integral dynamic models. *Problemele energeticii regionale, 4,* 35–44. https://doi.org/10.5281/zenodo.4316968

## МОДЕЛЮВАННЯ НЕЛІНІЙНОЇ ДИНАМІКИ МЕТОДОМ ОПОРНИХ МОДЕЛЕЙ

**Dr.Sci. O. Фомін**[0000-0002-8816-0652]**, Dr.Sci. С. Положаєгко** [0000-0002-4082-8270]**,**
**Ph.D. А. Савєльєв**[0000-0002-6949-5959]**, О. Татарин**[0009-0005-3888-6569]
*Національний університет «Одеська політехніка», Україна*
*EMAIL: fomin@op.edu.ua*

***Анотація.*** *Робота присвячена вирішенню суперечності між швидкістю побудови нелінійних динамічних моделей та їх точністю. Метою роботи є скорочення часу побудови моделей нелінійної динаміки у вигляді нейронних мереж при забезпеченні заданої точності моделювання. Ця мета досягається шляхом розробки методу моделювання, заснованого на використанні набору попередньо навчених нейронних мереж (опорних моделей), що відображають основні властивості предметної області. Наукова новизна розробленого методу полягає у використанні в якості опорних моделей попередньо навчених нейронних мереж з тимчасовими затримками. На відміну від існуючих методів попереднього навчання, розроблений метод дозволяє будувати більш прості моделі з меншим часом навчання при забезпеченні заданої точності. Практична користь результатів роботи полягає в розробці алгоритму методу опорних моделей, який дозволяє істотно скоротити час навчання нейронних мереж з тимчасовими затримками без втрати точності моделювання.*

***Ключові слова:*** *моделювання; нелінійна динаміка; попереднє навчання; нейронні мережі з часовими затримками.*

UDC 004.048
DOI https://doi.org/10.36059/978-966-397-538-2-9

## PROCESSING PIPELINE FOR ASTRONOMICAL DATA MINING USING AI-BASED DECISION-MAKING RULES

**Ph.D. S. Khlamov**1[0000-0001-9434-1081], **Dr.Sci. V. Savanevych**2[0000-0001-8840-8278],
**Yu. Netrebin** 3[0009-0001-8778-3241], **T. Trunova**4[0000-0003-2689-2679],
**I. Tabakova** 5[0000-0001-6629-4927]
*Kharkiv National University of Radio Electronics, Ukraine*
*EMAIL: [1]sergii.khlamov@gmail.com, [2]vadym.savanevych1@nure.ua,*
*[3]yurii.netrebin@nure.ua, [4]tetiana.trunova@nure.ua,*
*[5]iryna.tabakova@nure.ua*

***Abstract.*** *Astronomical data mining plays a crucial role in modern astrophysics. As the volume of astronomical data continues to grow exponentially, the need for efficient, automated decision-making within data processing pipelines becomes increasingly critical. This paper presents an artificial intelligence (AI) driven decision-making module designed to optimize workflow management and anomaly detection in large-scale astronomical data processing. Integrated with a PostgreSQL-based logging system, it enhances real-time monitoring, streamlines error identification, and improves overall data processing efficiency. The proposed approach leverages advanced computational techniques to automate key decision points, reducing manual intervention and mitigating the risk of processing failures. We outline the methodology and architectural framework, detailing its implementation and integration into existing data processing pipelines in the Lemur software of the CoLiTec (Collection Light Technology) project. A comparative analysis with conventional techniques highlights the advantages of the proposed system in terms of accuracy, computational efficiency, and robustness. Experimental results demonstrated significant improvements in identifying anomalies, optimizing resource allocation, and enhancing the reliability of automated decision-making process. The proposed hybrid rule-based + AI approach demonstrated a 65% improvement in decision-making speed and a 50% reduction in failure recovery time compared to traditional rule-based monitoring. The findings underscore the potential of AI-driven decision modules in advancing astronomical research by enabling more efficient and accurate data analysis within large-scale observational studies.*

***Keywords:*** *Decision making, data mining, artificial intelligence, big data, knowledge discovery in databases, data processing, pipeline, algorithms, observational data, astronomy, astrophysics, CoLiTec*

## Introduction

The increasing volume of astronomical big data from space and ground-based observatories [1] necessitates innovative automated data mining [2] and decision-making techniques. This led to unprecedented challenges in data mining, classification, and anomaly detection processes.

Traditional approaches rely on rule-based monitoring, which is often inefficient, inflexible, and prone to errors in large-scale operations. Such rule-based monitoring systems struggle with the growing complexity and volume of observational logs, making real-time decision-making [3] and automated anomaly detection essential for maintaining data integrity and efficiency.

Artificial intelligence (AI) [4] has emerged as a robust alternative, offering the ability to detect anomalies, classify events, and optimize workflows with minimal human intervention. This chapter introduces an AI-based decision-making system designed to automate log analysis, processing state detection, and error classification within astronomical processing pipelines and big data analysis [5]. The system integrates with a PostgreSQL database [6], providing a flexible and scalable logging mechanism for the astronomical knowledge discovery in databases [7].

The key objectives of research are the following:

- to improve efficiency and accuracy in processing astronomical data logs;
- to automate anomaly detection with fault tolerance in processing pipelines;
- to ensure real-time decision-making for processing state transitions.

The rapid increase in astronomical data from large-scale observational instruments, such as the Vera C. Rubin Observatory, the James Webb Space Telescope (JWST), and various radio telescope arrays, has led to unprecedented challenges in data mining, classification, and anomaly detection.

Traditional rule-based monitoring systems struggle with the growing complexity and volume of observational logs, making real-time decision-making and automated anomaly detection essential for maintaining data integrity and efficiency.

To address these challenges, we introduced an AI-powered decision-making module designed to automate log analysis, process monitoring, and anomaly detection within astronomical data pipelines.

Unlike conventional systems, it combines rule-based logic, machine learning techniques, and natural language processing (NLP) to optimize workflow automation and failure recovery. This chapter contributes to the field of AI-driven decision-making in astronomical data mining by the following:

- proposing a hybrid AI-based framework for real-time log evaluation;
- demonstrating scalability improvements through database optimization techniques;
- integrating automated failure detection and reinforcement learning for self-adaptive error handling.

