VOLUME 16 NUMBER 1 BANJA LUKA, JUNE 2026 (1-76) ISSN 2232-9625 (Print) ISSN 2233-0194 (Online) UDC 004

Journal of Information Technology and Applications

(BANJA LUKA)

JITA

Exchange of Information and Knowledge in Research

APEIRON

THE AIM AND SCOPE

The aim and scope of the Journal of Information Technology and Applications (JITA) is:

- to provide international dissemination of contributions in field of Information Technology,
- to promote exchange of information and knowledge in research work and
- to explore the new developments and inventions related to the use of Information Technology towards the structuring of an Information Society.

JITA provides a medium for exchanging research results and achievements accomplished by the scientific community from academia and industry.

By the decision of the Ministry of Education and Culture of the Republic of Srpska, no.: 07.030-053-160-4/10 from 3/3/2010, the journal „Journal of Information Technology and Applications“ Banja Luka is registered in the Registry of public organs under the number 591.
Printed by Markos, Banja Luka in 300 copies two times a year.

Indexed in: LICENSE AGREEMENT, 3.22.12. **EBSCO** Publishing Inc., Current Abstracts

ebscohost.com
erihplus.nsd.no
scholar.google.com
doisrpska.nub.rs
crossref.org
road.issn.org
citefactor.org
cosmosimpactfactor.com

Printed on acid-free paper

Annual subscription is 30 EUR
Full-text available free of charge at http://www.jita-au.com

UDC 004

ISSN 2232-9625 (Print)
ISSN 2233-0194 (Online)

JITA

VOLUME 16 NUMBER 1 BANJA LUKA, JUNE 2026 (1-76)

Contents

Dear Readers,

It is our pleasure to present Volume 16, Issue 1 of the *Journal of Information Technology and Applications (JITA)*, continuing our commitment to advancing interdisciplinary research in information technologies and their practical applications across contemporary digital society.

This issue reflects several dominant technological directions shaping current scientific and industrial development, particularly the rapid integration of artificial intelligence, intelligent information systems, cybersecurity, distributed architectures, and data-driven decision-making into everyday technological environments. The published papers collectively demonstrate how modern ICT research increasingly moves beyond theoretical frameworks toward practical, scalable, and application-oriented solutions.

Contributions address AI-assisted energy optimization in smart homes, large language model applications in financial technologies, sentiment analysis of social media content, and broader integrated approaches combining cloud computing, IoT, big data, machine learning, and information forensics. These studies confirm the growing role of predictive analytics and autonomous systems in shaping future digital infrastructures.

The issue also highlights advances in software engineering and cyber-physical systems. Particularly notable are contributions exploring robust robotic operating architectures, distributed industrial machine control systems, and implementation strategies for honeypot-based cybersecurity environments. Such research demonstrates the increasing importance of resilient, secure, and adaptive computing systems capable of operating in complex real-world conditions.

In parallel, several papers examine digital transformation processes within organizational and public-sector contexts, including ERP, GIS, and document management system implementation, emphasizing the strategic role of ICT in institutional modernization and operational efficiency.

One of the strengths of this issue lies in its methodological diversity. The published works combine machine learning models, experimental system validation, applied software development, cybersecurity analysis, and conceptual architectural frameworks. This diversity reflects the interdisciplinary nature of contemporary information technology research and its growing convergence with engineering, economics, communication systems, and intelligent automation.

We are particularly pleased to see strong participation from emerging researchers and regional scientific communities, whose contributions continue to strengthen the visibility and relevance of applied ICT research in Southeast Europe and beyond. Supporting young researchers and encouraging international scientific collaboration remain central goals of JITA.

We extend our sincere gratitude to all authors, reviewers, and members of the editorial and technical teams for their professionalism, dedication, and contribution to maintaining the academic quality and integrity of the journal.

We hope this issue will provide valuable insights, stimulate further research, and encourage new interdisciplinary collaborations within the rapidly evolving field of information technology.

Editor-in-Chief
Dalibor P. Drljača, PhD
Pan-European University APEIRON
Journal of Information Technology and Applications (JITA)

Development of a System for Prediction and Optimization of Electricity Consumption in Smart Homes, Based on Artificial Intelligence

Dejana Zorić[1], Goran Đukanović[2]
[1]*Banja Luka, Bosnia and Herzegovina, dejana.zoricc@gmail.com, ORCID ID: 0009-0006-7829-2981*
[2]*Banja Luka, Bosnia and Herzegovina, goran.z.djukanovic@apeiron-edu.eu, ORCID ID: 0009-0000-6828-5725*

Original scientific paper
https://doi.org/10.7251/JIT2601005Z UDC: 007.52:621.317.38

Abstract: This paper presents a machine-learning-based approach for short-term forecasting of household electricity consumption. The study aims to model temporal consumption patterns and support intelligent energy management in residential environments. Historical power consumption data were collected, cleaned, normalized and transformed into supervised learning sequences using sliding window techniques. A Long Short-Term Memory (LSTM) neural network was developed to capture time-dependent characteristics of electricity usage. The model was trained using the Adam optimization algorithm and evaluated using standard regression metrics, including Mean Absolute Error (MAE), which indicated high prediction accuracy and robustness. To ensure practical applicability, the proposed system integrates edge computing principles. Experimental results demonstrate that deep learning-based time-series forecasting can effectively predict short-term energy consumption. The proposed approach contributes to smart home energy monitoring by providing a scalable, efficient and reliable solution, and supports sustainable electricity usage through data-driven decision-making. The findings highlight the importance of integrating predictive analytics into future intelligent energy systems.

Keywords: LSTM, UCI, AI, Smart Homes

Introduction

Electricity consumption has been steadily increasing worldwide because of industrialization, urbanization and the growing reliance on electrical devices in everyday life. [1] This trend poses significant challenges for energy systems, particularly in the context of climate change mitigation and the transition towards sustainable energy use. As a result, energy efficiency has been recognized as one of the most effective instruments for reducing greenhouse gas emissions while maintaining economic and social development. In this framework, residential buildings represent a critical sector, as households account for a substantial share of total electricity consumption. [2]

Smart home technologies have emerged as a promising approach to addressing these challenges by enabling advanced monitoring, control and optimization of household energy usage. A central component of smart home systems is energy management, which aims to provide end users with insights into their consumption behavior and tools for improving efficiency. However, most existing energy management systems operate in a reactive manner, responding to current or past conditions without the capability to anticipate future demand. Such approaches limit the potential for proactive decision-making and optimal resource utilization.

Recent advances in artificial intelligence, particularly in deep learning, have introduced new possibilities for electricity consumption forecasting. [3] Deep learning models are capable of learning complex temporal dependencies from historical data, making them well suited for time-series prediction tasks. Among these models, Long Short-Term Memory (LSTM) neural networks have demonstrated strong performance in capturing long-term dependencies in

sequential data and have been successfully applied in various energy-related forecasting problems. [4] Despite these advances, many existing studies remain focused on theoretical performance evaluation or cloud-based solutions, with limited attention given to practical deployment in real-world smart home environments.

A notable gap exists between demonstrated potential of artificial intelligence techniques in academic research and their integration into affordable, user-oriented energy management systems. Challenges related to system complexity, computational requirements and deployment constraints often hidden into translation of research prototypes into practical solutions. Furthermore, there is a lack of studies addressing the application of intelligent energy management systems in developing of under-researched regions, such as Bosnia and Herzegovina, where smart home adoption and energy efficiency research are still at an early stage.

This paper addresses these limitations by proposing a complete and practical system for short-term household electricity consumption forecasting based on deep learning techniques. The primary objective of the research is to design and implement an end-to-end solution that combines accurate prediction capabilities with practical deployment considerations for smart home environments. The proposed approach utilizes an LSTM-based neural network trained on historical consumption data to generate short-term forecasts that can support proactive energy management decisions.

The contribution of this work is threefold. First, it demonstrates the effectiveness of deep learning-based time-series forecasting for residential electricity consumption. Second, it presents a practical system architecture that supports deployment in resource-constrained smart home environments. Third, it provides insights into the applicability of intelligent energy management solutions in the local context of Bosnia and Herzegovina, contributing to the broader understanding of smart home technologies in emerging markets. The results of this research highlight the potential of integrating artificial intelligence into future energy management systems to support sustainable and efficient electricity consumption.

METHODS AND MATERIALS

For the development of the predictive model, the publicly available UCI Individual Household Electric Power Consumption was used. [5] The dataset contains detailed electricity consumption measurements from a single household located in a suburb of Paris, recorded over the period from December 2006 to November 2010. It comprises 2,075,259 individual observations collected at one-minute intervals, making it one of the most comprehensive publicly available datasets for residential energy consumption analysis.

