary matrix is ready for direct application of mining algorithms (such as FP-Growth, Eclat, and HUI-Miner).

4.3 Mining Algorithms

1. FP-Growth

The preparation process of FP-Growth begins with a pre-prepared binary 0/1 matrix, where the columns (features) are filtered to features with reasonable prevalence by excluding extremely rare ($\leq 5\%$) and extremely common ($\geq 95\%$) features. The number is then limited to about 40 features to reduce noise and improve performance. Then, the minimum support is adaptively adjusted according to the sample size to prevent patterns from being missed in small samples. Next, the frequent itemsets are extracted using the FP-Growth algorithm, and association rules are generated starting with a lift threshold ($\text{Lift} \geq 1.0$). If the number of rules is limited, confidence ≥ 0.1 is used as an alternative. Finally, the rules are ranked by lift, and a standard output is generated that includes the top rules, execution time, memory consumption, number of rules, and a summary of metrics (average support/confidence/lift) for comparison with other algorithms. Table 1 shows the FP-Growth process.

Table (1): FP-Growth Processes/ (a) Represents the Main Data after Converting to Binary Data, (b) Represents the 1's counts at each column and their frequencies

symptom_chest_pain	risk_smoking	dx_angina
1	0	1
0	1	1
1	0	1
1	1	1
0	0	1
0	1	1
1	0	1
0	1	1
1	0	1
0	1	1

(a)

	count_true	Frequency
symptom_chest_pain	5	0.5
risk_smoking	5	0.5
dx_angina	8	0.8
overweight	1	0.1
ecg_q_wave	9	0.9

(b)

2. *Eclat*

The preparation process of Eclat begins with a pre-prepared binary 0/1 matrix, where features with "medium" prevalence are selected (for example, between 0.1 and 0.9) and a maximum of about 25 features to reduce noise and cost. Each row (patient) is converted into a transaction that includes only the names of the columns with a value of 1 (e.g., ["symptoms_chest_pain", "risk_smoking", "dx_angina",...]), and the transaction can be stored as a comma-separated string, as shown in Table 2. After constructing the transaction representation, Eclat relies on the vertical format (TID-lists) and explores the space using a depth-first search approach; it calculates the supports by intersecting the identifier lists to efficiently extract the frequent itemsets without generating explicit candidates. The procedure is followed by generating association rules from each frequent itemset (calculating confidence from the support of the antecedent and lift from the support of the consequent), then

ranking the rules, usually by confidence, and outputting the final package that includes the measured rules, ready for comparison with other algorithms.

Table (2): Eclat Processes/ (a) Represents a sample of Binary Column, (b) Represents the Transaction Structure

symptom_chest_pain	risk_smoking	dx_angina
1	0	1
0	1	1
1	1	0
0	0	0
1	0	1

TID	Items
0	['symptom_chest_pain', 'dx_angina']
1	['risk_smoking', 'dx_angina']
2	['symptom_chest_pain', 'risk_smoking']
3	['symptom_chest_pain', 'dx_angina']

3. *HUI-Miner*

The preparation process of HUI-Miner begins with the pre-prepared binary 0/1 matrix. The columns most clinically relevant (marked with words like "critical," "severe/high," "risk_," "dx_," and "symptom_") are selected, with the option to add the most frequent ones if needed. To control costs, the list is usually limited to about 30 features. Then, the binary values are converted into a utility matrix: the columns are assigned non-cumulative priority weights (only the first match is applied): critical $\times 10$, then severe/high $\times 7$, then risk or dx_ $\times 5$, then symptom_ $\times 3$, and everything else $\times 1$, meaning that each 1 in the column is converted to the corresponding utility weight, while 0 remains unchanged, as shown in Table 3. Then, the high-utility items/groups are identified according to an overall threshold, and the sum of the high-utility items is extracted first. Next, association rules are generated between the highest utility items, calculating support, confidence, and lift, and ordering the rules by utility (returning the number of rules, execution time, memory

consumption, and a summary of metrics such as average support/confidence/lift). In this way, HUI-Miner does not rely solely on the presence of the item but highlights clinically significant patterns even if they are less common, thanks to the weighting of critical and clinically heavy items.

Table (3): HUI-Miner processes/ (a) a sample of Binary Columns Before Multiplying by Weights, (b) After Multiplying by Weights

Columns	Patient A	Patient B
lab_troponin_critical_high	1	0
symptom_chest_pain	1	1
risk_smoking	0	1
dx_angina	1	1

Columns	Patient A	Patient B
lab_troponin_critical_high	10	0
symptom_chest_pain	3	3
risk_smoking	0	5
dx_angina	5	5

The rules extracted from the above algorithms are evaluated by applying the bootstrap technique, which involves sampling with replacement from the available rules for several iterations (30 iterations). This technique aims to evaluate the rules' stability and determine their strength by calculating the means and standard deviations of confidence, support, and lift. After calculating these values, the rules are classified as "strong" if they exceed the initial criteria for confidence, support, and lift, and as "temporary" if they do not meet these criteria but require further processing. The temporary rules are not entirely excluded but are placed in the buffer for later processing using the composite score, which is calculated for each rule by averaging the three values (confidence, support, and lift). Then, a threshold for the composite score is determined based on the arithmetic mean of all the temporary rules, where the rules that exceed this threshold are saved and added to the strong rules. After adding the rules from the buffer, the final verification of the strong rules is done by reducing the confidence if the rule has perfect confidence and very low support. Finally, the means and standard deviations are calculated via bootstrap for confidence, support, and lift, and the rules that exceed the specified criteria are added to the final set. This flow

produces more reliable and interpretable rules that support diagnostic decision-making, reduce uncertainty, and pave the way for seamless integration with interactive medical aids.

4.4 Knowledge base

The knowledge base was built using a sequential methodology that begins with extracting clinical facts from records and unstructured texts—such as age, gender, body mass index, smoking status, lab results, and ECG—through an information extraction step that transforms free-text statements into organized features. These facts are then categorized into the normal and abnormal classes based on their clinical significance. Then, IF-THEN rules are formulated, linking a single fact or a set of facts using logical connectors (AND/OR/NOT) to generate diagnostic or therapeutic outcomes. In this way, the knowledge base is formed by facts and rules that are later used in case classification and generating recommendations. Due to the medical nature, the rules can be encoded in standard executable formats within clinical decision support systems, facilitating sharing, reuse, verification, and governance. With this structure, the knowledge base—composed of extracted facts (normal/abnormal) and interpretative rules—provides a high level of traceability and clinical justification when classifying coronary artery disease cases and making diagnostic and therapeutic decisions.

5. Experimental Results

Three approaches for extracting association rules have been adopted: FP-Growth, ECLAT, and HUI-Miner—each with a dedicated data processing method. The process begins by converting the previously extracted facts in the form of (entity-attribute-value) into a binary/transaction representation suitable for mining. Then, the rule generation is executed for each approach individually. After generation, a unified filtering is applied to the rules according to fixed criteria: a minimum support threshold, confidence adapted to the sample size, and a lift condition > 1 to retain the strong and most significant rules. The resulting rules are evaluated according to the following criteria:

1. Support: The proportion of transactions that contain the set of concerned items. For the rule $A \Rightarrow B$, the support is defined as the support of both the antecedent and the consequent combined as computed in Equation (1) [32].

$$P(A \cap B) = \frac{\text{No. of Transactions including } A \text{ and } B}{\text{Total No. of Transactions}} \quad (1)$$

Where A and B are two separate sets of items that appear at the same time in a transaction.