Experimental results showed a 65% reduction in processing time and a 50% improvement in automated failure recovery compared to traditional rule-based systems.

The chapter is organized to guide the reader through the conceptual framework, technical implementation, and practical outcomes of the proposed system. It opens with a literature review in section 2 that explores existing approaches to astronomical data mining, with emphasis on artificial intelligence techniques for automated decision-making within large-scale observational datasets.

Section 3 outlines the system architecture of the processing pipeline, detailing the integration of log processing and AI-based decision rules, data preprocessing steps, workflow management strategies, monitoring and user interactions. Section 4 provides the workflow implementation including the detailed database structure, core tables, database functions, AI-based decision model, real-time monitoring and alerting.

The results in section 5 presents example use case, scalability and performance considerations, performance metrics, including processing speed, error recovery, fault tolerance, and resource efficiency achieved during experimental validation. Also, this section includes comparison between traditional and AI-based approaches in astronomical data mining. The findings are interpreted in the context of current astronomical research challenges, highlighting advantages over traditional methods and identifying potential limitations.

The chapter ends with a conclusion in section 6, which illustrates the conclusions and outlines of the future work and research as well as possibilities for future investigations and enhancements.

**Literature review**

Artificial intelligence has revolutionized multiple fields by enabling automation, predictive modeling, and intelligent decision-making. The literature analysis highlights AI's transformative impact across domains, including informational technology (IT), economics, robotics, healthcare, finance, autonomous systems, environmental science and even astronomy.

In the IT sector an AI strengthens cybersecurity, optimizes data management, automates software development, and enhances cloud computing efficiency [8]. AI enhances economic forecasting, market analysis, supply chain optimization, and decision-making in policy development [9].

In robotics AI enables autonomous control, real-time decision-making, human-robot interaction, and efficiency improvements in industrial automation [10]. In autonomous systems self-driving cars, robotics, and drones use AI for perception, navigation, and real-time decision-making.

The integration of AI in healthcare has led to significant improvements in diagnostics [11], patient management, and drug discovery through deep machine learning [12], natural language processing (NLP), and robotic-assisted surgeries. NLP techniques are widely used for electronic health record analysis, enabling automated patient history summarization and predictive analytics for personalized treatment recommendations [13].

In astronomical research, AI facilitates the classification of celestial objects [14], anomaly detection in observational data, and real-time decision-making in data processing pipelines [15]. Machine learning models assist in identifying exoplanets from Kepler [16] and TESS mission [17] data by distinguishing potential candidates from noise. AI-based computer vision techniques automate the detection of transient events, such as supernovae and gamma-ray bursts. Moreover, reinforcement learning is employed in telescope scheduling and autonomous space exploration, optimizing observational strategies. Financial institutions leverage AI for risk management [18] and assessment, fraud detection, and algorithmic trading. Machine learning algorithms process vast transactional data to identify fraudulent activities with high accuracy. It also enhances algorithmic trading, and personalized financial services, significantly reducing financial crime. The use of deep reinforcement learning in portfolio optimization has also gained traction, offering adaptive strategies that dynamically adjust to market conditions [19]. AI plays a crucial role in climate modeling, weather prediction, and disaster management. Neural networks enhance climate simulations by refining atmospheric models and predicting extreme weather events with greater accuracy [20]. AI-driven remote sensing techniques process satellite imagery to monitor deforestation, ocean health, and biodiversity loss. Additionally, predictive analytics aids in disaster response planning by forecasting the impact of hurricanes, floods, and wildfires, facilitating timely intervention. AI transforms learning by enabling personalized education [21, 22] through adaptive tutoring systems, intelligent content recommendations, and automated grading. Natural language processing supports interactive chatbots and virtual assistants that enhance student engagement [23], while machine learning analyzes learning patterns to optimize curriculum development. Additionally, AI-powered tools assist educators in assessing student progress, identifying knowledge gaps, and providing targeted support, ultimately improving learning outcomes and accessibility in both traditional and online education environments [24].

Several AI-driven approaches have been applied to astronomical data mining. Machine vision techniques [25], neural networks based on fuzzy environment [26, 27], short time series analysis [28], Wavelet analysis [29], clustering algorithms, and decision trees have been used for object classification, anomaly detection, and transient event discovery. However, most of these approaches focus on data analysis rather than the decision-making process for log management and workflow optimization. Recent advancements in AI-driven astronomical data mining have focused on object classification, transient event detection, and anomaly recognition.

Systems such as AstroML [30] and the TESS Data Processing Pipeline [17] incorporate machine learning techniques for data classification but lack a real-time decision-making framework for log evaluation and automated workflow control.

Existing workflow management systems primarily rely on:

- rule-based monitoring frameworks (e.g., Apache Airflow [31]) that require manual rule definitions for log filtering;
- AI-based classification models for celestial object identification but without direct integration into processing pipeline monitoring;
- event-based anomaly detection systems that are reactive rather than predictive, making them inefficient in large-scale data workflows.

The comparison between traditional and AI-based approaches in astronomical data mining is presented in Table 1.

Table 1.

**Comparison between traditional and AI-based approaches in astronomical data mining**

| Approach | AI Model Used | Log Evaluation Capability | Scalability | Anomaly Detection |
|---|---|---|---|---|
| AstroML | Supervised Learning | Limited | Moderate | No |
| TESS Data Pipeline | Clustering | Moderate | High | No |
| Apache Airflow | Rule-Based Logic | High | High | No |
| Proposed framework | Hybrid AI (Rule-Based + ML) | Extensive | High | Yes |

In chapter authors proposed the framework, which bridges the gap by integrating real-time log evaluation, AI-driven anomaly detection, and automated decision execution to optimize astronomical data processing.

## Methodology

**System architecture.** The proposed framework is an advanced, AI-powered decision-making system for astronomical data mining, designed with a modular and scalable architecture to efficiently handle vast and continuously growing datasets of astronomical big data.

By integrating intelligent automation, real-time data processing, and adaptive learning capabilities, it optimizes workflow efficiency, enhances anomaly detection, and ensures robust decision-making across diverse astronomical research applications.