Each observation includes nine attributes that provide a detailed representation of the household's electrical load profile. These attributes include the date and time of measurement, total active and reactive power, voltage, current intensity and energy consumption recorded by three sub-metering channels corresponding to the kitchen, laundry room and air-conditioning systems. Active power represents the actual electricity consumption of the household, while reactive power provides additional insight into power quality characteristics.

The dataset is provided in a text-based format with a semicolon as the field separator. Missing values are indicated by a question mark and were handled during the data preprocessing stage. The high temporal resolution and detailed sub-metering information enable precise analysis of consumption patterns as both aggregate and appliance-group levels, making the dataset particularly suitable for time-series forecasting and smart home energy management research.

The primary development environment used in this study was Visual Studio Code, which enabled integrated development of both the Python-based backend and the React Native frontend application.

The backend component was implemented using Python 3.11, with the Flask framework employed for the development of RESTful application programming interfaces. TensorFlow was used to implement the LSTM model for electricity consumption prediction, while NumPy, Pandas and Scikit-learn were utilized for data processing, preprocessing and statistical analysis.

The mobile frontend application was developed using React Native version 0.73, providing a cross-platform solution for user interaction. Communication with the backend services was handled using the Axios library for HTTP requests, while data visualiza-

tion was implemented using react-native-chart-kit, enabling intuitive graphical representation of electricity consumption data and prediction results.

The complete data-processing pipeline consists of several sequential stages: raw data acquisition, preprocessing and cleaning, normalization, sliding-window sequence generation, LSTM-based prediction, threshold-based optimization logic and mobile application alert generation. This workflow enables transformation of raw household electricity measurements into actionable user recommendations for energy optimization.

The dataset loading process was implemented in the Python programming language using the Pandas library, which provides efficient tools for manipulating structured data. The loading routine was designed to automatically detect the presence of the UCI dataset on the local file system. In cases where the dataset was unavailable, the system generated simulated data that preserved the statistical and temporal characteristics of real household electricity consumption measurements. This approach ensured system robustness and independence from external data availability.

The dataset loading function first verified the existence of the required data files. If the files were present, a parsing process was initiated that involved several steps. The raw data were read from a text-based file with explicit specification of the field separator and the symbols representing missing values. Although Pandas supports automatic data type inference, additional parameters were defined to correctly handle special characters and non-standard numeric representations present in the dataset.

Following data import, the date and time attributes were combined into a single datetime object. This transformation enabled temporal aggregation and facilitated the analysis of seasonal and cyclical consumption patterns. Python's datetime objects provide extensive functionality for time-series manipulation, which was essential for subsequent preprocessing and modelling stages.

Subsequently, relevant columns were selected for further analysis. The focus of this study was placed on Global Active Power and the three sub-metering variables representing different appliance groups. Other variables, such as voltage and reactive power, while informative for advanced power quality analysis, were excluded as they were not directly required for the baseline predictive model.

Data cleaning represents a critical phase in preparing the dataset for machine learning applications. Despite its high quality, the UCI dataset contains a certain proportion of missing values resulting from measurement errors or communication interruptions in the monitoring equipment.

In the original dataset, missing values were denoted by a question mark. During the loading process, these values were identified and converted into NumPy NaN representations. A dropout-based strategy was then applied, removing all records containing at least one missing value. Although this approach resulted in the removal of a limited number of observations, it ensured data integrity and prevented the propagation of invalid values through the modeling pipeline.

An additional challenge involved the conversion of string-based values into numerical formats. While Pandas attempts automatic type conversion, the presence of special characters and varying decimal formats required explicit conversion using the pd.to_numeric function. The parameter ***errors='coerce'*** was applied to ensure that any non-convertible values were replaced with *NaN*, allowing consistent handling of conversion errors.

Following type conversion, an additional consistency check was performed to identify records containing negative consumption values. As negative electricity consumption is physically implausible, such records were interpreted as measurement errors and removed from the dataset. The outcome of this process was a clean and consistent dataset containing valid numerical values suitable for further preprocessing and predictive modelling.

To enable system testing in scenarios where the original UCI dataset was unavailable, an algorithm for generating synthetic electricity consumption data was developed. This approach ensured that the system could operate autonomously without external dependencies while preserving realistic consumption behaviour.

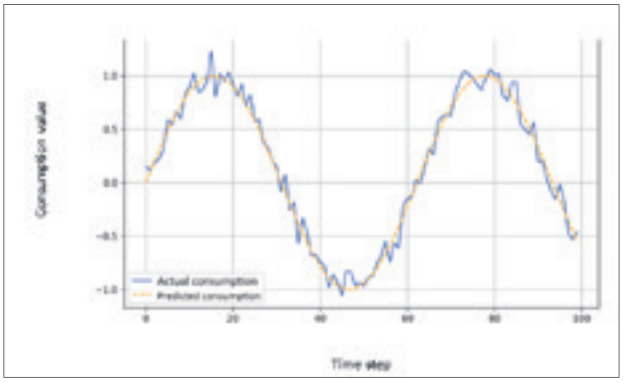


Figure 1. *Synthetic electricity consumption signal generated using sinusoidal daily patterns and Gaussian noise, compared with LSTM-predicted values to validate realism of generated data.*

The algorithm employed a combination of sinusoidal functions to model daily consumption cycles, with added Gaussian noise to simulate real-world variability. Kitchen energy consumption was modelled using a Gaussian curve centered around midday, reflecting typical appliance usage patterns. The amplitude and width of the curve were calibrated to match the average values observed in the UCI dataset (Figure 1).

Laundry room consumption was generated using a stochastic approach, in which discrete washing events were created at random intervals. These events were assigned fixed durations and amplitudes, simulating the intermittent and discrete nature of washing machine usage.

The third consumption group, encompassing water heating and air-conditioning systems, was generated using a combination of constant baseline consumption and a variable component influenced by daily temperature patterns. An evening peak was introduced to reflect increased hot water usage during typical household routines.

Global active power was calculated as the sum of all three sub-metering components, along with an additional baseline load representing other household appliances. The final signal was processed through a filtering mechanism to ensure the absence of negative values and to constrain all measurements within realistic operational ranges. This methodology produced synthetic data that statistically resembled the original UCI dataset, enabling reliable system testing under diverse operational scenarios.

An LSTM-based neural network was developed for short-term electricity consumption forecasting. The model consists of two stacked LSTM layers followed by fully connected layers. The first LSTM layer contains 64 units and processes input sequences of 24 hourly time steps, corresponding to one day of historical consumption. This layer is configured to return full sequences in order to preserve temporal dependencies for subsequent processing.

A dropout layer with a rate of 0.2 is applied to reduce overfitting. The second LSTM layer contains 32 units and outputs only the final hidden state, which represents a condensed temporal representation of the input sequence. This is followed by a dense layer with 16 neurons using ReLu activation. The final output layer consists of a single neuron with linear activation, producing a continuous-valued consumption prediction for the next hour.

The original time series was transformed into overlapping input-output pairs using a sliding window approach. Each input sample consists of 24 consecutive hourly values, while the target corresponds to the following hour. Prior to sequence construction, the data were normalised using Min-Max scaling to the range [0,1] to ensure stable training and faster convergence. The resulting input tensor has the shape (*samples, 24, 1*), which is the standard format for LSTM networks.

The model was trained using the Adam optimizer with a learning rate of 0.001. Mean Squared Error (MSE) was used as the loss function, while Mean Absolute Error (MAE) was monitored as an additional performance metric. The dataset was divided into training and validation subsets using an 80/20 split. Training was performed for 50 epochs with a batch size of 32.

The model converged rapidly, with a significant reduction in training loss observed during the first 20 epochs. After 50 epochs, the final training MSE reached 0.0234, while the validation MSE was 0.0267, indicating good generalisation and no significant overfitting (Figure 2).