2. Confidence: The probability of the occurrence of B given the occurrence of A; that is, among all those who have A, as computed in Equation (2) [33].

$$P(B|A) = \frac{\text{No.of Transactions including A and B}}{\text{No.of Transactions including A}} \tag{2}$$

3. Lift: It measures the strength of the correlation compared to what we would expect if A and B were independent, as computed in Equation (3) [34].

$$\text{Lift}(A \rightarrow B) = \frac{\text{Supp}(A \cup B)}{\text{Supp}(A) \times \text{Supp}(B)} \tag{3}$$

If lift > 1: Positive correlation (the occurrence of B is more likely when A is present than we would expect by chance).

If lift=1: Independence (no predictive value).

If lift < 1: negative correlation.

Table 4: Evaluation Criteria for Association Rules Techniques

algorithms	Overall score	Rule generation	Quality score	Performance Score	Efficiency	Reliability Score	No. of Rules	Avg. Confidence	Avg. Support	Avg. Lift	Execution Time	Memory Usage	Rank
Eclat	0.8995	1.0000	0.8328	0.7485	1	1.0000	5000	1.0000	0.3900	1.3376	0.6393	0.0022	1
HUI-Miner	0.8849	0.9994	0.6719	0.9759	1	0.8968	497	0.6742	0.3999	1.0225	0.0612	0.0026	2
FP-Growth	0.7403	1.0000	0.8362	0.0000	1	0.8939	500	0.7033	0.1791	2.5589	2.5419	0.0034	3

Table 4 provides a comparison between Eclat, HUI-Miner, and FP-Growth across quality, performance, and reliability indicators with a final ranking: Eclat tops the list with an overall score of 0.8995, thanks to complete rule generation (500/500), reliability and confidence = 1.0, average support ≈0.39, and lift ≈1.33, with good performance (time ≈0.639s and memory ≈0.0022MB). HUI-Miner follows with a score of 0.8849, attributed to its high speed (≈0.061 s), generation efficiency of 0.994, and support of ≈0.399; however, it has lower confidence (0.674), lift close to independence (≈1.02), and lower reliability (0.8968). FP-Growth comes in third with a score of 0.7403 due to a longer execution time (≈2.542 s) and higher memory usage, despite its superiority in lift (≈2.55) and a confidence level of 0.703 with lower support (≈0.179). Therefore, Eclat is the most balanced and reliable; HUI-Miner is suitable when speed and coverage are preferred. FP-

Growth is useful for discovering the strongest associations with an acceptance of slower speed and lower coverage.

6. RESULTS AND DISCUSSION

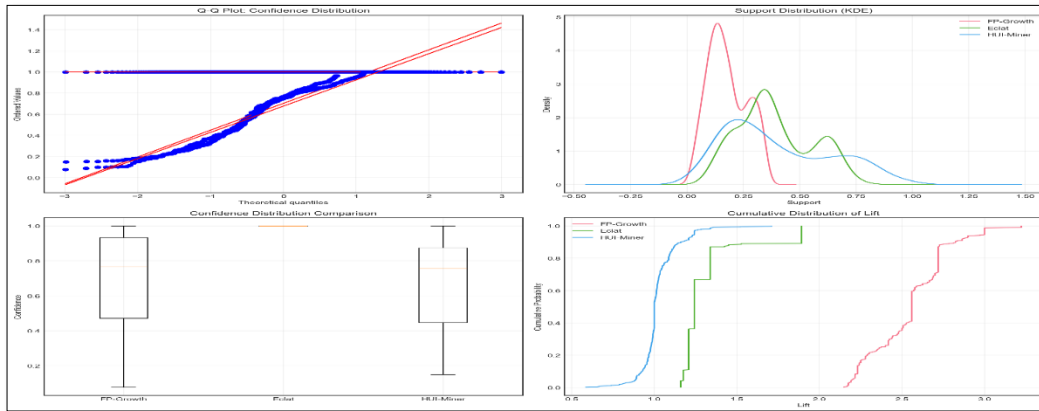


Figure 2: Confidence, Support, and Lift Distribution Comparison

Figure 2 shows that Eclat offers the highest confidence with a very narrow variance (median ≈ 1.0 and a slim box), indicating semi-deterministic and specialized rules with a moderate lift (ECDF curve stops near ~ 1.5). In contrast, FP-Growth is characterized by an average confidence and wider dispersion. Still, it achieves the highest lift values (The ECDF curve is strongly skewed to the right and extends beyond 3) compared to relatively lower support (support density is centered around ~ 0.25). HUI-Miner demonstrates broader and right-skewed coverage/support (approximately 0.25 to 0.9), along with lower confidence and higher variance, which improves as support increases; the lift is typically low to moderate (around 1.3 to 1.4). The Q-Q plot confirms the non-normality of the confidence distribution and the presence of a "ceiling effect" at 1.0 due to the density of Eclat values. Therefore, FP-Growth is the most suitable for identifying the strongest associations, Eclat is ideal for achieving nearly certain and homogeneous confidence, while HUI-Miner is the best when the goal is to maximize coverage and practical feasibility.

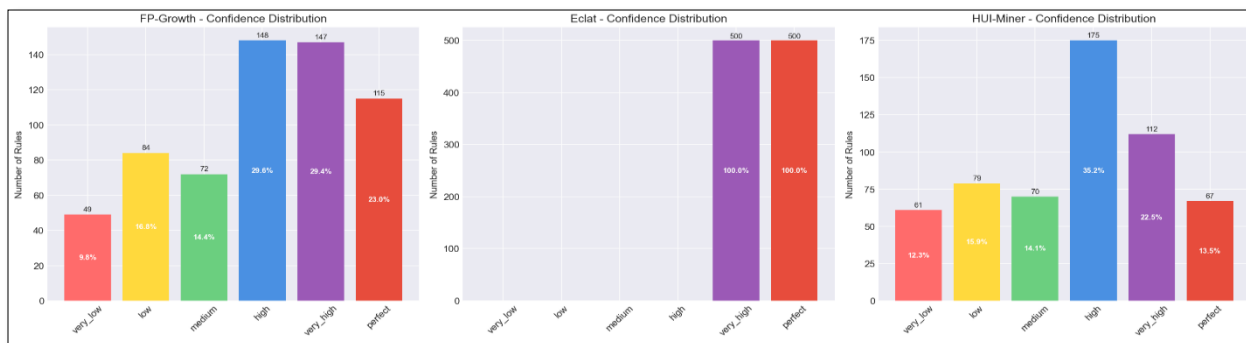


Figure 3: Confidence Distribution Comparison

Figure 3 summarizes the distribution of confidence for the three algorithms' rules across six categories: for FP-Growth, the output is distributed across the spectrum but is concentrated at the top: high ≈ 148 (29.6%), very_high ≈ 147 (29.4%), and perfect ≈ 115 (23.0%), with a noticeable tail toward the lower categories (low ≈ 84 , medium ≈ 72 , and very_low ≈ 49). As for Eclat, its rules appear almost entirely in the very_high and perfect categories (confidence ~ 1.0 practically). HUI-Miner provides a broader distribution, peaking at high with approximately 175 (35.2%), followed by very_high at approximately 112 (22.5%) and perfect at approximately 67 (13.5%), while also showing a clear presence of lower categories: very_low at approximately 61, low at approximately 79, and medium at approximately 70. Therefore, Eclat generates quasi-deterministic rules, FP-Growth tends toward high/full confidence with a low tail, and HUI-Miner distributes rules across the spectrum with a peak at high confidence.