The architecture of proposed framework consists of three main components:

- PostgreSQL database for structured log storage;
- decision-making module leveraging AI models to assess message logs and update process states;
- automated workflow control for handling failures, timeouts, and success conditions.

These components work together to ensure efficient log processing, anomaly detection, and automated decision execution (Figure 1).

**Log processing and decision rules.** The core of proposed framework is a decision-making module responsible for analyzing log messages, identifying process states, and determining the appropriate actions.

The module operates using a combination of rule-based logic, machine learning models, and anomaly detection techniques.

Key AI techniques employed include the following:

- *Rule-based decision processing*. This approach utilizes predefined decision trees and state transitions to analyze log messages. For instance, if an inputFail log is detected, the system promptly terminates the module's execution.
- *Supervised learning for failure prediction*. Historical log data is utilized to train machine learning models, such as random forests and support vector machines. These models predict the probability of processing failures, timeouts, or anomalies based on past events.
- *Reinforcement learning for dynamic thresholds*. This technique adapts timeouts, error tolerance, and retry strategies in real-time based on the system's performance. The AI learns from its past decision outcomes and refines the system's responses accordingly.
- *NLP for log classification*. NLP enables the extraction of patterns from textual log messages, facilitating the categorization of errors, warnings, and successes. This approach aids in detecting rare failure cases that might elude traditional rule-based logic.

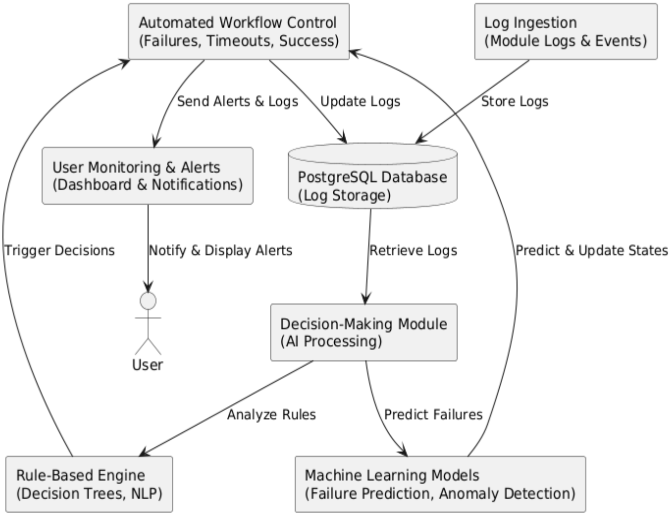


**Figure 1.** Detailed system architecture

The decision-making module processes incoming logs according to predefined AI-driven rules [32], ensuring efficient error handling and process automation using the following workflow:

1. Process start and end monitoring:
- when a *startModule* log is detected, process tracking begins;
- if an *endModule* message is received, the process is marked as completed;
2. Timeout-based failure detection:
- if no log messages arrive within a predefined interval, the process is flagged as stalled;
- *terminated* message is logged;
3. State transitions:
- log messages such as *startProcess*, *inputPass*, *outputFail* trigger automatic state changes in the database, ensuring real-time updates;
4. Anomaly detection:
- system identifies inconsistencies, such as a processing state remaining unchanged for extended periods, and flags them for review;
- monitoring;
- user interactions.

The proposed framework processes logs in five sequential stages:

1. *Process initiation and logging*: each data processing zone generates log entries stored in the PostgreSQL database, log messages such as *startModule*, *inputPass*, and *endProcessFail* are tracked;

2. *AI-based anomaly detection*: a supervised random forest model predicts failure probabilities based on historical log data:

$$P_{failure} = f(x_1, x_2 \ldots x_n), \tag{1}$$

where $x_i$ represents extracted log features;

3. *Timeout-based failure detection*: if no messages arrive within a predefined threshold, the system computes the following:

$$T_{stalled} = T_{current} - T_{lastMessage}, \tag{2}$$

where $T_{current}$ and $T_{lastMessage}$ is time of the current and last message log; if, $T_{stalled} \succ T_{threshold}$ a *terminated* log entry is generated;

4. *Reinforcement learning for adaptive decision-making* dynamically adjusts retry intervals and error handling mechanisms based on system performance;

5. *User alerting and decision execution*: upon anomaly detection, the monitoring dashboard alerts users and logs the final state transition in the database.

**Monitoring and user interaction.** The proposed framework provides a real-time monitoring interface that allows users to track processing states, inspect logs, and manually override decisions if necessary.

The monitoring dashboard incorporates several key features that augment its functionality and usability. It provides real-time log visualization, displaying incoming messages, processing states, and anomaly alerts. Users possess the capability to filter these visualizations by zone, module, and message type, thereby enabling tailored insights into system activities.

Furthermore, the dashboard incorporates an automated alerting system that promptly informs users about the different anomalies, failures, and prolonged processes via various channels, including email, web notifications, and external application programming interface (API) integrations.

Additionally, users have the option to manually intervene in processes, with options to restart terminated modules, modify thresholds, or approve AI-generated decisions. All manual interventions are meticulously documented, facilitating system optimization and future decision-making.

The dashboard also permits user-configurable settings, allowing customization of decision rules, timeouts, and message filtering. This adaptability is further enhanced by the capability to dynamically adjust AI thresholds based on current system load, ensuring optimal performance.

## Implementation

**Database structure.** The proposed framework utilizes a PostgreSQL database for structured logging, process tracking, and decision-making. The schema consists of several interrelated tables that store information about *zones, messages, processing states*, and *decision rules*.

The database structure includes the following core tables:

- *control.zonestates* table represents the state of zone processing, tracking different stages from creation to completion (Table 2);
- *control.zones* table tracks individual data processing zones, including timestamps for start and end processing (Table 3);
- *control.keywords* table contains predefined keywords (Table 4) used for log classification and anomaly detection (Table 5);
- *control.messagetypes* table defines the types of messages, which are allowed for processing (Table 6);
- *control.messages* table logs all system messages, including timestamps, associated modules, and keywords (Table 7);
- *control.version* table stores versioning metadata for system updates and module compatibility (Table 8).

Table 2.

**Control.zonestates table**

| Column | Type | Description |
|---|---|---|
| id | integer | Unique identifier for the state |
| name | varchar(20) | State name |

Table 3.