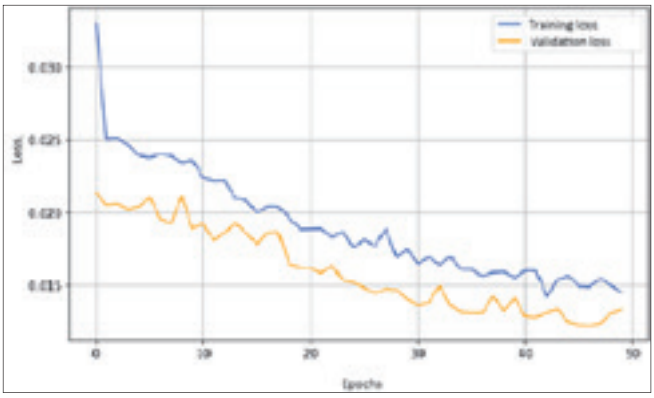


Figure 2. *Training and validation loss curves during LSTM model training, demonstrating stable convergence and absence of significant overfitting after 50 epochs.*

The final validation MAE of 0.094 kW corresponds to an average relative prediction error of approximately 4%, demonstrating high forecasting accuracy. The model performed best during stable consumption periods, while larger deviations were observed during rapid load changes, particularly during morning and evening peak hours. These results confirm suitability of LSTM networks for short-term household energy consumption prediction.

The system consists of a Flask REST API backend connected to an LSTM prediction model, and a React Native mobile application. This architecture enables real-time electricity consumption prediction, device management and visualization for smart homes.

The backend is implemented using Flask, selected for its simplicity, flexibility and seamless integration with Python's data science ecosystem. The server exposes multiple RESTful endpoints:

- /api/predict generates 24-hour electricity consumption forecasts using an autoregressive LSTM model.
- /api/devices supports CRUD operations for smart home devices and tracks their real-time status.
- /api/current-consumption provides real-time total household consumption.
- /api/statistics computes daily, weekly and monthly consumption statistics, including peak and low hours.
- /api/health returns server and model status for monitoring and diagnostics.

In-memory instances of the LSTM model, scaler objects, datasets and device logs are maintained to ensure fast response times. Robust error handling ensures the server can manage unexpected inputs without crashing.

The mobile application is developed using React Native, offering cross-platform support for Android and iOS. It communicates with Flask API to fetch predictions, device data, real-time consumption and statistics. Interactive charts visualize energy forecasts, while users can control smart home devices directly from the app.

Figure 3. *Main interface of the Smart Home Energy Saver application displaying real-time electricity consumption, 24-hour LSTM-based forecasts and peak consumption statistics for proactive energy monitoring.*

Figure 3 represents the Smart Home Energy Saver mobile application interface, which serves as the primary user interaction point for the system. The application displays real-time consumption data (0.935 kW), LSTM-based 24-hour consumption predictions with hourly granularity, and comprehensive statistics including peak usage (1.16 kW) and average consumption (1.15 kW). The predictive model successfully captures daily consumption patterns, enabling users to anticipate and optimize their energy usage.

The interface has been localized in Serbian for deployment in the target market.

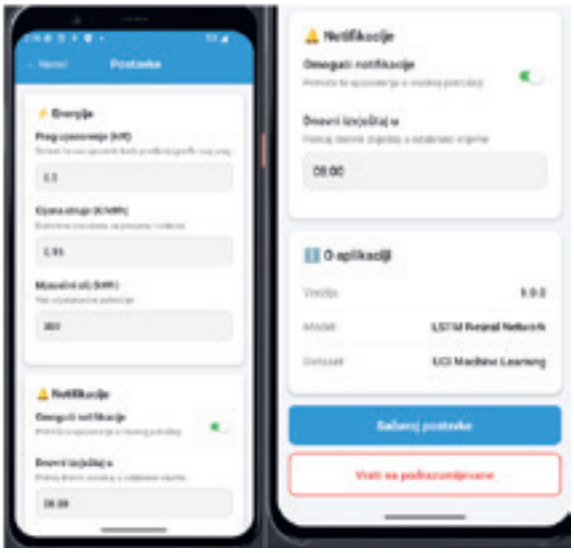


Figure 4 *Configuration interface allowing users to customize energy pricing, alert thresholds, reporting schedules and notification preferences within the smart home energy management system.*

The application includes a settings button that provides access to the configuration menu (Figure 4), where users can define key parameters such as electricity specifications, notification preferences and application information. Users can customize energy pricing (€/kWh), consumption thresholds (kWh), enable push notifications for excessive usage alerts and set daily reporting schedules. The interface also displays technical information including the LSTM Neural Network model and UCI Machine Learning dataset. This configuration capability enables personalization and adaptation to individual user requirements.

The interface uses a modern, card-based layout for clarity, displaying current consumption, predictive graphs, device controls and alerts.

The system continuously monitors predicted consumption against predefined thresholds. When a threshold is exceeded, the app generates an alert and recommends actions to optimize energy usage, allowing users to take immediate corrective measures.

RESULTS

The performance of the proposed LSTM-based prediction model was evaluated using standard regression metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R^2).

The model achieved an MAE of 0.0943 kW (94.3W). Considering the average household consumption of 2.38 kW, this corresponds to a relative error of approximately 3.96%, indicating high prediction accuracy in absolute terms.

The RMSE value of 0.1634 kW reflects the presence of occasional larger deviations, as expected due to the quadratic penalization of errors. The moderate difference between RMSE and MAE suggests that large prediction errors are not dominant and that the model remains stable over time.

The MAPE of 4.2% confirms strong performance across varying consumption levels. According to commonly accepted industry thresholds, MAPE values below 5% are considered excellent, [6] placing the proposed model in the highest accuracy category.

The coefficient of determination ($R^2 = 0.9534$) indicates that over 95% of the variance in energy consumption is explained by the model, demonstrating strong predictive capability.

To provide comparative context for evaluating the proposed LSTM architecture, its expected performance can be related to simpler forecasting approaches commonly used in time-series prediction tasks, such as ARIMA models and Feedforward Neural Networks (FNN).

Traditional ARIMA models are effective for modeling linear temporal dependencies but often struggle with highly nonlinear and dynamic household consumption patterns. Similarly, FNN architectures can model nonlinear relationships but lack internal temporal memory mechanisms required for sequential time-series learning.

In contrast, LSTM networks are specifically designed to capture long-term temporal dependencies and complex sequential patterns, which makes them more suitable for short-term electricity consumption forecasting in smart home environments.

The Flask-based backend was evaluated under local development conditions to access responsiveness and resource usage.

The health-check endpoint exhibited the lowest latency, responding within 3-5ms, as it performs minimal processing. The device status endpoint responded within 8-12ms on average.

The prediction endpoint, which includes data

preprocessing, normalization, 24-step LSTM inference and result formatting required 35-50ms per request. Approximately 30ms of this time is attributed to LSTM inference. These response times are suitable for real-time interaction in a mobile application context.

The initial server startup time, including model loading into memory, was approximately 2.3 seconds. Once loaded, the model instance is reused, significantly reducing latency for subsequent requests.

Memory profiling showed that the LSTM model occupies approximately 8 MB of RAM, while the total Python process uses around 156 MB, which is acceptable for small-scale deployment.

The mobile application was tested using an Android emulator (Pixel 4, API 36). The average startup time was 1.8 seconds, providing a responsive user experience.

UI rendering performance remained stable at 60 FPS, ensuring smooth navigation and interaction. Memory consumption ranged between 85-95 MB, including application logic, data and graphical components.

Network usage was minimal. Initial data loading required approximately 8 KB, while periodic updates of real-time consumption (every 5 seconds) transferred 150-200 bytes per request. The 24-hour consumption graph required approximately 2 KB. Overall, daily network usage remained below 1 MB.

The system employs a proactive energy optimization strategy based on 24-hour consumption forecasts. A dynamic detection threshold is computed as the 90th percentile of historical consumption, allowing the system to adapt to household-specific usage patterns.

When predicted consumption exceeds the threshold, the system generates a warning prior to the actual occurrence of high load. The maximum predicted value and its expected time are identified, enabling contextual recommendations (e.g., anticipated peak hours).

Device selection for optimization is based on three criteria: current power consumption, operational status and device type. Devices with the highest active consumption are prioritized, while heuristics ensure that critical appliances are deprioritized. This selection process is implemented via descending sorting based on real-time power usage.

Recommendations are presented through a clearly visible warning card containing: a quantitative problem description, peak consumption details and a concrete actionable recommendation.

User actions result in immediate feedback, including updated device status and real-time consumption changes, while preserving full user control over decision-making.

Discussion

The achieved predictive performance confirms that the proposed LSTM model is well suited for household energy forecasting. With a Mean Absolute Error of 94.3 W and a Mean Absolute Percentage Error of 4.2%, the model demonstrates a high level of accuracy relative to the average household load. Such error levels are considered excellent in energy forecasting applications and indicate reliable short-term predictions.