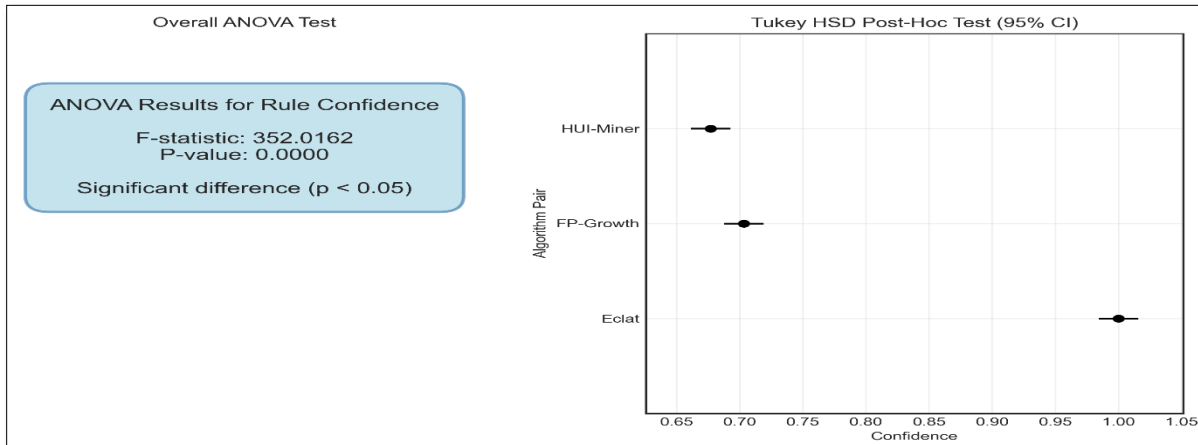


Figure 4: ANOVA and Post-Hoc Analysis of Rule Confidence

Figure 4 illustrates the results of the comparative analysis of the confidence of association rules using one-way ANOVA

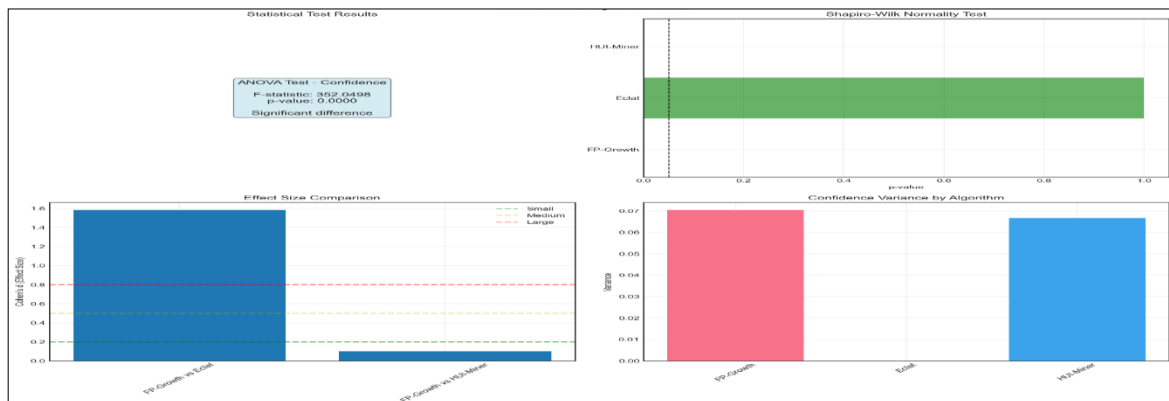


Figure 5: Statistical Analysis Results

y ANOVA followed by the Tukey HSD test. The high F value of 352.0162 with $p = 0.0000$ indicates that the variance between the means of the algorithms exceeds the variance within each algorithm, reflecting significant differences. Therefore, the interpretation is done using the Tukey test, which displays 95% non-overlapping confidence intervals. The Eclat algorithm leads with an average confidence close to 1.0 and a narrow interval, indicating the generation of strong and specialized rules. In contrast, the FP-Growth algorithm has an average confidence of approximately 0.70, which balances the extraction of strong and moderate patterns without the need to generate candidates, resulting in reduced confidence but increased diversity. As for HUI-Miner, it records a slightly lower average (≈ 0.65) due to its focus on maximizing utility/weight rather than just confidence, resulting in the extraction of high-value practical rules even if they do not raise the overall average confidence.

Figure 5 shows that the confidence means differ significantly among the three algorithms according to the ANOVA analysis ($F \approx 352.05$, $p = 0.0000$). In the Shapiro-Wilk test, the confidence of Eclat appears nominally normal ($p \approx 1$) but is almost constant around 1.0, which limits the usefulness of the test and justifies reinforcing the conclusion with a non-parametric test like Kruskal-Wallis; also, the p-values for both FP-Growth and HUI-Miner exceed 0.05, indicating the acceptance of the normality assumption. The effect size (Cohen's d) reveals a significant practical difference between Eclat and FP-Growth ($d \approx 1.58$) compared to a slight, almost negligible difference between FP-Growth and HUI-Miner ($d \approx 0.08$). In terms of variance, it is approximately 0.076 in FP-Growth (wider fluctuation), ≈ 0.000 in Eclat (high and uniform confidence), and ≈ 0.071 in HUI-Miner (medium fluctuation). Therefore, Eclat provides high and uniform reliability, while FP-Growth offers a balance with greater variability in results. HUI-Miner falls between the two with an advantage when adjusting weights.

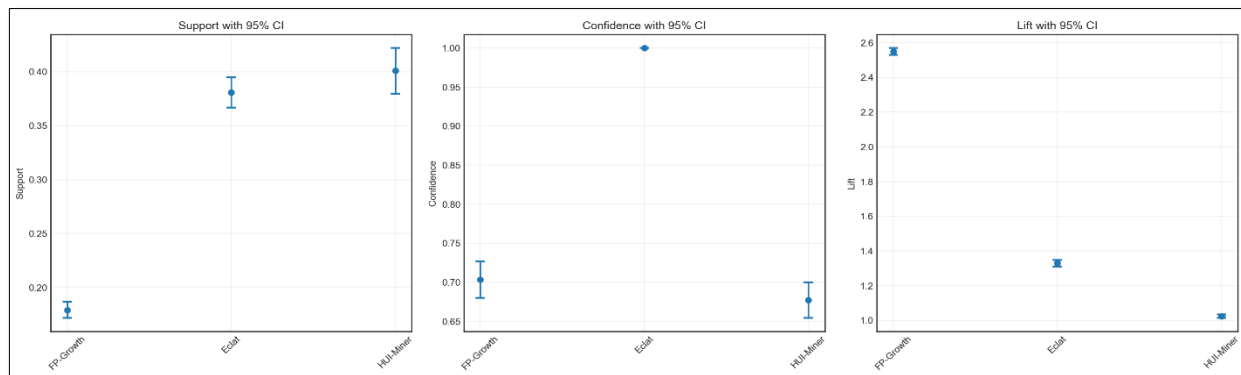


Figure 6: Confidence Intervals for Rule Metrics

Figure 6 illustrates the differences in metrics among the three algorithms as follows: in terms of support, HUI-Miner ($\approx 0.37-0.39$) leads with narrow confidence intervals, followed by Eclat

($\approx 0.20-0.25$), and then FP-Growth ($\approx 0.18-0.20$). In terms of confidence, Eclat excels with near-complete confidence ($\approx 0.95-1.00$) and narrow confidence intervals, while FP-Growth achieves a satisfactory average confidence (≈ 0.70), followed by HUI-Miner ($\approx 0.60-0.65$). In terms of lift, FP-Growth leads ($\approx 2.0-2.5$), indicating stronger associations than random, followed by HUI-Miner ($\approx 1.1-1.2$), and then Eclat ($\approx 1.05-1.15$). Therefore, FP-Growth is the best for discovering the strongest associations, Eclat has the highest confidence, and HUI-Miner has the widest coverage.