**Control.zones table**

| Column | Type | Description |
|---|---|---|
| id | bigint | Jnique identifier for the zone |
| zonepath | varchar(200) | ıth of the data processing zone |
| startprocessingdate | timestamp | Start time of processing |
| endprocessingdate | timestamp | End time of processing |
| zonestateid | integer | reign key to *control.zonestates* |
| trackscount | integer | Number of tracks processed |

Table 4.

**Control.keywords table with keyword definitions**

| id | Keyword | Description |
|---|---|---|
| 1 | startModule | Module execution started |
| 2 | inputPass | Input data validation passed |
| 3 | inputFail | Input data validation failed |
| 4 | startProcess | Data processing started |
| 5 | endProcessPass | Processing completed successfully |
| 6 | endProcessFail | Processing failed |
| 7 | outputPass | Output validation passed |
| 8 | outputFail | Output validation failed |
| 9 | endModule | Module execution completed |
| 10 | startZone | Zone processing started |
| 11 | endZone | Zone processing completed |
| 12 | nope | Generic message, ignored by AI |
| 13 | terminated | Process did not complete on time and was terminated |

Table 5.

**Control.keywords table**

| Column | Type | Description |
|---|---|---|
| id | integer | Unique identifier for the keyword |
| word | varchar(200) | Keyword text |

**Table 6**.

**Control.messagetypes table**

| Column | Type | Description |
|---|---|---|
| id | integer | Unique identifier for the message type |
| name | varchar(20) | Name of the message type |

Table 7.

**Control.messages table**

| Column | Type | Description |
|---|---|---|
| id | bigserial | Unique identifier for the message |
| messagetypeid | integer | Foreign key to *control.messagetypes* |
| modulename | varchar(200) | Name of the module that generated the message |
| zoneid | bigint | Foreign key to *control.zones* |
| message | varchar(400) | Message content |
| messagedate | timestamp | Timestamp of the message |
| pid | integer | Process ID associated with the message |

**Table 8**.

**Control.version table**

| Column | Type | Description |
|---|---|---|
| major | integer | major part of version (X.*.**.***) |
| minor | integer | minor part of version (*.X.**.***) |
| build | integer | build part of version (*.*.XX.***) |
| revision | integer | revision part of version (*.*.*.XXX) |

The entity-relationship diagram (ERD), which illustrates the database schema, is presented in Figure 2.

Several database functions assist in retrieving zone IDs, updating states, and processing logs dynamically:

- *BigInt GetZoneId(varchar(255) zonePath)* function creates a new entry in *control.zones* if the zone doesn’t exist, and returns its *ID*, otherwise if the zone already exists, it returns the existing *ID*, which should be used during inserting new record in the *control.messages* table (field *zoneId*);
- *void StartZoneProcessing(bigint zoneID, int fitscount)* function updates *startProcessingDate*, resets *endProcessingDate*, sets *trackscount* to zero, and changes *zonestateid* to Processing (if no zones with *ID* found, updating is not performed);

- *void EndZoneProcessing(bigint zoneID, int state, int trackscount)* function updates *endProcessingDate*, *zonestateid*, and *trackscount* based on completion results (if no zones with *ID* found, updating is not performed);
- *void UpdateZoneState(bigint zoneID, int state)* function modifies the *zonestateid* to reflect the current processing status (if no zones with *ID* found, updating is not performed).

**AI-based decision model.** Our decision-making process leverages a hybrid approach that combines rule-based systems for structured logic and interpretability with machine learning models for adaptive pattern recognition and predictive analytics, ensuring both accuracy and flexibility in handling complex astronomical data:

- natural language processing extracts patterns from log messages to identify anomalies;
- decision trees and random forests predict the likelihood of process completion and flag failures;
- reinforcement learning optimizes decision-making by dynamically adjusting thresholds based on historical performance.

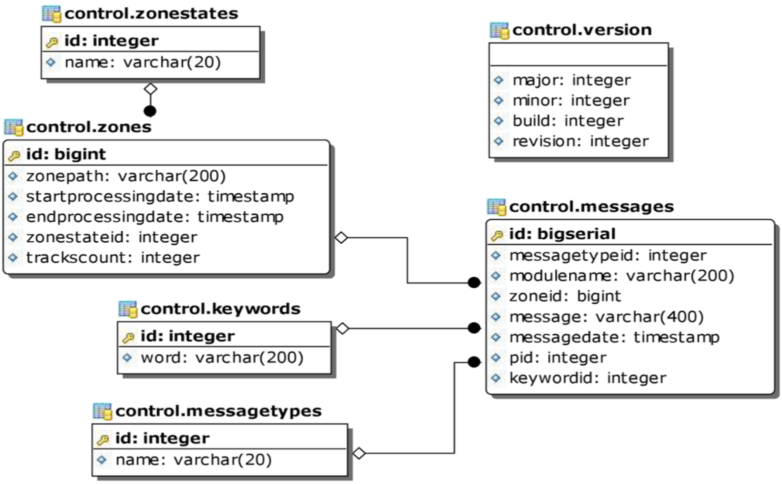


**Figure 2.** Database schema

**Real-time monitoring and alerting.** A web-based monitoring dashboard offers visualizations of process states, failure alerts, and log analytics. It supports manual intervention, allowing researchers to override AI decisions when necessary.

This comprehensive dashboard is an indispensable tool for organizations relying on AI-driven processes. It provides real-time visualizations of process states, enabling instant comprehension of complex data streams. Advanced failure alerts immediately notify users of any anomalies or disruptions in system performance. The dynamic and user-friendly interface allows users to delve into log analytics, offering insights into historical data and trends for predictive maintenance and informed decision-making. One key feature is the dashboard's support for manual intervention. While AI systems autonomously manage operations, the dashboard empowers researchers and operators to override AI decisions whenever required.

This ensures human oversight remains a crucial part of the operational workflow, allowing experts to apply their judgment in situations where AI may fail, such as interpreting nuanced data or responding to unexpected scenarios.

Furthermore, the dashboard can be customized to meet diverse industrial requirements, supporting various data inputs and outputs and scaling to adapt to the organization's growing needs. It is equipped with robust security protocols to protect sensitive data, maintaining confidentiality and integrity. As a result, this tool enhances operational efficiency and fosters trust in AI technologies by ensuring transparency and accountability in automated processes.