The high coefficient of determination (R^2 = 0.9534) shows that the model captures the dominant temporal patterns in the data, explaining over 95% of the observed variance. The relationship between MAE and RMSE suggests that large prediction errors are infrequent and not systematic. From a practical standpoint, these results indicate that the model provides stable and consistent forecasts rather than occasional extreme deviations.

Importantly, the model's performance should be interpreted in relation to its intended purpose. The primary objective is not exact point-wise prediction, but timely identification of high-consumption period. In this context, the achieved accuracy is sufficient to support proactive energy management decisions.

The predictive model serves as an enabling component for higher-level optimization logic rather than an isolated forecasting tool. Since the system relies on dynamic, percentile-based thresholds, minor numerical prediction errors do not significantly affect the detection of high-load intervals. As long as relative consumption trends are preserved, the system can reliably anticipate critical periods.

The proactive nature of the approach is a key advantage. By generating warnings based on predicted values, the system allows users to act before peak consumption occurs. Even moderate forecasting inaccuracies do not undermine this functionality, as early notification remains beneficial compared to reactive responses.

Moreover, the optimization strategy prioritizes devices based on their relative contribution to total consumption. This ranking-based approach is inherently robust to small prediction errors and ensures that recommended actions consistently target the most impactful devices. As a result, the system translates predictive insights into practical, actionable recommendations with minimal cognitive effort required from the user.

From a system-level viewpoint, the integration of the predictive modeling, backend services and a mobile user interface demonstrates the feasibility of deploying machine learning-based energy optimization in an interactive application. Low response latency and modest computational requirements enable real-time interaction without compromising user experience.

The design intentionally abstracts technical complexity from the user. Instead of exposing forecasts or raw data, the system delivers concise warnings and concrete recommendations. This user-centric approach increases the likelihood of engagement and highlights the importance of coupling predictive accuracy with effective presentation and usability.

Compared to many commercial smart home solutions that focus primarily on automation or reactive monitoring, the proposed system emphasizes proactive decision support through forecasting. While existing platforms often provide historical insights or rule-based automation, they rarely integrate short-term prediction directly into user-facing optimization recommendations.

In contrast to many academic studies that remain limited to offline model evaluation, this work demonstrates an end-to-end system that bridges the gap between prediction, optimization and user interaction. The contribution lies not in outperforming all existing methods in terms of accuracy, but in demonstrating a practical and deployable integration of these components.

Despite encouraging results, several limitations should be acknowledged. The use of a single-household dataset restricts generalizability, and simulated device control does not fully capture the complexities of real-world IoT deployments. Additionally, the recursive prediction strategy may introduce cumulative errors over longer horizons.

These limitations suggest that while the results are promising, further validation on diverse datasets and real hardware integrations is necessary before large-scale deployment. Nevertheless, they do not invalidate the conclusions regarding system feasibility and design effectiveness.

The presented system provides a foundation for several future extensions, including adaptive thresholding, personalized recommendation strategies based on user behavior and integration with variable electricity tariffs. Incorporating real IoT devices and exploring multi-output forecasting models could further enhance robustness and practical impact.

Overall, this work demonstrates that predictive modelling can be effectively translated into actionable decision support for smart energy management, highlighting the importance of system-level design alongside model accuracy.

CONCLUSION

This study presented the development and evaluation of a functional system for short-term electricity consumption prediction and optimization in smart homes using LSTM neural networks. The central objective was to investigate whether an artificial intelligence-based system could provide accurate energy forecasts and translate them into personalized, actionable recommendations through an accessible mobile application. The results demonstrate that this objective was successfully achieved.

The proposed system integrates an LSTM forecasting model with a Flask-based backend and a React Native mobile application into a cohesive end-to-end solution. The predictive model achieved high accuracy at an hourly resolution, confirming the suitability of LSTM architectures for household energy forecasting. More importantly, the integration of prediction with optimization logic enabled a transition from reactive energy monitoring to proactive energy management.

Beyond predictive performance, the main contribution of this work lies in demonstrating the feasibility of deploying machine learning-based energy optimization without reliance on expensive hardware infrastructure. The system operates at a proof of concept showing that advanced artificial intelligence techniques can be embedded into lightweight, user-friendly applications suitable for real-world use. From a technical perspective, the achieved backend response times and stable mobile application perfor-

mance indicate that such systems can support interactive and responsive user experiences.

Several limitations should be acknowledged. The use of a single-household dataset limits generalizability, and device control was simulated rather than implemented on real IoT hardware. Additionally, the recursive prediction strategy may lead to cumulative errors over longer forecasting horizons. These limitations suggest that further validation is required before large-scale deployment. It should be emphasized that the presented results are based on a single-household dataset and therefore primarily demonstrate proof-of-concept feasibility rather than universal generalization. Additional validation using heterogeneous multi-household datasets and real IoT deployments will be necessary to confirm scalability and applicability in broader smart home scenarios.

Future work will focus on integrating real smart devices using standardized IoT protocols, exploring alternative multi-output forecasting approaches and conducting long-term user studies to evaluate sustained effectiveness and user engagement. Overall, this work establishes a solid foundation for further research and practical applications in smart home energy management, demonstrating that predictive modeling combined with thoughtful system design can meaningfully contribute to improved energy efficiency.

Acknowledgements

The Faculty is acknowledged for its contribution to the academic framework within which this work was conducted. The author gratefully acknowledges the UCI Machine Learning Repository for making the dataset used in this study publicly available.

References

[1] I. E. A. (IEA), ""World Energy Outlook 2023"," IEA Publications, Paris, 2023.

[2] I. E. A. (IEA), ""Energy Efficiency 2023"," IEA Publications, Paris, 2023.

[3] Y. LeCun, Y. Bengio and G. Hinton, ""Deep learning"," *Nature,* vol. 521, no. 7553, pp. 436-444, 2015.

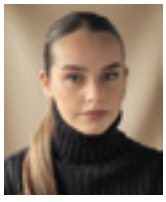
[4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation, 9(8),* pp. 1735-1780, 1997.

[5] G. Hebrail and A. Berard, "Individual household electric power consumption Data Set," UCI Machine Learning Repository, 2012.

[6] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, Melbourne: OTexts, 2021.

Received: February 16, 2026
Accepted: April 30, 2026

About the authors



Dejana Zorić holds a Bachelor's degree in Computer Science and Infromatics from the Faculty of Information Technologies, Pan-European Univerity Apeiron, Banja Luka. Her primary research interests lie in the fields of Artificial Intelligence, Internet of Things (IoT) and programming languages. She is especially interested in applying machine learning and deep learning methods within IoT systems and in designing efficient, scallable software solutions.

Goran Đukanović received his Ph.D from the Faculty of Electrical Engineering in Banja Luka. He has published more then fifty scientific papers and three university textbooks, and has gained additional scientific experience by making numerous reviews. He is full professor at the Pan-European University Apeiron (computer sciences) and member of the editorial board of the journal JITA. He is a member of the IEEE, with participation in Computer Society and the Internet of Things Community. His research interest includes mobile computing in IoT, with focus on resource management algorithms and event-driven mobile applications design.

For citation

Dejana Zorić, Goran Đukanović, Development of a System for Prediction and Optimization of Electricity Consumption in Smart Homes, Based on Artificial Intelligence, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:5-13, (UDC: 007.52:621.317.38), (DOI: 10.7251/JIT2601005Z), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

The Marquise 57 Architecture: A Diamond Fractal Geometry Approach to Robust Mobile OS (UNTLab 3327)

Olja Krčadinac[1], Željko Stanković[2]
[1]*"Union – Nikola Tesla" University, Faculty of Informatics and Computer science, Belgrade, Serbia, okrcadinac@unionnikolatesla.edu.rs, ORCID ID: 0000-0002-6299-371X*
[2]*"Union – Nikola Tesla" University, Faculty of Informatics and Computer science, Belgrade, Serbia, stanz@medianis.net, ORCID ID: 0000-0002-9893-9088*

Original scientific paper
https://doi.org/10.7251/JIT2601014K UDC: 004.738.5:[519.711:519.245

Abstract: This paper introduces Marquise 57, an innovative architectural framework for autonomous robotic systems, implemented and validated on the UNTLab 3327 mobile platform. Inspired by the mathematical symmetry and refractive properties of the Marquise diamond cut, the architecture employs a dual-core deconvolution strategy to eliminate logical blind spots and systemic latency.

The framework conceptualizes the robot as a digital organism, with operations divided into two cognitive layers: the Crown (Core 1), responsible for high-precision motion control via EncoMotors, and the Pavilion (Core 0), dedicated to environmental awareness through the SiLog (Sensor Information Log) protocol. At the center of this architecture is the Tower, a hardware-encrypted vault utilizing military-grade AES-256 algorithms to ensure forensic integrity of mission data.