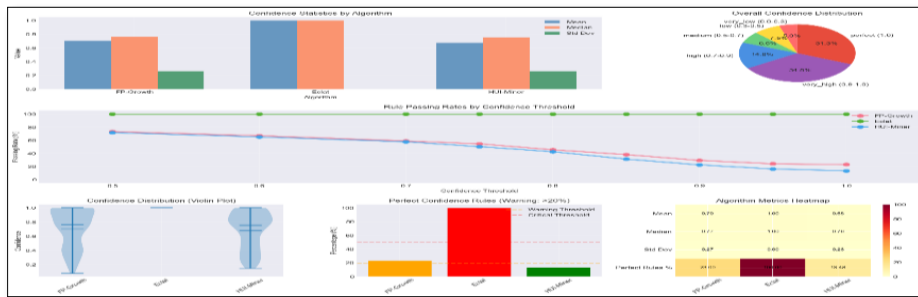


Figure 7: Confidence Analysis Dashboard

Figure 7 summarizes the following: Eclat achieves a mean and median of approximately 1.00 with a deviation of approximately 0.00, indicating almost constant confidence, and it passes 100% of the rules for all thresholds from 0.5 to 1.0; additionally, the percentage of rules with complete confidence is approximately 100%. FP-Growth: Average 0.70, median 0.77, and deviation 0.27; the pass rate decreases from $\sim 70\%$ (0.5) to $\sim 23\%$ (1.0), and the complete rules are approximately 23%. HUI-Miner: Average 0.68, median 0.76, deviation 0.26; the pass rate drops from $\sim 70\%$ to $\sim 15\%$ at 1.0, and the complete rules $\approx 13.5\%$. The pie chart shows that most rules fall within the range of 0.9–1.0 (perfect $\approx 31.3\%$ and very_high $\approx 34.8\%$), and the violin plots display a sharp cluster near 1.0 for Eclat compared to wider distributions for FP-Growth and HUI-Miner. Eclat provides near-certain and stable confidence; HUI-Miner has broader coverage but is more sensitive to threshold adjustments. FP-Growth balances confidence and diversity with a significant share of complete rules.

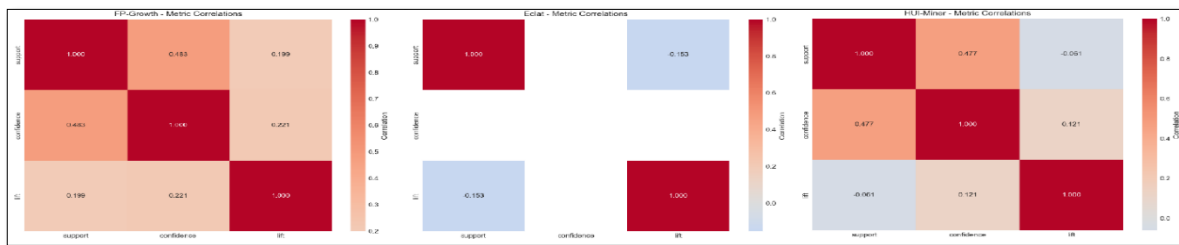


Figure 8: Correlation Analysis for Rule Metrics

Figure 8 indicates that the relationships of the metrics vary depending on the algorithm: in FP-Growth, support is moderately positively correlated with confidence (0.46), while the correlation of lift with both support and confidence remains weak (0.20 and 0.22), meaning that broader rules are not necessarily the highest in lift. In Eclat, there is a slight negative correlation between support and lift (-0.15), with other correlations being nearly nonexistent, which is a natural effect of precise and specialized rules resulting from the vertical TID representation. As for HUI-Miner, it shows a moderate positive correlation between support and confidence (0.48), accompanied by a slight negative correlation between support and lift (-0.06) and a very weak correlation between confidence and lift (~0.12), suggesting that higher lift often manifests in less common patterns. Therefore, FP-Growth balances coverage and accuracy with a reasonable amount of lift, while Eclat is suitable when high accuracy is preferred, even at the expense of support. On the other hand, HUI-Miner provides broader coverage/utilization with weight calibration.

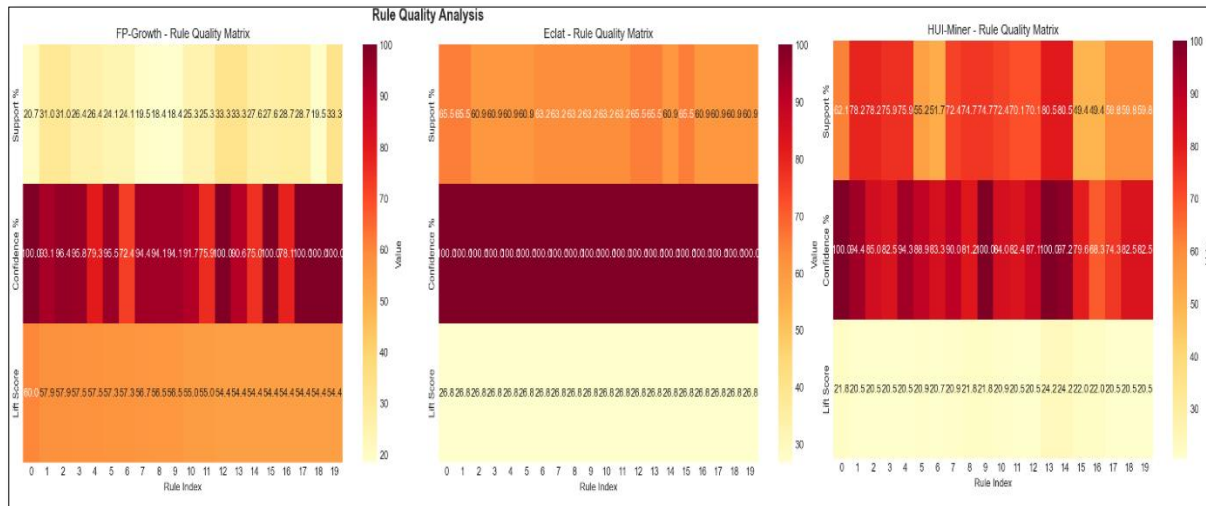


Figure 9: Rules Quality Analysis

Figure 9 summarizes how the quality of rules varies with different algorithms: FP-Growth extracts infrequent strong patterns (support around 18–33%) with a high, nearly constant lift (around 54–60) and high confidence with local fluctuations; Eclat provides medium to high support (around 60–66%) and nearly 100% confidence thanks to TID representation, but with medium and constant lift (around 26–27), which requires verification to avoid general rules; HUI-Miner offers broader coverage with relatively high support (around 49–80%) and high but fluctuating confidence, with the lowest lift (around 20–25) because the utility criterion tends to frequent rules. The strongest associations are often with FP-Growth, the highest accuracy with Eclat, and the broadest practical coverage with HUI-Miner.