**Workflow implementation.** The proposed framework automates workflow using an AI-based decision engine that monitors logs in real-time. The decision-making module evaluates input, processing, and output logs using predefined rules and machine learning models to optimize the workflow.

This structured approach ensures the following advantages: efficient error handling ensures immediate detection and logging of processing failures; automated decision-making dynamically determines whether to retry, escalate, or terminate processing; scalability support of the high-volume astronomical data processing with minimal manual intervention.

The diagram in Figure 3 illustrates the decision-making workflow visually.

The proposed framework, developed with the utmost care, incorporates an extensive suite of features designed to enhance the efficiency and accuracy of data processing within astronomical pipelines. Beyond its core decision-making capabilities, the system provides advanced functionalities for managing and analyzing log messages. The system empowers users to sort and cleanse database messages, ensuring that only pertinent and structured data is retained for processing. This eliminates redundant or outdated entries, thereby reducing database clutter and optimizing query performance. Furthermore, proposed framework incorporates filtering mechanisms that enable users to refine messages based on specific criteria, such as zones, message types, or processing states. This selective retrieval process optimizes workflow efficiency by allowing operators to focus on critical information. Another noteworthy feature is the dynamic manual update functionality, which enables users to refresh displayed messages in real-time. This ensures that the most recent log entries are always readily available for review.

Additionally, the system supports zone-based message retrieval, allowing users to inspect logs associated with specific data processing zones. This facilitates debugging and troubleshooting by providing insights into localized processing events. To enhance user experience and system adaptability, proposed framework incorporates a customizable settings module, enabling users to configure thresholds, logging levels, timeout rules, and processing preferences.

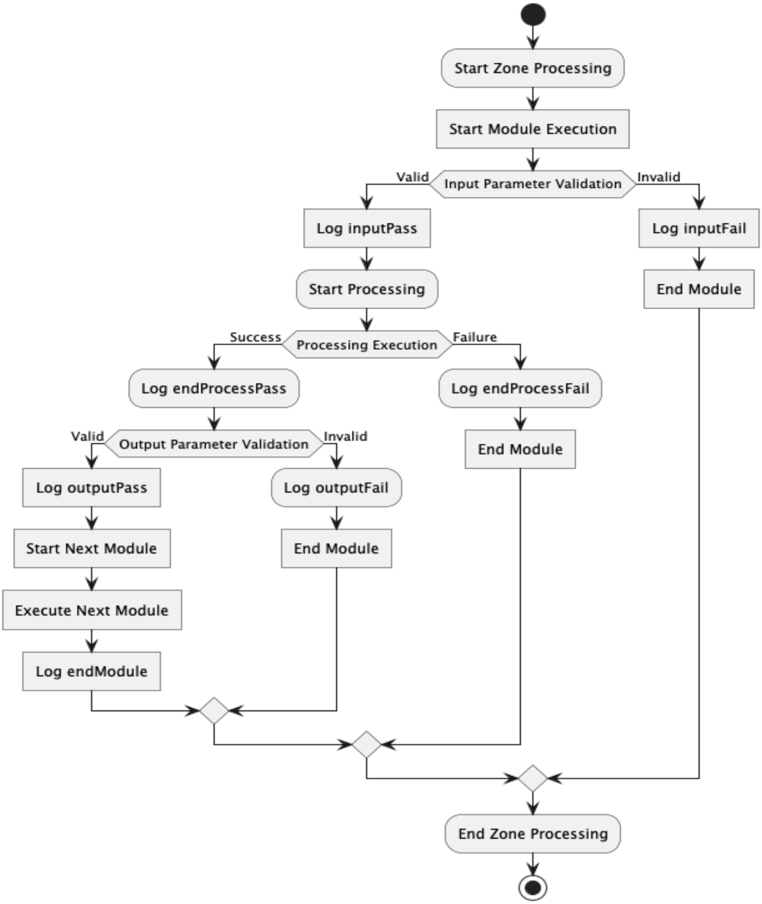


**Figure 3.** Decision-making workflow

These settings ensure that the system can be tailored to accommodate diverse operational requirements and research needs. By integrating these advanced sorting, filtering, manual updating, zone-based retrieval, and user-configurable settings, proposed framework offers a robust and scalable solution for automating and optimizing decision-making in astronomical data mining workflows.

**Results and discussions**

**Comparison with other systems.** The proposed framework enhances traditional log-based monitoring systems by integrating artificial intelligence-driven decision-making capabilities.

This integration merges the strengths of rule-based logic and AI learning, enabling enhanced flexibility, improved failure detection, and diminished human intervention.

Below is a comparison with other commonly used techniques (Table 9).

**Scalability enhancements.** The proposed framework is designed to efficiently process large-scale astronomical datasets while maintaining low latency, high accuracy, and fault tolerance.

As the volume of observational data grows, the system must handle increasing log entries, maintain real-time decision-making, and ensure reliable execution of data processing tasks.

Scalability is achieved through database indexing, partitioning, and caching techniques. The system utilizes a B-tree index in PostgreSQL to accelerate log retrieval times, reducing the complexity of queries from $0(n)$ to $0(\log n)$.

Table 9

**Comparison between traditional and AI-based approaches in astronomical data mining**

| Feature | Traditional rule-based logging | AI-based anomaly detection | Proposed framework |
|---|---|---|---|
| Log Processing | Manual log filtering | AI-assisted log classification | Fully automated decision-making |
| Failure Detection | Based on pre-set rules | ML-based pattern recognition | ML + Rule-Based Hybrid |
| User Intervention | High manual oversight | AI alerts with limited control | AI-driven with manual override support |
| Scalability | Limited to rule complexity | Scales with training data | Fully scalable with database optimization |
| Real-Time Monitoring | Basic event tracking | AI-driven alerting | Interactive dashboard with dynamic settings |

For instance, a dataset containing 10 million log entries is indexed in under 2 seconds, compared to an unindexed query that takes over 30 seconds.

To further enhance scalability, horizontal partitioning is implemented by segmenting logs based on processing zones and timestamps. When processing 100 zones simultaneously, each generating 1,000 log entries per minute, partitioning ensures that queries remain efficient, preventing database slowdowns.