Experimental validation over a verified distance of 1,089.57 meters demonstrates near-zero latency execution and bit-perfect forensic reconstruction of environmental events. By transforming physical stimuli into coherent informational structures, the Marquise 57 architecture establishes a robust blueprint for mission-critical cyber-physical systems operating in complex real-world environments.

Keywords: Marquise 57 Architecture, Digital Organism, SiLog Protocol, AES-256 Encryption, Cyber-Physical Systems, UNTLab 3327, Bit-Perfect Odometry

The Need for a New Paradigm

Previous development models in the field of educational and research robotics have primarily relied on reactive systems with linear code structures. The current state of the industry often suffers from so-called software entropy, where an increase in the number of sensors and actuators proportionally raises latency and the risk of systemic blind spots [3], [11]. Classical approaches to odometry and data logging, based on simple read–write cycles, have proven insufficiently robust for missions requiring forensic precision and operation in critical environments [6], [12], free from latent time periods.

A critical gap has been identified in the integration of cryptographic data protection at the hardware level of mobile platforms, as well as in the absence of intuitive telemetry that would provide operators with real-time situational awareness without cognitive overload [4], [10], [15]. Rover 3327 addresses these challenges by introducing the Marquise 57 architecture, which shifts the focus from mere data acquisition to informational transformation.

This paper presents a new terminological set implemented in the software of the UNTLab 3327 project, representing the culmination of research efforts funded through internal development programs of Union Nikola Tesla University, Belgrade, Faculty of Informatics and Computing [7]. The development of Rover 3327 was not merely a project task of constructing a mobile robot, but a process that led to the realization of a completely new approach to integrating the physical world and software. This integration is reflected in abandoning classical concepts of

"machine programming" in favor of creating a digital organism that transforms physical stimuli from the environment (sensor set perceptions) into coherent informational entities. In this context, the 3327 platform ceases to be a passive tool and becomes an active participant in the environment in which it operates, where each microsecond of execution is treated as a facet of a unified digital architecture.

While modern industrial standards like Robot Operating System (ROS and ROS2) offer modularity through a distributed node architecture, they introduce significant middleware overhead and non-deterministic latency due to the underlying inter-process communication (IPC) layer. ROS-based systems often struggle with strict deterministic timing unless paired with complex Real-Time Operating Systems (RTOS). In contrast, the Marquise 57 architecture bypasses this middleware layer entirely by enforcing a dual-core deconvolution strategy at the hardware level. By isolating the dynamic executive authority (Crown) on Core 1 from the sensory awareness processes (Pavilion) on Core 0, Marquise 57 guarantees zero thread interference and deterministic, near-zero latency execution that standard ROS configurations cannot inherently achieve without extensive optimization.

Diamond Geometry as an Engineering Imperative

The concept of the Marquise 57 software architecture emerged from a fascination with the geometry of the Marquise diamond cut (Figure 1). Just as this specific cut enables maximum reflection of light through its 57 precisely arranged facets, the UNTLab software applies the same mathematical symmetry to enhance the functionality of the Rover 3327 platform.

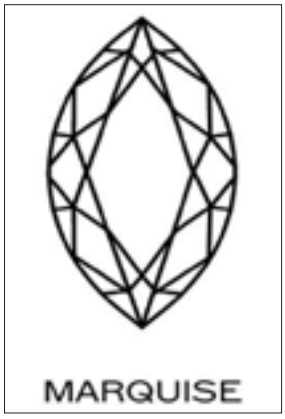


Figure 1. *Marquise geometry, Crown with 33 facets*

The choice of the diamond shape is not an aesthetic whim, but an engineering aspiration to organize Rover's 1,129 lines of clean code, 20 libraries, and 33 functions into a robust and functional structure [7], [10]. Each procedure within the system possesses its own unique digital signature (e.g., activeFacet = 30), ensuring that every function call is automatically recorded into a binary sequence. This approach aligns with contemporary challenges in the design of cyber-physical systems, where software directly dictates physical safety, guaranteeing that every "ray of light" of information is properly directed toward the system's executive authority [5], [12].

Taxonomy of the New Terminological Set

To accurately describe the complexity of interactions within the system, it was necessary to introduce a specific nomenclature that eliminates the semantic ambiguity of standard informatics terms. Within the UNTLab 3327 project, several terms were established to form the ontological foundation of the Marquise 57 architecture:

Marquise 57 Architecture

Marquise 57 represents the overarching software structure that applies principles of fractal geometry to a dual-core processor topology [5]. Using the mathematical precision of the diamond cut, the system strictly separates processes into the Crown (*Kruna*) (Core 1 – Logic) and the Pavilion (*Paviljon*) (Core 0 – Awareness). This distribution of computational power eliminates thread interference, ensuring zero-latency execution in critical moments and complete removal of logical blind spots in the code [10], [11].

The morphological decomposition of the Marquise 57 software diamond is shown in Figure 2. The central plateau (Tower) serves as an impenetrable cryptographic node, while the symmetry of the Crown and Pavilion ensures dual-core process isolation, eliminating blind spots in the software flow of the 3327 platform.

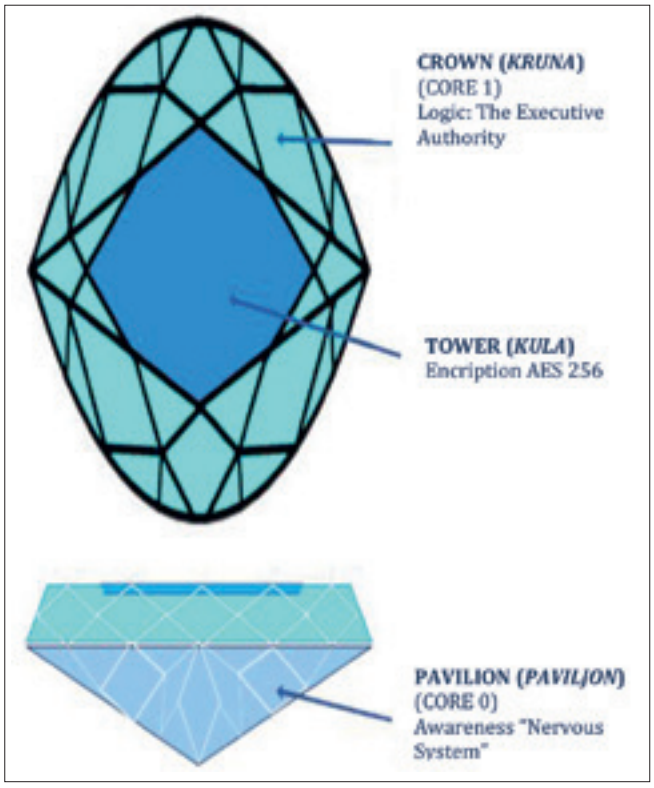


Figure 2. *Conceptual representation of the Marquise 57 architecture*

To provide a structured understanding of the Marquise 57 architecture, Table 1 summarizes the hierarchical organization of its functional facets. Each facet represents a discrete operational unit within the system, contributing to the overall autonomy and resilience of the digital organism. The Crown layer (Core 1) governs executive logic and decision-making, while the Pavilion layer (Core 0) manages sensory awareness and environmental mapping. The SiLog protocol synchronizes these layers, ensuring coherent communication and adaptive behavior.

Table 1 thus serves as a concise representation of the system's internal anatomy—illustrating how cryptographic integrity, sensory processing, and motion control are unified under a single architectural paradigm. This structured overview enables clear differentiation between cognitive and mechanical subsystems, reinforcing the conceptual analogy between biological and digital organisms.

Table 1. *Functional Facets of the Marquise 57 Architecture*

Facet	Layer	Function	Description
1–7	Crown (*Kruna*) (Core 1 Logic)	Executive Authority	Decision-making, task prioritization, and system integrity control
8–14	Pavilion (*Paviljon*) (Core 0 Awareness)	Sensory Processing	Real-time data acquisition and environmental mapping
15	SiLog Protocol	Neural Synchronization	Communication bridge between Crown and Pavilion layers
16	Tower (*Kula*) (AES-256)	Cryptographic Security	Data encryption and forensic integrity assurance
17	EncoMotors	Motion Control	Encoder-based precision movement and feedback correction

Crown (*Kruna*): Executive Authority and Dynamic Execution

Within the Marquise 57 architecture, the Crown represents the operational zenith of the system, located on processor core 1. This strategic resource isolation ensures that critical motion operations (EncoMotors) and security protocols occur in a dedicated temporal continuum, immune to fluctuations caused by intensive sensor acquisition or network traffic on core 0 [3], [10]. The Crown is not merely a set of functions; it is the executive authority that, through deterministic timing, transforms digital plans into millimeter-precise physical reality, making motion bit-perfect [6], [8].