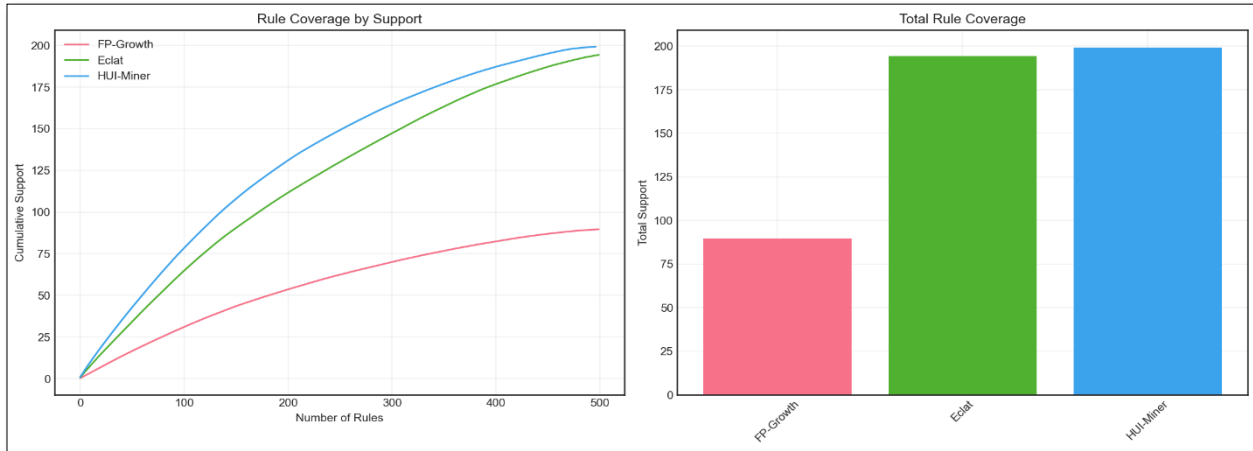


Figure 10: Rules Coverage Analysis

Figure 10 indicates that HUI-Miner achieves the highest early coverage due to a steep cumulative slope—because it prefers high-support rules—then its curve begins to flatten around 300 rules due to rule overlap and diminishing returns. In contrast, Eclat rises at a nearly linear pace and continues to add "new coverage" for a more extended period because its rules are less overlapping. Still, it ends slightly lower than HUI-Miner in total. As for FP-Growth, its slope is slower, and it ends with the lowest overall coverage, as it focuses on the strongest associations (higher lift), often with less support. It is worth noting that the accumulation here is the sum of support, not the union of cases; therefore, duplicate counting may occur in the overlap. Consequently, HUI-Miner achieves the highest overall coverage, followed closely by Eclat, and then FP-Growth.

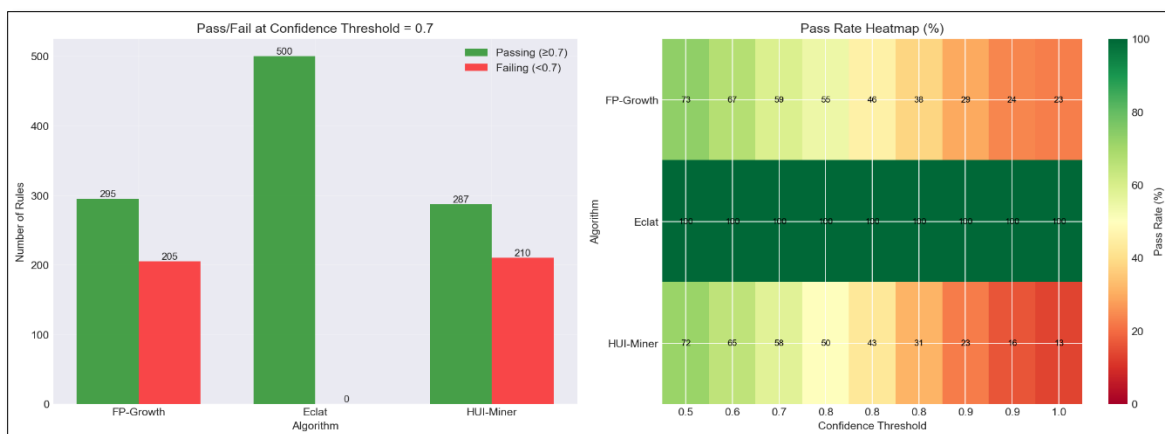


Figure 11: Thresholds of Passing/Failing the Rules

Figure 11 summarizes the success rates according to confidence thresholds for three algorithms: at the 0.7 threshold, FP-Growth succeeds at approximately 59% (295/500), Eclat at 100%

(500/500), and HUI-Miner at approximately 58% (287/497). The heatmap shows that Eclat maintains 100% across the thresholds 0.5→1.0. In comparison, the success of FP-Growth gradually decreases from 73% (0.5) to 23% (1.0), and HUI-Miner follows a similar pattern but with a steeper decline at higher thresholds (72% → 13%). Therefore, Eclat is the most stable for strict thresholds, while FP-Growth and HUI-Miner are close around 0.6–0.7. However, raising the threshold above 0.8 harms HUI-Miner more; thus, the choice depends on balancing confidence strictness and the number of accepted rules.

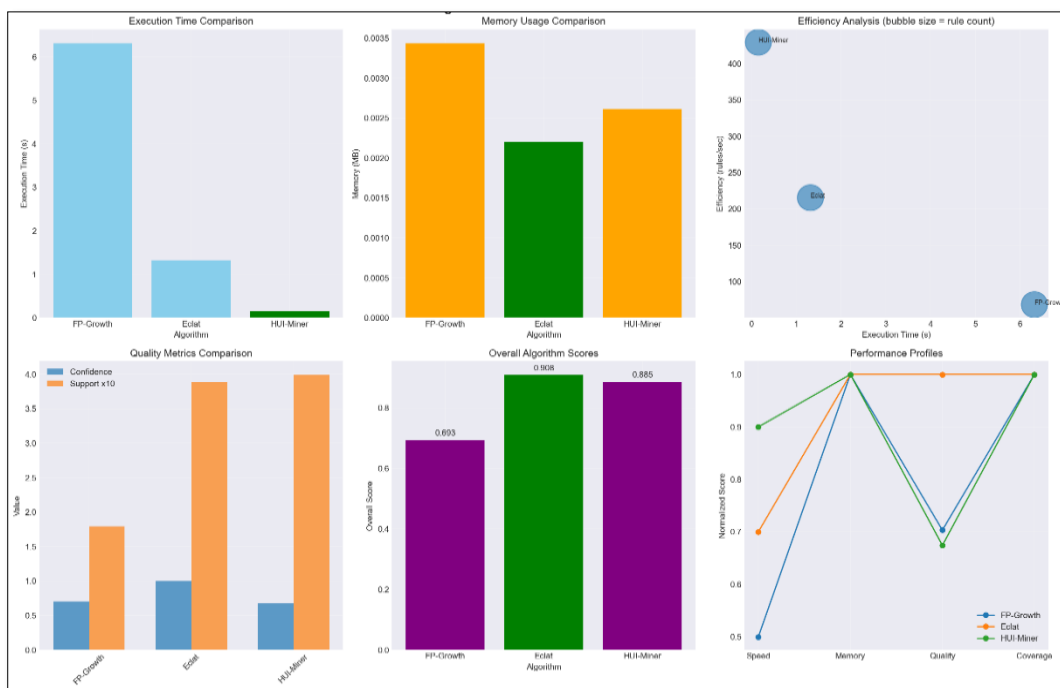


Figure 12: Algorithm Performance Dashboard

Figure 12 shows that HUI-Miner is the fastest (≈ 0.18 seconds), followed by Eclat (≈ 1.3 seconds), and then FP-Growth (≈ 6.3 seconds). Eclat is the least memory-consuming (≈ 0.0022 MB), followed by HUI-Miner (≈ 0.0026 MB), and then FP-Growth (≈ 0.0034 MB). In terms of efficiency (rules/second), HUI-Miner leads ($\approx 420\text{--}430$), followed by Eclat (≈ 215), while FP-Growth comes last (≈ 80). In terms of quality, Eclat achieves near-complete confidence (~ 1.00) with support ≈ 0.39 , while HUI-Miner provides the highest coverage with support ≈ 0.405 and confidence ≈ 0.68 . FP-Growth has an average confidence of ≈ 0.70 and support of ≈ 0.17 . The overall result confirms the superiority of Eclat (0.908), followed by HUI-Miner (0.885), and then FP-Growth (0.693). Practically: Use Eclat to achieve the best overall balance, start with HUI-Miner if you desire a quick return and higher speed, and include selected rules from FP-Growth when the goal is to maximize the strength of association, even with lower coverage.