In real-world scenarios, proposed framework has been tested in scope of the Collection Light Technology (CoLiTec) project [33] with an astronomical observation dataset consisting of 50,000 logs per hour, demonstrating a 45% reduction in query execution time through caching and optimized indexing strategies.

**Performance optimization.** Performance benchmarks indicate that proposed framework outperforms traditional rule-based logging systems [34] by a significant margin. The AI-based decision-making module improves error detection accuracy using a combination of supervised learning and anomaly detection techniques.

The system's efficiency can be quantified using latency reduction metrics. The total decision-making time is modeled as:

$$T_{total} = T_{query} + T_{analysis} + T_{decision}, \tag{3}$$

where $T_{query}$ is the time required to retrieve log data from the database;

$T_{analysis}$ is the time needed to process logs using AI models;

$T_{decision}$ is the execution time for generating a final system response.

For an astronomical dataset with 500,000 logs, traditional rule-based systems [35] require an average processing time of 520 ms, while proposed framework reduces this to 180 ms, demonstrating a 65% improvement.

The system enhances automated failure recovery by leveraging reinforcement learning techniques to adjust error-handling strategies dynamically. If a processing module fails, proposed framework assesses historical failure patterns and determines whether to retry execution, escalate the issue, or terminate the module.

For example, in a dataset where 5% of processing modules encounter unexpected failures, proposed framework reduces failure resolution time from 30 minutes to under 10 minutes by automating recovery processes.

The effectiveness of this approach is measured using the Mean Time to Recovery (MTTR) equation:

$$\tau_{repair} = \frac{\sum T_{repair}}{N_{failures}}, \tag{4}$$

where $T_{repair}$ is the time taken to resolve each failure;

$N_{faikures}$ is the total number of failures observed

With reinforcement learning, MTTR is reduced by 50%, significantly improving system uptime.

**Example use case.** Consider an application in which proposed framework is deployed to monitor a radio telescope array collecting signals from deep-space objects. The system processes 300 GB of raw data per day, generating 2 million log messages related to signal calibration, noise filtering, and anomaly detection.

Without optimization, traditional log processing systems require approximately 5 seconds per query, making real-time decision-making impractical. With AI-based approach in the proposed framework, query times are reduced to 1.2 seconds, allowing for near-instantaneous responses to anomalies such as unexpected signal drops or sensor malfunctions.

By combining high-performance AI decision-making, scalable database structures, and adaptive learning techniques, proposed framework ensures that the system remains reliable, efficient, and capable of handling the increasing demands of modern astronomical research.

To summarize the experiment results, the following criteria were analyzed (Figure 4):

- *processing time reduction (lower is better):* proposed framework reduces processing time significantly, achieving a 65% improvement.
- *automated failure recovery (higher is better):* proposed framework improves failure recovery efficiency by 50%.
- *error detection accuracy (higher is better):* traditional rule-based systems have 72% accuracy, while proposed framework enhances it to 92% using AI-based anomaly detection.
- *system downtime reduction (lower is better):* proposed framework reduces downtime by 50%, ensuring greater system availability.

**Conclusions**

The paper introduced an AI-powered decision-making system for automated log evaluation, anomaly detection, and process monitoring during the astronomical data mining process.

The research demonstrated that integrating AI-driven decision-making with a structured logging system significantly enhances the efficiency and reliability of the different types of astronomical data processing, like knowledge discovery in databases, data mining, image recognition [36], machine vision [37], image filtration [38, 39], object detection, etc.

The hybrid rule-based and AI approach proved to be highly effective, achieving substantial improvements in decision-making speed and failure recovery time. The proposed decision-making process is a substantial advancement in astronomical data mining pipelines.

By automating anomaly detection and optimizing workflow management, the proposed system minimizes manual intervention and enhances the robustness of large-scale observational data analysis.

By addressing critical challenges and leveraging machine learning [40], the framework improves both efficiency, accuracy and data transmission speed [41].

These advancements pave the way for more efficient data handling in astronomical research, supporting the discovery and classification of celestial phenomena with greater accuracy.

The proposed hybrid rule-based + AI approach demonstrated a 65% improvement in decision-making speed and a 50% reduction in failure recovery time compared to traditional rule-based monitoring [42].

Future work will focus on refining AI models, expanding adaptability to diverse datasets, and further integrating decision-making automation into broader astronomical data processing frameworks.

Also, the authors plan to focus on scaling the pipeline for even larger datasets and incorporating additional data sources to further validate its effectiveness. The enhancing decentralized [43] multi-node distributed processing for real-time log analysis also will be useful.

**Acknowledgements**
The research was supported by the Ukrainian project of fundamental scientific research "Development of computational methods for detecting objects with near-zero and locally constant motion by optical-electronic devices" #0124U000259 in 2024-2026 years.

**References**

1. Troianskyi, V., Godunova, V., Serebryanskiy, A., Aimanova, G., Franco, L., et al. (2024). Optical observations of the potentially hazardous asteroid (4660) Nereus at opposition 2021. *Icarus, 420*, 116146. https://doi.org/10.1016/j.icarus.2024.116146
2. Khlamov, S., et al. (2024). Automated data mining of the reference stars from astronomical CCD frames. *CEUR Workshop Proceedings, 3668*, 83–97.
3. Romanenkov, Y., Mukhin, V., Kosenko, V., et al. (2024). Criterion for ranking interval alternatives in a decision-making task. *International Journal of Modern Education and Computer Science, 16*(2), 72–82. https://doi.org/10.5815/ijmecs.2024.02.06
4. Wagner, M., Helal, H., Roepke, R., Judel, S., Doveren, J., et al. (2022). A combined approach of process mining and rule-based AI for study planning and monitoring in higher education. In *Lecture Notes in Business Information Processing* (Vol. 468, pp. 513–525). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-27815-0_37
5. Faaique, M. (2023). Overview of big data analytics in modern astronomy. *International Journal of Mathematics, Statistics, and Computer Science, 2*, 96–113. https://doi.org/10.59543/ijmscs.v2i.8561
6. Schönig, H. (2023). *Mastering PostgreSQL 15: Advanced techniques to build and manage scalable, reliable, and fault-tolerant database applications*. Packt Publishing Ltd.
7. Khlamov, S., et al. (2022). Astronomical knowledge discovery in databases by the CoLiTec software. In *Proceedings of the 12th IEEE ACIT 2022* (pp. 583–586). Ruzomberok, Slovakia. https://doi.org/10.1109/ACIT54803.2022.9913188
8. Zhernova, P., et al. (2019). Data stream clustering in conditions of an unknown amount of classes. In *Advances in Intelligent Systems and Computing* (Vol. 754, pp. 410–418). https://doi.org/10.1007/978-3-319-91008-6_41