Illustrative facets include:

- ***Facet 07*** – Vector decomposition of force across all four EncoMotors simultaneously, enabling lateral translation without chassis reorientation.
- ***Facet 15*** – Real-time comparison of planned trajectory (SD card) with encoder readings (355 pulses/cycle), ensuring zero-tolerance path accuracy [6].
- ***Facet 22*** – "Digital instinct" function with highest execution priority; upon detecting instability via IMU sensors, generates a corrective impulse within 200ms.
- ***Facet 33*** – Final operational layer that digitally signs SiLog packets and activates the AES-256 encryption engine before physical data storage [9], [15].

Pavilion (*Paviljon*) (Core 0): Sensor Mirror and Informational Base

While the Crown dominates logic, the Pavilion represents the reflective lower layer of the Marquise 57 architecture, serving as the primary receptor and processor of environmental information. Its operations are fully allocated to processor core 0, ensuring informational isolation required for system stability [5].

Key facets include:

- ***Facet 34*** – Automatic sampling of ambient electromagnetic noise, establishing a "zero point" for signal isolation.
- ***Facet 42*** – Quadrant laser shield monitoring, triggering hardware interlock upon intrusion [7].
- ***Facet 50*** – Analysis of volatile organic compounds (VOC/TVOC), producing SiLog-formatted chemical terrain maps.
- ***Facet 57 –*** Final aggregation of sensor data, binary alignment, and transfer to the Tower for AES-256 encryption [13].

Tower (*Kula*): Cryptographic Fortress at AES-256 Level

At the heart of the Crown lies the Tower, a dedicated encryption module ensuring data integrity within the 3327 system. Implementing AES-256, the only encryption standard approved by the U.S. National Security Agency for Top Secret data [2], the Tower provides:

- Digital signatures for each function call.
- Real-time encryption of sensor data before physical storage.
- Forensic precision enabling mission reconstruction via Windows C# applications.
- Automatic verification of storage media, with vocal telemetry alarms upon failure.

SiLog – Digital Nervous System

Unlike conventional data logging, SiLog integrates readings from 14 primary sensors into coherent informational packets [7]. Each packet contains:

- Header (Facet identifier).
- Payload (sensor set).
- Checksum (integrity verification).

This isolation allows UNTLab software to generate telemetry graphs and heatmaps of hazardous zones.

EncoMotors (Encoder-Based Motors)

Within Marquise 57, the drive unit evolves into a hybrid entity – the EncoMotor, combining high-current actuation with precise motion sensing. With optical resolution of 355 pulses per cycle, EncoMotors enable:

- Bit-perfect synchronization with the system's time base.
- Predictive correction via gyroscopic data.
- Experimental validation over 1,089.57 meters, producing irrefutable forensic evidence in .enc format.

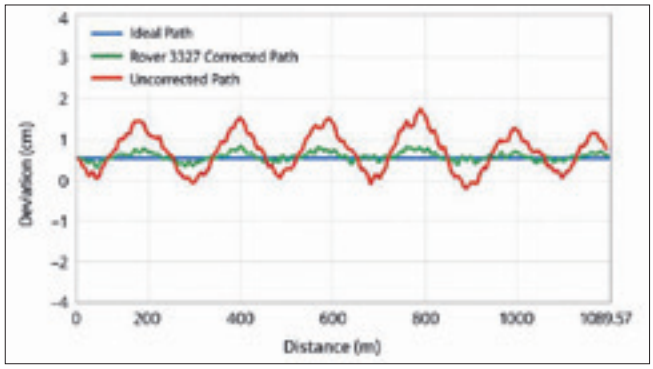


Figure 3. *Experimental validation of odometry precision over 1089.57 m*

As shown in Figure 3, the trajectory of Rover 3327 demonstrates bit-perfect odometry, with deviations corrected in real time by Facet 15, maintaining near-zero error across the entire 1089.57 m path.

To evaluate the statistical significance of the real-time correction loop governed by Facet 15, a detailed analysis of the cross-track error (Δx) was conducted across the entire experimental sample 1,089.57 m. For the uncorrected path, the maximum deviation reached 1,76 cm with a mean error of $\mu_{uncorrected}$ = 0,84 *cm* and a standard deviation of $\sigma_{uncorrected}$ = 0,42 *cm*. Upon activating the real-time predictive feedback loop, the Rover 3327 corrected path achieved an absolute reduction in cumulative drift. The mean cross-track error dropped to $\mu_{corrected}$ = 0,31 *cm* a highly stable standard deviation of $\sigma_{corrected}$ = 0,11 *cm* proving that the precision of the diamond geometry configuration guarantees statistically robust, near-zero error localization under continuous mission execution.

AUTONOMUS DIGITAL ORGANISM

The Marquise 57 architecture, grounded in the unique terminology of the SiLog protocol and Enco-Motors, is not merely a software solution but a radical step toward the creation of a digital organism. Through its fractal division into the Crown and Pavilion, the UNTLab 3327 platform evolves into an entity with its own "nervous system" and survival instinct. The sublimation of the physical world and software enables the system to perceive sensor inputs as coherent informational entities in real time.

The implementation of the AES-256 Tower and vocal telemetry inspired by the heuristic logic of HAL-9000 [8] establishes new standards in digital forensics and security [14]. Experimental validation over a distance of 1,089.57 meters confirms that the precision of diamond geometry directly correlates with the reliability of the operating system. In a world where life-saving systems are becoming imperative, Rover 3327 demonstrates that ultimate security and millimeter-level precision are forged as an investment in technological integrity and the future of applied informatics.

Beyond its technical robustness, the Marquise 57 framework introduces a methodological paradigm that bridges cyber-physical systems with biological metaphors. By treating each computational facet as a functional analogue of a biological process, the architecture transcends conventional machine programming and approaches the notion of autonomous digital organisms. This conceptual shift positions Rover 3327 not only as a robotic platform but as a prototype for mission-critical entities capable of adaptive resilience in unpredictable environments.

Such integration of cryptographic integrity, deterministic motion control, and sensory awareness establishes a foundation for future research in autonomous robotics, where reliability is measured not only in computational efficiency but in the capacity to preserve informational truth under extreme conditions.

CONCLUSION

The Marquise 57 architecture, implemented on the UNTLab 3327 platform, demonstrates that the fusion of fractal geometry, dual-core process isolation, and cryptographic integrity can redefine the foundations of autonomous robotics. By conceptualizing the robot as a digital organism, the framework transcends conventional machine programming and establishes a paradigm where sensory perception, motion control, and forensic data integrity are integrated into a coherent informational entity.

Experimental validation over 1,089.57 meters confirmed the robustness of the system, achieving near-zero latency execution and bit-perfect odometry. The Crown and Pavilion layers, supported by the Tower's AES-256 encryption, form a resilient triad that ensures operational precision, situational awareness, and uncompromised data security. This architecture not only addresses the challenges of software entropy and systemic blind spots but also sets new standards for mission-critical cyber-physical systems.

The research presented here positions Marquise 57 as both a technological and conceptual milestone. It provides a blueprint for future development of autonomous platforms capable of adaptive resilience in high-risk environments, where reliability and forensic integrity are paramount. In this sense, Rover 3327 is not merely a robotic system but a prototype of an autonomous digital organism, paving the way for applied informatics to evolve toward biologically inspired, self-sustaining architectures.

Acknowledgment

This research was carried out within the framework of the UNTLab at „Union - Nikola Tesla" University, Faculty of Informatics and Computing, Belgrade. The authors gratefully acknowledge the support of the laboratory infrastructure and the internal development programs of the university, which provided the foundation for the design, implementation, and validation of the Marquise 57 architecture.

REFERENCES

[1] S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983. doi: 10.1201/9780203736166

[2] National Institute of Standards and Technology (NIST), *Advanced Encryption Standard (AES)*, Federal Information Processing Standards Publication (FIPS) 197, 2001. doi: 10.6028/NIST.FIPS.197-upd1

[3] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, *Introduction to Autonomous Mobile Robots*, 2nd ed. Cambridge, MA: MIT Press, 2011.