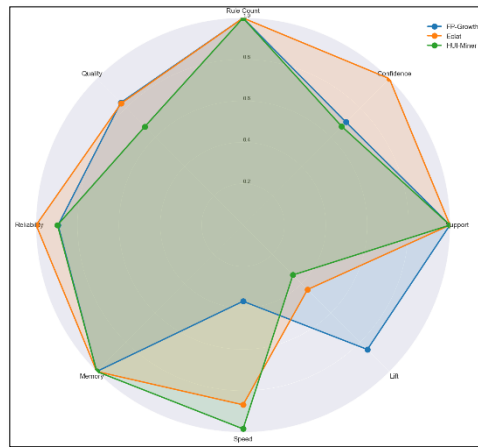


Figure 13: Comprehensive Algorithm Performance Radar

Figure 13 illustrates Eclat's superiority in confidence (≈ 1.0), number of rules, reliability, and quality, with good efficiency in speed and memory, while its lift rate remains average (~ 0.45). HUI-Miner excels in speed (~ 0.95) and coverage/support (≈ 1.0) with efficient memory and a large number of rules, but it records lower lift and quality/reliability than Eclat (0.35 and 0.6–0.9). As for FP-Growth, it features the highest lift rate (~ 0.75), good quality (~ 0.8), and moderate reliability (~ 0.85), but it is the slowest and appears relatively less efficient in memory. Additionally, its support is usually narrower because it captures less common strong patterns. Therefore, Eclat is the best for precision, consistency, and resource efficiency. At the same time, HUI-Miner excels in speed and broader scope, and FP-Growth is the ideal choice in terms of strength of association.

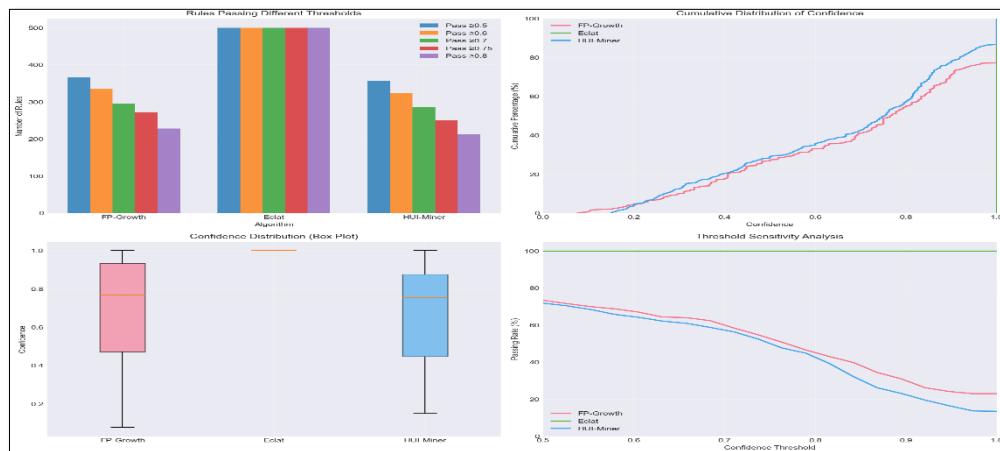


Figure 14: Threshold Effect on the Rules' Quality

Figure 14 summarizes the effect of raising the confidence threshold on the quality of rules for three algorithms. Eclat remains at nearly optimal performance across all thresholds (confidence ≈ 1.0 and support 100%). At the same time, the number of accepted rules for FP-Growth and HUI-Miner gradually decreases from about 350–360 rules at 0.50 to around 230–250 at 0.80. The ECDF curve shows that Eclat values cluster at 1.0, and HUI-Miner tends to have lower to medium confidence levels compared to FP-Growth, which retains a higher mass near the larger values without reaching perfection. The box plots show almost no dispersion for Eclat at 1.0 compared to a wider variance for FP-Growth (median ≈ 0.76) and HUI-Miner (median ≈ 0.74). On the sensitivity curve, Eclat remains at 100%, while FP-Growth drops from $\sim 70\%$ (0.50) to $\sim 25\%$ near 1.0, and HUI-Miner declines faster ($\sim 70\% \rightarrow \sim 15\%$). Eclat provides near-certain confidence and consistent performance, HUI-Miner offers broader coverage with higher threshold sensitivity, and FP-Growth supplies a better balance than HUI-Miner at high thresholds.

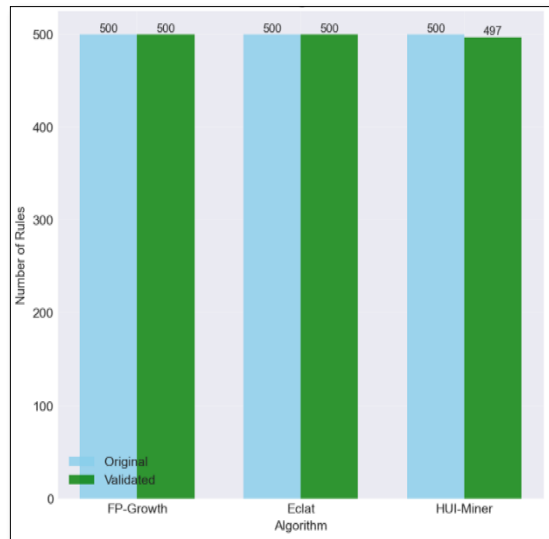


Figure 15: Rules Number before and after Verification for Each Algorithm

Figure 15 illustrates the comparison of the number of rules before and after verification for each algorithm; FP-Growth retained all its rules (500 original \rightarrow 500 verified, 100% retention), as did Eclat (500 \rightarrow 500, 100% retention), while HUI-Miner recorded a very slight decline (500 \rightarrow 497, losing three rules and retaining $\approx 99.4\%$). The result indicates that the rules of FP-Growth and Eclat passed the verification completely, reflecting high stability within the verification data/tests. At the same time, the slight deficiency in HUI-Miner may be due to low-support/trust boundary rules that did not recur in the verification set or to the removal of conflicts/duplicates during filtering.

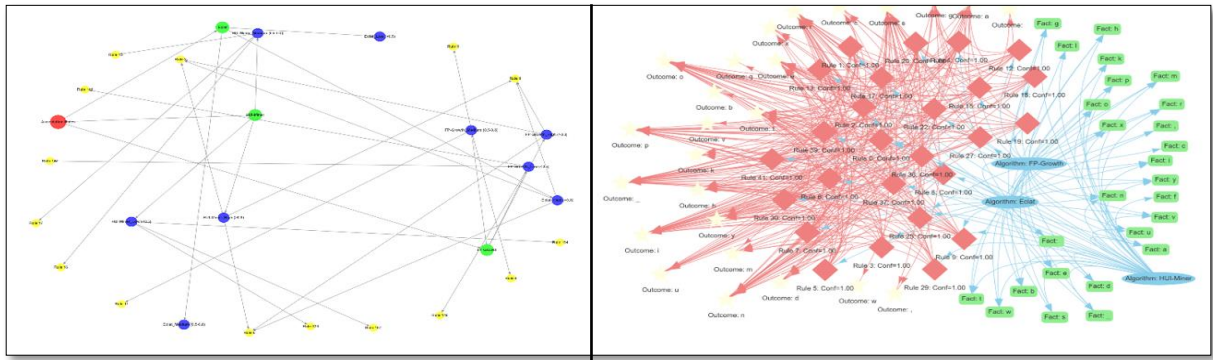


Figure 16: Hierarchical Rule Structure

Figure 16 displays a hierarchical network of association rules with two levels of clarification: In the left panel, the diagram starts from a red root, Association Rules, branching into green nodes representing the algorithms (Eclat, FP-Growth, HUI-Miner), then blue nodes for confidence layers (High >0.8, Medium 0.5–0.8, Low <0.5), and finally yellow nodes for individual rules, noting that the size of the nodes reflects the level of aggregation, not a scale value. The connections are denser around Eclat–High because the vertical representation and intersection of TID sets produce highly correlated antecedents, leading to many high-confidence rules. In contrast, the branches of HUI-Miner are distributed in low/medium because it prioritizes utility/coverage over maximizing confidence. The branches of FP-Growth appear less numerous in Medium/High—and occasionally Low—because it captures strong but less common associations. As for the right panel, it adds the semantic meaning to the rule paths: from each algorithm, the arrows point to facts/attributes (Facts), then to rules distinguished by confidence values (Conf), and end at outcomes (Outcomes). Any path is read as follows: Algorithm → Fact(s) → Rule (Conf=...) → Outcome. In this formulation, the left side highlights the structure and organization, while the right side shows how facts are clinically linked to rules and outcomes.