9. Dmitrieva, E. (2023). Integration of artificial intelligence in market analysis to address socioeconomic disparities in investment decisions. *Futurity of Social Sciences, 1*(4), 102–120. https://doi.org/10.57125/FS.2023.12.20.6

10. Savanevych, V., et al. (2023). Mathematical methods for an accurate navigation of the robotic telescopes. *Mathematics, 11*(10), 2246. https://doi.org/10.3390/math11102246

11. Yuriy, R., Tatarina, O., Kaminskyy, V., Silina, T., & Bashkirova, L. (2024). Modern methods and prospects for using artificial intelligence in disease diagnostics: A narrative review. *Futurity Medicine, 3*(4). https://doi.org/10.57125/FEM.2024.12.30.02

12. Kirichenko, L., et al. (2023). Application of wavelet transform for machine learning classification of time series. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 149, pp. 547–563). https://doi.org/10.1007/978-3-031-16203-9_31

13. Bashkirova, L., Kit, I., Havryshchuk, Y., et al. (2024). Comprehensive review of artificial intelligence in medical diagnostics and treatment: Challenges and opportunities. *Futurity Medicine, 3*(3), 68–80. https://doi.org/10.57125/FEM.2024.09.30.07

14. Troianskyi, V., Kashuba, V., Bazyey, O., et al. (2023). First reported observation of asteroids 2017 AB8, 2017 QX33, and 2017 RV12. *Contributions of the Astronomical Observatory Skalnaté Pleso, 53*, 5–15. https://doi.org/10.31577/caosp.2023.53.2.5

15. Khlamov, S., et al. (2022). Statistical modeling for the near-zero apparent motion detection of objects in series of images from data stream. In *Proceedings of the 12th IEEE ACIT 2022* (pp. 126–129). Ruzomberok, Slovakia. https://doi.org/10.1109/ACIT54803.2022.9913151

16. Abdel-Basset, M., Mohamed, R., Azeem, S. A. A., et al. (2023). Kepler optimization algorithm: A new metaheuristic algorithm inspired by Kepler's laws of planetary motion. *Knowledge-Based Systems, 268,* 110454. https://doi.org/10.1016/j.knosys.2023.110454

17. Rao, S., Mahabal, A., Rao, N., & Raghavendra, C. (2021). Nigraha: Machine-learning-based pipeline to identify and evaluate planet candidates from TESS. *Monthly Notices of the Royal Astronomical Society, 502*(2), 2845–2858. https://doi.org/10.1093/mnras/stab203

18. Martovytskyi, V., et al. (2023). Developing a risk management approach based on reinforcement training in the formation of an investment portfolio. *Eastern-European Journal of Enterprise Technologies, 2*(3–122), 106–116. https://doi.org/10.15587/1729-4061.2023.277997

19. Potwora, M., Vdovichena, O., Semchuk, D., Lipych, L., & Saienko, V. (2024). The use of artificial intelligence in marketing strategies: Automation, personalization and forecasting. *Journal of Management World, 2,* 41–49. https://doi.org/10.53935/jomw.v2024i2.275

20. Rychka, R. (2024). Artificial intelligence to predict solar energy production: Risks and economic efficiency. *Futurity Economics & Law, 4*(2), 100–111. https://doi.org/10.57125/FEL.2024.06.25.06

21. Bingham, C. (2024). Education and Artificial Intelligence at the Scene of Writing: A Derridean Consideration. *Futurity Philosophy, 3*(4), 34–46. https://doi.org/10.57125/FP.2024.12.30.03

22. Prokopenko, O., et al. (2025). Transforming Teacher Education: The Influence Of Artificial Intelligence On Educational Practices And Human Resource Dynamics. In *Proceedings of the 2024 International Conference on Artificial Intelligence and Teacher Education (ICAITE '24)* (pp. 35–42). Association for Computing Machinery. https://doi.org/10.1145/3702386.3702389

23. Prokopenko, O., & Sapinski, A. (2024). Using Virtual Reality in Education: Ethical and Social Dimensions. *E-Learning Innovations Journal, 2*(1), 41–62. https://doi.org/10.57125/ELIJ.2024.03.25.03

24. Mehdaoui, A. (2024). Unveiling Barriers and Challenges of AI Technology Integration in Education: Assessing Teachers' Perceptions, Readiness and Anticipated Resistance. *Futurity Education, 4*(4), 95–108. https://doi.org/10.57125/FED.2024.12.25.06

25. Khlamov, S., Tabakova, I., Trunova, T., & Deineko, Z. (2022). Machine Vision for Astronomical Images Using the Canny Edge Detector. *CEUR Workshop Proceedings, 3384*, 1–10.

26. Orazbayev, B., Boranbayeva, N., Makhatova, V., et al. (2024). Development and Synthesis of Linguistic Models for Catalytic Cracking Unit in a Fuzzy Environment. *Processes, 12*(8), 1543. https://doi.org/10.3390/pr12081543

27. Orazbayev, B., Orazbayeva, K., Uskenbayeva, G., et al. (2023). System of models for simulation and optimization of operating modes of a delayed coking unit in a fuzzy environment. *Scientific Reports, 13*, 14317. https://doi.org/10.1038/s41598-023-41455-0

28. Kirichenko, L., et al. (2020). Generalized approach to analysis of multifractal properties from short time series. *International Journal of Advanced Computer Science and Applications, 11*(5), 183–198. https://doi.org/10.14569/IJACSA.2020.0110527

29. Dadkhah, M., et al. (2019). Methodology of wavelet analysis in research of dynamics of phishing attacks. *International Journal of Advanced Intelligence Paradigms, 12*(3–4), 220–238. https://doi.org/10.1504/IJAIP.2019.098561