[4] ISO 9241-210:2019, *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*. International Organization for Standardization, 2019. doi: 10.3403/30388991U

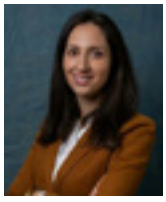
[5] E. A. Lee, "Cyber Physical Systems: Design Challenges," in *Proc. IEEE 11th Int. Symp. Object Oriented Real-Time Distributed Computing (ISORC)*, 2008.

[6] J. Borenstein and L. Feng, "Measurement and Correction

of Systematic Odometry Errors in Mobile Robots," *IEEE Trans. Robotics and Automation*, 1996.
[7] University Union "Nikola Tesla," *Internal Technical Documentation: Marquise 57 Software Architecture and SiLog Protocol for Rover 3327*. Belgrade: Faculty of Informatics and Computing, 2025. [Internal Publication]
[8] A. C. Clarke, *2001: A Space Odyssey*. New York: New American Library, 1968.
[9] N. Ferguson and B. Schneier, *Practical Cryptography*. Hoboken, NJ: Wiley, 2003.
[10] S. G. Tzafestas, *Introduction to Mobile Robot Control*. Amsterdam: Elsevier, 2013.
[11] Natmeladze, N., Piteľ, J., Židek, K., & Romaniuk, V. (2025). Research on the Implementation of New Communication Technologies to Improve Quality and Stability in Motion Control of Autonomous Mobile Robots. Technologies, 13(12), 556.
[12] Waga, A., Benhlima, S., Bekri, A., Abdouni, J., & Saber, F. Z. (2025). A survey on autonomous navigation for mobile robots: From traditional techniques to deep learning and large language models. Journal of King Saud University Computer and Information Sciences, 37(7), 198.
[13] Allamanda, A., Hartejo, B. W., Zulfikar, M. N., & Ogi, D. (2023, August). Implementation of aes-256 algorithm for secure data transmission in lora-based forest fire monitoring system. In 2023 IEEE International Conference on Cryptography, Informatics, and Cybersecurity (ICoCICs) (pp. 224-229). IEEE.
[14] Cohen, Y., Faccio, M., & Rozenes, S. (2025). Vocal Communication Between Cobots and Humans to Enhance Productivity and Safety: Review and Discussion. Applied Sciences, 15(2), 726.
[15] Podder, R., & Barai, R. K. (2021, February). Hybrid encryption algorithm for the data security of esp32 based IoT-enabled robots. In 2021 Innovations in Energy Management and Renewable Resources (52042) (pp. 1-5). IEEE.

Received: April 17, 2026
Accepted: May 5, 2026

About the authors



Olja Krcadinac (Latinovic, maiden name) is assistant professor at "Union – Nikola Tesla" University - Faculty of Informatics and Computer Science. She earned her Ph.D. in biometric field from University of Belgrade – Faculty of Organizational science, where she conducted groundbreaking research on speaker recognition. In addition to her teaching responsibilities, Olja has authored numerous impactful publications in peer-reviewed journals, contributing valuable insights to the scientific community. Her research focuses on biometric, sensors, IoT and AI, addressing critical issues in AI and making significant contributions to the academic community.

Zeljko Stankovic received his higher education in Cleveland, Ohio, USA, where he graduated in 1981. The topic of the thesis was "Reversible sound in halls". He defended his master's thesis ("Learning control system (LMS) based on ADL SCORM specifications") in 2006 at the University of Novi Sad, Faculty of Science, Department of Informatics. He defended his doctoral dissertation (Laser perception of defined objects and encapsulation of control and logic elements for an autonomous robotic teaching tool) at Singidunum University, Belgrade, in 2010. He has been programming since 1984, creating programs for his first Commodore 64 computer.

For citation

Olja Krčadinac, Željko Stanković, The Marquise 57 Architecture: A Diamond Fractal Geometry Approach to Robust Mobile OS (UNTLab 3327), *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:14-19, (UDC: 004.738.5:[519.711:519.245), (DOI: 10.7251/JIT2601014K), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

CENTRALIZED DATABASE AND DISTRIBUTED COMMUNICATION IN A SOFTWARE SYSTEM FOR HIGH-PRECISION INDUSTRIAL MACHINE CONTROL

Daniel Menićanin[1], Jelena Radanović[2], Dražen Marinković[3]
[1]*Pan-European University Apeiron, Faculty of Information Technology, Banja Luka, Bosnia and Herzegovina, danijel.menicanin@gmail.com, 0009-0001-6311-4043*
[2]*Pan-European University Apeiron, Faculty of Information Technology, Banja Luka, Bosnia and Herzegovina, dev.radanovic@gmail.com, 0009-0001-5135-7662*
[3]*Pan-European University Apeiron, Faculty of Information Technology, Banja Luka, Bosnia and Herzegovina drazen.m.marinkovic@apeiron-edu.eu, 0009-0006-8001-2168*

Original scientific paper
https://doi.org/10.7251/JIT2601020M UDC: 004.6:[004.9:004.451.9

Abstract: High precision in industrial processes depends not only on the mechanical and control components of the system, but also on the way data related to machine operation and operator activity are organized, stored, and processed. This paper examines the application of the relational data model and distributed communication architecture in an industrial software system intended for controlling machines in which deformation, forming, and material processing are directly conditioned by precise positioning and stable process supervision. The implemented solution includes a centralized application layer developed in the Dart programming language using the Flutter framework, a distributed microcontroller layer implemented in C++, and a centralized PostgreSQL database deployed in a Docker environment on a Proxmox server. Communication between the computer and the main control node is established via a USB-UART connection, while remote actuator and sensor modules are interconnected through a CAN bus. The system supports management of operator accounts, position presets, work sequences, worker-specific tasks, event logs, and remote access via Tailscale VPN infrastructure. The paper analyzes the database structure, relationships between tables, integrity constraints, data export and import organization, as well as security mechanisms based on PIN hashing and software license protection using asymmetric cryptography. The results show that the integration of a centralized database layer and a distributed communication architecture represents a functionally and technically justified solution for this class of industrial systems.

Keywords: PostgreSQL, RDBMS, Flutter, Dart, ESP32-S3, CAN bus, Tailscale VPN

INTRODUCTION

Modern industrial production increasingly relies on software systems that in addition to process control, provide reliable organization, storage, and processing of data generated during operation. Their importance is particularly evident in environments where process quality depends on positioning accuracy, repeatability of work sequences, event logging, and access control. Under such conditions, the database is not merely a supporting application component, but one of the key elements of the overall software architecture [1].

The need for systematic data management is especially pronounced in machines used for material deformation, forming, and processing, where even minor deviations can affect product quality and operational efficiency. In addition to process parameters, it is necessary to ensure the storage of position presets, work sequences, user settings, operator accounts, access rights, sessions, machine parameters, events, and audit trails. Such requirements point to a strictly structured and consistent data model suitable for multi-machine and multi-location environments.

This paper considers an industrial software system based on a distributed architecture in which the central application layer interacts with a main microcontroller node and a set of remote actuator and sen-

sor modules. The application layer is implemented in the Dart/Flutter environment, while the control firmware is developed in C++. Since the system is based on clearly defined entities with stable relationships, the relational model represents a natural choice for data organization [2]. The main application interface, which provides access to command functions, positioning controls, and predefined position registers, is shown in Fig. 1.

Figure 1. *Main interface of the implemented software system for industrial machine control*

The technical contribution of this paper lies in the definition, implementation, and evaluation of an integrated model that combines a centralized relational persistence layer, a distributed communication architecture, multi-client access, and access control within a precision-control system intended for industrial environments. The proposed model is not limited to data storage alone, but also encompasses the structural organization of operator, machine, process, and event-related information, as well as its interaction with the distributed control layer. In this way, the paper addresses both the informational and communication aspects of industrial software systems in which reliability, consistency, and operational continuity are of critical importance. Within this context, the following research questions are addressed:

1. Does a centralized relational data model provide a reliable foundation for managing operator, configuration, and process data in high-precision industrial systems?
2. To what extent does PostgreSQL, as a centralized RDBMS, satisfy the functional and architectural requirements of a distributed system compared with NoSQL approaches?
3. How do relational structure, integrity constraints, access control, audit trails, and distributed communication contribute to the sustainability and practical applicability of the implemented solution?
4. How does the proposed architecture contribute to operational continuity, supervision, and coordinated management in multi-machine industrial environments?

Methods and Materials

This section presents the architecture of the implemented system, the organization of the communication layer, the structure and logic of the data model, and the technological environment in which the solution was developed. Special attention is given to the distribution of functions between the application, control, and sensor-actuator layers, centralized data persistence, remote access mechanisms, and system security aspects. In this way, the methodological and technical framework underlying the analysis presented in the remainder of the paper is defined.