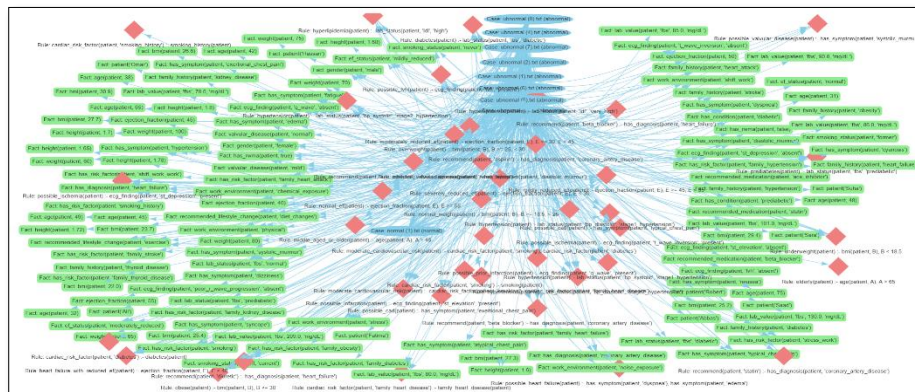


Figure 17: Medical Knowledge Base

Figure 17 represents a knowledge base built on the sequence "Facts \Rightarrow Rules \Rightarrow Classification (normal/abnormal)"; the green nodes are clinical facts extracted from patient files (age, gender, BMI, smoking status, family history, FBS values, ECG results, symptoms, etc.), and the red diamonds are rules of the IF...THEN type that combine multiple facts to produce a single conclusion (for example, smoking + diabetes + ECG changes \rightarrow possible ischemia, or very low ejection fraction \rightarrow severely reduced ejection fraction, or CAD diagnosis \rightarrow recommendation (statin)); the arrow entering the rule means that the fact is part of its premise, and the arrow exiting goes to the conclusion. As for the blue oval shapes at the top of the diagram, they represent the conditions/classifications of the patients (abnormal or normal) and are connected by arrows when the conditions of the rule are met in the patient's file. Thus, the interpretation path is always green fact \rightarrow red rule \rightarrow blue state/result. A dense concentration of links is observed in the middle of the diagram around high-impact facts such as smoking, diabetes/prediabetes, obesity or underweight, age ≥ 65 , family history of heart disease, hypertension, ECG changes, and reduced ejection fraction; these feed diagnostic rules (possible ischemia/heart failure/valvular disease...) and therapeutic rules (recommend beta-blocker/statin/ACE-inhibitor) often toward abnormal nodes, while reassuring facts such as non-smoking and normal ECG/EF lead to normal nodes.

7. CONCLUSIONS

This study addresses the problem of early diagnosis of coronary artery disease (CAD) in a heterogeneous clinical data environment. It proposes a practical solution based on extracting association rules from a patient's "facts file" after cleaning and converting them into a structured binary representation, using three algorithms: FP-Growth, Eclat, and HUIM. The rules are generated and filtered according to support/confidence/lift criteria to retain the strongest rules supporting the diagnosis. Eclat demonstrated the highest reliability, achieving an average confidence close to 1.0 with near-zero variance, passing all thresholds (0.5–1.0) with an average support of ≈ 0.39 and a moderate lift of ≈ 1.3 . ANOVA/Tukey tests confirmed its significant superiority and large effect size compared to FP-Growth and HUIM. The rules generated by Eclat are preserved with the clinical facts that constitute them within a knowledge base that considers the contextual variation (environmental, demographic, and lifestyle) between Iraqi regions, thereby supporting more accurate diagnostic decisions. Future directions include measuring union coverage instead of cumulative support, evaluating the practical impact of the rules on doctors' decisions and patient outcomes, adopting periodic updates to the rules (incremental mining) while monitoring clinical pattern drift, as well as temporal modeling of vital signs and integration with conversational AI systems.

REFERENCES

- [1] Hikmawati E, Maulidevi NU, Surendro K. Minimum threshold determination method based on dataset characteristics in association rule mining. *Journal of Big Data*. 2021 Nov 25;8(1):146. DOI: <https://doi.org/10.1186/s40537-021-00538-3>.
- [2] Khorev V, Krivoš-Belluš R, Pristaš V, Shridhara MG, Antoni LU, Krídlo O, Krajči S. Comparison of Two Methods of Association Rules Based on Running Time for Different Number of Transactions and Support Values. In *2024 International Conference on Emerging eLearning Technologies and Applications (ICETA) 2024 Oct 24* (pp. 321-325). IEEE. DOI: [10.1109/ICETA63795.2024.10850832](https://doi.org/10.1109/ICETA63795.2024.10850832).
- [3] Das A, Jana S, Ganguly P, Chakraborty N. Application of association rule: Apriori algorithm in E-Commerce. In *2021 Innovations in Energy Management and Renewable Resources (52042) 2021 Feb 5* (pp. 1-7). IEEE. DOI: [10.1109/IEMRE52042.2021.9386737](https://doi.org/10.1109/IEMRE52042.2021.9386737).
- [4] Xu R, Luo F. Risk prediction and early warning for air traffic controllers' unsafe acts using association rule mining and random forest. *Safety science*. 2021 Mar 1; 135:105125. DOI: <https://doi.org/10.1016/j.ssci.2020.105125>.
- [5] Alsaeedi HA, Alhegami AS. An Incremental Interesting Maximal Frequent Itemset Mining Based on FP-Growth Algorithm. *Complexity*. 2022;2022(1):1942517. DOI: <https://doi.org/10.1155/2022/1942517>.
- [6] Krishnan MS, Nair AS, Sebastian J. Comparative analysis of apriori and ECLAT algorithm for frequent itemset data mining. In *Ubiquitous Intelligent Systems: Proceedings of ICUIS 2021 2021 Oct 9* (pp. 489-497). Singapore: Springer Singapore. DOI: https://doi.org/10.1007/978-981-16-3675-2_37.
- [7] Lin JC, Djenouri Y, Srivastava G, Fourier-Viger P. Efficient evolutionary computation model of closed high-utility itemset mining. *Applied Intelligence*. 2022 Jul;52(9):10604-16. DOI: <https://doi.org/10.1007/s10489-021-03134-3>.
- [8] Thomas A, Sangeetha S. Semi-supervised, knowledge-integrated pattern learning approach for fact extraction from judicial text. *Expert Systems*. 2021 May;38(3): e12656. DOI: <https://doi.org/10.1111/exsy.12656>.
- [9] Ji S, Pan S, Cambria E, Marttinen P, Yu PS. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*. 2021 Apr 26;33(2):494-514. DOI: [10.1109/TNNLS.2021.3070843](https://doi.org/10.1109/TNNLS.2021.3070843).