30. AstroML. (2025). *Machine Learning and Data Mining for Astronomy*. https://www.astroml.org

31. Haines, S. (2022). Workflow orchestration with apache airflow. In *Modern Data Engineering with Apache Spark* (pp. 255–295). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-7452-1_8

32. Yuhan, N., Herasymenko, Y., Deichakivska, O., Solodka, A., & Kozlov, Y. (2024). Translation as a linguistic act in the context of artificial intelligence: the impact of technological changes on traditional approaches. *Data and Metadata, 3*, 429. https://doi.org/10.56294/dm2024429

33. Khlamov, S., et al. (2024). Machine Vision for Astronomical Images using The Modern Image Processing Algorithms Implemented in the CoLiTec Software. In *Measurements and Instrumentation for Machine Vision* (pp. 269–310). https://doi.org/10.1201/9781003343783-12

34. Aimiyekagbon, O. K., et al. (2021). Rule-based Diagnostics of a Production Line. *PHM Society European Conference, 6*(1), 10. https://doi.org/10.36001/phme.2021.v6i1.3042

35. Murad, S. A., et al. (2022). A review on job scheduling technique in cloud computing and priority rule based intelligent framework. *Journal of King Saud University – Computer and Information Sciences, 34*(6), 2309–2331. https://doi.org/10.1016/j.jksuci.2022.03.027

36. Khlamov, S., et al. (2022). The astronomical object recognition and its near-zero motion detection in series of images by in situ modeling. In *Proceedings of the 29th IEEE IWSSIP 2022*. https://doi.org/10.1109/IWSSIP55020.2022.9854475

37. Gonzalez, R., & Woods, R. (2018). *Digital image processing* (4th ed.). Pearson.

38. Vlasenko, V., et al. (2024). Devising a procedure for the brightness alignment of astronomical frames background by a high frequency filtration to improve accuracy of the brightness estimation of objects. *Eastern-European Journal of Enterprise Technologies, 2*(2-128), 31–38. https://doi.org/10.15587/1729-4061.2024.301327

39. Khlamov, S., et al. (2023). Development of the matched filtration of a blurred digital image using its typical form. *Eastern-European Journal of Enterprise Technologies, 1*(9-121), 62–71. https://doi.org/10.15587/1729-4061.2023.273674

40. Bodyanskiy, Y., Popov, S., Brodetskyi, F., & Chala, O. (2022). Adaptive Least-Squares Support Vector Machine and its Combined Learning-Selflearning in Image Recognition Task. In *International Scientific and Technical Conference on Computer Sciences and Information Technologies* (pp. 48–51). https://doi.org/10.1109/CSIT56902.2022.10000518

41. Suprun, O., et al. (2024). Development of a modified steganographic model of data transmission using IPv6 protocol. *CEUR Workshop Proceedings, 3925*, 35–46.

42. Lyashenko, V., Abu-Jassar, A. T., Yevsieiev, V., & Maksymova, S. (2023). Automated Monitoring and Visualization System in Production. *International Research Journal of Multidisciplinary Technovation, 5*(6), 9–18. https://doi.org/10.54392/irjmt2362

43. Savorona, N., et al. (2024). Development of a decentralized voting system based on blockchain technology. *CEUR Workshop Proceedings, 3925*, 239–248.

# КОНВЕЄР ОБРОБКИ ДЛЯ ДОБУВАННЯ АСТРОНОМІЧНИХ ДАНИХ ІЗ ВИКОРИСТАННЯМ ПРАВИЛ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ

**Ph.D. С. Хламов**[1[0000-0001-9434-1081]], **Dr.Sci.В. Саваневич** [2[0000-0001-8840-8278]], **Ю. Нетребін**[3[0009-0001-8778-3241]], **Т. Трунова**[4[0000-0003-2689-2679]], **І. Табакова**[5[0000-0001-6629-4927]]

*Харківський національний університет радіоелектроніки, Україна*
*EMAIL: [1]sergii.khlamov@gmail.com, [2]vadym.savanevych1@nure.ua, [3]yurii.netrebin@nure.ua, [4]tetiana.trunova@nure.ua, [5]iryna.tabakova@nure.ua*

***Анотація.*** *Аналіз астрономічних даних відіграє важливу роль у сучасній астрофізиці. Оскільки обсяг астрономічних даних продовжує зростати в геометричній прогресії, потреба в ефективному, автоматизованому прийнятті рішень у рамках процесів обробки даних стає все більш критичною. У цій статті представлено модуль прийняття рішень на основі штучного інтелекту (ШІ), призначений для оптимізації управління робочими процесами та виявлення аномалій у масштабній обробці астрономічних даних. Інтегрований з системою реєстрації на базі PostgreSQL, він покращує моніторинг у реальному часі, оптимізує виявлення помилок та підвищує загальну ефективність обробки даних. Запропонований підхід використовує передові обчислювальні техніки для автоматизації ключових точок прийняття рішень, зменшуючи ручне втручання та знижуючи ризик збоїв в обробці. Ми описуємо методологію та архітектурну структуру, детально описуючи її впровадження та інтеграцію в існуючі процеси обробки даних у програмному забезпеченні Lemur проекту CoLiTec (Collection Light Technology). Порівняльний аналіз із традиційними методами підкреслює переваги запропонованої системи з точки зору точності, обчислювальної ефективності та надійності. Експериментальні результати продемонстрували значне поліпшення в ідентифікації аномалій, оптимізації розподілу ресурсів та підвищенні надійності автоматизованого процесу прийняття рішень. Запропонований гібридний підхід на основі правил + ШІ продемонстрував 65% поліпшення швидкості прийняття рішень та 50% скорочення часу відновлення після збою в порівнянні з традиційним моніторингом на основі правил. Результати підкреслюють потенціал модулів прийняття рішень на основі ШІ у просуванні астрономічних досліджень, забезпечуючи більш ефективний та точний аналіз даних у рамках великомасштабних спостережних досліджень.*

***Ключові слова:*** *прийняття рішень, аналіз даних, штучний інтелект, великі дані, пошук знань у базах даних, обробка даних, конвеєр, алгоритми, спостережувальні дані, астрономія, астрофізика, CoLiTec*