- [10] Thurachon W, Kreesuradej W. Incremental association rule mining with a fast incremental updating frequent pattern growth algorithm. *IEEE Access*. 2021 Apr 8; 9:55726-41. **DOI:** [10.1109/ACCESS.2021.3071777](https://doi.org/10.1109/ACCESS.2021.3071777).
- [11] Patel P, Sivaiah B, Patel R. Relevance of frequent pattern (FP)-growth-based association rules on liver diseases. In *Intelligent Systems: Proceedings of ICMIB 2021 2022 May 4* (pp. 665-676). Singapore: Springer Nature Singapore. **DOI:** https://doi.org/10.1007/978-981-19-0901-6_58.
- [12] Cheng Z, Fang W, Shen W, Lin JC, Yuan B. An efficient utility-list based high-utility itemset mining algorithm. *Applied Intelligence*. 2023 Mar; 53(6):6992-7006. **DOI:** <https://doi.org/10.1007/s10489-022-03850-4>.
- [13] Wu H, Xiao Z. A knowledge graph-based analytical model for mining clinical value of drug stress echocardiography for diagnosis, risk stratification and prognostic evaluation of coronary artery disease. *International journal of cardiology*. 2023 Sep 15; 387:131107. **DOI:** <https://doi.org/10.1016/j.ijcard.2023.05.057>.
- [14] G. Vineela, M. Alekhya, A. Jayanth, J. M. V. L. Harshitha, V. P. K. S., and V. K. K. K, An Experimental Analysis of Association Rule Mining Algorithms to Extract Strong and Interesting Association Rules. *Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 2024*, pp. 1-7, **DOI:** 10.1109/ICAIT61638.2024.10690451.
- [15] Sun Z, Dong W, Shi J, Huang Z. Interpretable disease progression prediction based on reinforcement reasoning over a knowledge graph. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2024 Mar 29;54(3):1948-59. **DOI:** [10.1109/TSMC.2023.3331847](https://doi.org/10.1109/TSMC.2023.3331847).
- [16] Marques JG, Carvalho BM, Guedes LA, Da Costa-Abreu M. Using association rules to obtain sets of prevalent symptoms throughout the COVID-19 pandemic: an analysis of similarities between cases of COVID-19 and unspecified SARS in São Paulo-Brazil. *International Journal of Environmental Research and Public Health*. 2024 Sep 1;21(9):1164. **DOI:** <https://doi.org/10.3390/ijerph21091164>.
- [17] Aldino AA, Pratiwi ED, Sintaro S, Putra AD. Comparison of market basket analysis to determine consumer purchasing patterns using FP-Growth and Apriori algorithm. In *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE) 2021 Oct 27* (pp. 29-34). IEEE. **DOI:** [10.1109/ICOMITEE53461.2021.9650317](https://doi.org/10.1109/ICOMITEE53461.2021.9650317).

- [18] Jang HJ, Yang Y, Park JS, Kim B. Fp-growth algorithm for discovering region-based association rule in the iot environment. *Electronics*. 2021 Jan;10(24):3091. **DOI:** <https://doi.org/10.3390/electronics10243091>.
- [19] Mudumba B, Kabir MF. Mine-first association rule mining: An integration of independent frequent patterns in distributed environments. *Decision Analytics Journal*. 2024 Mar 1; 10:100434. **DOI:** <https://doi.org/10.1016/j.dajour.2024.100434>.
- [20] Mohapatra D, Tripathy J, Mohanty KK, Nayak DS. Interpretation of optimized hyperparameters in associative rule learning using eclat and apriori. In 2021, the 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021 Apr 8 (pp. 879-882). IEEE. **DOI:** [10.1109/ICCMC51019.2021.9418049](https://doi.org/10.1109/ICCMC51019.2021.9418049).
- [21] Singla A, Gandhi P. A Comprehensive Study and Analysis of Frequent Itemsets Mining Algorithms using Diverse Real Datasets. In 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT) 2023 Nov 23 (pp. 174-180). IEEE. **DOI:** [10.1109/ICAICCIT60255.2023.10466004](https://doi.org/10.1109/ICAICCIT60255.2023.10466004).
- [22] Zhao Y, Lv Y, Zeng J, Dong Y, Fang H, Yu P, Xu S. Mining fault association rules in the perception layer of electric power sensor network based on improved Eclat. In 2021 International Wireless Communications and Mobile Computing (IWCMC) 2021 Jun 28 (pp. 418-423). IEEE. **DOI:** [10.1109/IWCMC51323.2021.9498688](https://doi.org/10.1109/IWCMC51323.2021.9498688).
- [23] Liu Y, Yang X, Liu K, Xu R, Pian Y, Liu S. Mining of dynamic traffic-meteorology-atmospheric pollutant association rules based on Eclat method. *Atmospheric Pollution Research*. 2024 Dec 1;15(12):102305. **DOI:** <https://doi.org/10.1016/j.apr.2024.102305>.
- [24] Dahiya V, Dalal S. EAHUIM: Enhanced absolute high utility itemset miner for big data. *International Journal of Information Management Data Insights*. 2022 Apr 1;2(1):100055. **DOI:** <https://doi.org/10.1016/j.jjime.2021.100055>.
- [25] Verma A, Dawar S, Kumar R, Navathe S, Goyal V. High-utility and diverse itemset mining. *Applied Intelligence*. 2021 Jul;51(7):4649-63. **DOI:** <https://doi.org/10.1007/s10489-020-02063-x>.
- [26] Singh K, Kumar R, Biswas B. High average-utility itemsets mining: a survey. *Applied Intelligence*. 2022 Mar;52(4):3901-38. **DOI:** <https://doi.org/10.1007/s10489-021-02611-z>.
- [27] Tung NT, Nguyen LT, Nguyen TD, Vo B. An efficient method for mining multi-level high utility itemsets. *Applied Intelligence*. 2022 Mar;52(5):5475-96. **DOI:** <https://doi.org/10.1007/s10489-021-02681-z>.

- [28] Wu X, Duan J, Pan Y, Li M. Medical knowledge graph: Data sources, construction, reasoning, and applications. *Big data mining and analytics*. 2023 Jan 26;6(2):201-17. **DOI:** [10.26599/BDMA.2022.9020021](https://doi.org/10.26599/BDMA.2022.9020021).
- [29] Mohssen HM, Safi HH. Chatbot in the E-Service of Mental Health Using the Reprogramming of the GPT-2 Model. *Journal of International Crisis and Risk Communication Research*. 2024;7(S9):984. **DOI:** 10.63278/jicrcr.vi.643.
- [30] Peng C, Xia F, Naseriparsa M, Osborne F. Knowledge graphs: Opportunities and challenges. *Artificial intelligence review*. 2023 Nov; 56(11):13071-102. **DOI:** <https://doi.org/10.1007/s10462-023-10465-9>.
- [31] Mohssen HM, Safi HH. An Efficient Mental Healthcare Chatbot Using Reprogramming Transformer Model Based on a Fine-Tuned GPT-2. In *International Conference on Data-Processing and Networking*. 2024 Oct 25 (pp. 69-83). Singapore: Springer Nature Singapore. **DOI:** https://doi.org/10.1007/978-981-96-3102-5_5.
- [32] Wang L, Guo Y, Guo Y, Xia X, Zhang Z, Cao J. An improved eclat algorithm based association rules mining method for failure status information and remanufacturing machining schemes of retired products. *Procedia CIRP*. 2023 Jan 1; 118:572-7. **DOI:** <https://doi.org/10.1016/j.procir.2023.06.098>.
- [33] Liu Y, Zhao L. Development of E-Commerce Recommendation System Based on FP-Growth Algorithm. In *2023 3rd International Signal Processing, Communications and Engineering Management Conference (ISPCEM)*. 2023 Nov 25 (pp. 378-383). IEEE. **DOI:** [10.1109/ISPCEM60569.2023.00074](https://doi.org/10.1109/ISPCEM60569.2023.00074).
- [34] Hassan MM, Karim A, Mollick S, Azam S, Ignatious E, Al Haque AF. An apriori algorithm-based association rule analysis to detect human suicidal behaviour. *Procedia Computer Science*. 2023 Jan 1; 219:1279-88. **DOI:** <https://doi.org/10.1016/j.procs.2023.01.412>.