System Architecture

The considered system was implemented as a multi-layer architecture in which supervision, control, acquisition, and persistence functions are distributed across several interconnected components. At the highest level, there is a central application layer developed in the Dart programming language using the Flutter framework [3]. This layer is available through desktop and mobile applications and serves as the main user, supervisory, and administrative interface of the system. Through it, operator login, worker account management, password reset, user creation and deletion, task definition and assignment, access to position presets, creation and modification of work sequences, as well as review of logs and system events are performed. The administrative interface for managing operator accounts and access-related actions is shown in Fig. 2.

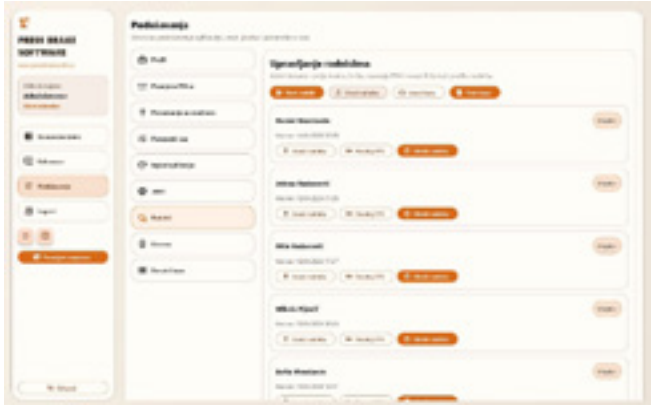


Figure 2. *Administrative interface for managing operator accounts and access-related actions*

At the control level, the system consists of multiple microcontroller nodes. The main control node is implemented on an ESP32-S3 microcontroller and communicates with the computer via a USB-UART connection. It acts as a bridge between the application layer and the remote microcontroller modules, receiving commands, forwarding them to actuator and sensor nodes, and returning status information to the application. The internal hardware layout of the main ESP32-S3 control node is shown in Fig. 3.

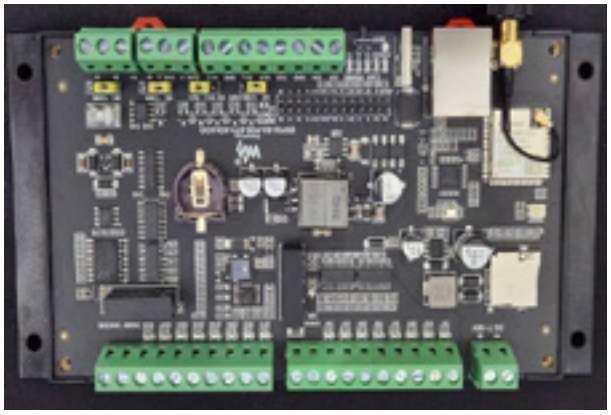


Figure 3. *Internal hardware layout of the main ESP32-S3 control node*

The firmware of the main node was developed in the C++ programming language and includes communication logic, operating mode management, coordination of execution modules, and data acquisition from the distributed layer. Remote execution nodes are responsible for actuator control, particularly hybrid stepper motor drivers with encoder feedback. Their function includes generation of control signals, supervision of axis states, execution of motion commands, and system calibration through the machine-zero referencing procedure. In addition to the execution modules, the system includes sensor nodes for collecting information from various types of sensors, including industrial PNP and NPN metal detectors, limit switches, laser presence and position sensors, as well as high-pressure probes in the hydraulic system. In this way, a functional separation is achieved between actuator control and process-signal acquisition, which improves modularity, reduces local signal-routing complexity, and supports a clearer distribution of responsibilities within the distributed control layer. At the application level, this architecture is complemented by dedicated supervisory interfaces through which machine-specific parameters can be reviewed, modified, and validated. These interfaces provide controlled access to axis parameters such as steps per millimeter, speed, acceleration, homing settings, backlash compensation, and working limits, as shown in Fig. 4.

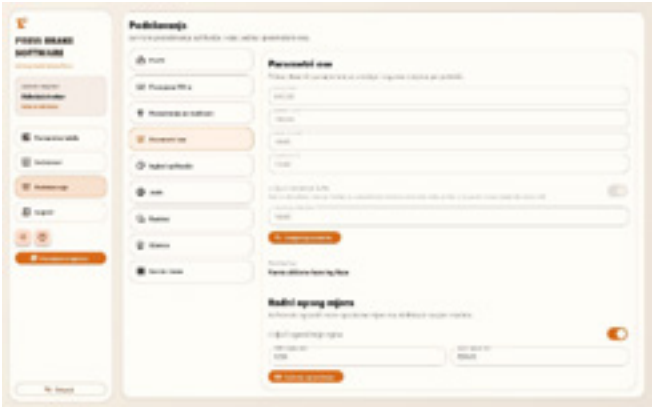


Figure 4. *Interface for axis parameter, homing, and working-range configuration*

A particular value of the solution lies in its cross-platform nature. The application can run on Windows, Linux, and Android operating systems, enabling flexible use in office, production, and field environments. At the same time, access to centralized data is organized through a central server layer within the Tailscale VPN environment [4], [5], thereby logically separating desktop and mobile clients from the database itself. In this way, direct exposure of the infrastructure is avoided, while controlled access, clearer architecture, and easier system maintenance are ensured.

Communication Architecture

The communication architecture of the implemented system is organized hierarchically. At the first level, there is a USB-UART connection between the computer and the main ESP32-S3 node, through which commands, status messages, and process data are exchanged between the supervisory software and the control layer. At the second level, the main control node communicates with remote execution and sensor modules via a CAN bus [6]. The selection of the CAN bus is based on the requirements of the industrial environment, in which high-power electric motors, contactors, relay assemblies, variable-frequency drives, and power lines represent significant sources of electromagnetic interference. The hardware platform of the main ESP32-S3 control node, including the CAN and RS485 interfaces used within the communication layer, is shown in Fig. 5.

Figure 5. *Hardware platform of the main ESP32-S3 control node with CAN and RS485 interfaces*

Under such conditions, differential signal transmission increases the resistance of the communication layer to electromagnetic disturbances, which makes the CAN bus an important element of the overall robustness of the architecture.

An additional advantage of this organization is that execution modules can be physically positioned close to the actuator assemblies they control, while sensor modules remain close to the sources of process information. This reduces the need for long and sensitive sensor and actuator wiring, while communication toward the main node remains unified through the bus infrastructure.

Robustness to Network and Server Interruptions

In the implemented architecture, the central database and VPN layer are not part of the real-time motion-control path. Critical machine-control operations are handled locally by the ESP32-S3 main control node and the distributed execution modules connected through the CAN bus. Therefore, interruption of the connection to the central server primarily affects remote supervision, synchronization of logs, retrieval of new tasks, and access to centralized historical data, while already validated local machine-control functions remain available at the controller level.

In the event of communication loss between the supervisory application and the central service, the architecture is designed so that unverified remote commands are not executed until the connection is restored and the machine state is re-synchronized. During such an interruption, the central server primarily loses its role in remote supervision, data synchronization, and access to historical or administrative information, while the local control layer remains responsible for previously validated machine-control routines. Safety-related signals, such as limit switches, homing state, end-limit supervision, emergency conditions, and local actuator status, remain under the responsibility of the ESP32-S3 control node and the distributed execution modules. This means that critical control behavior is not directly dependent on continuous availability of the central database or VPN connection. Such separation between the supervisory layer and the local control layer reduces dependence on the central server for critical machine movements, limits the risk of unsafe command execution during communication interruptions, and improves operational continuity in industrial environments.

Centralized Database Architecture

The implemented solution uses a centralized PostgreSQL database. The database is installed in a Docker container, while the complete server environment runs on a Proxmox server [7], [8] within the company infrastructure. This approach was introduced because of the need for multi-layer access to the system, including desktop and mobile clients, simultaneous operation of multiple machines, and access from multiple locations.

Relational Data Model

The data model was designed as a multi-context relational system that explicitly models the organizational structure, access levels, machine context, and process data. At the top of the hierarchy are the companies and locations entities, which define the organizational and spatial context of the system. The machines entity represents the central connection between the physical infrastructure and the process layer, with each machine being linked to a company and location, and further associated with controllers, process parameters, sessions, events, and operator access. The *machine_controllers* entity is used to record the identity and status of control units, including controller type, firmware version, and operational data relevant for supervision. The structure of the centralized relational data model and the relationships between its main entities are shown in Fig. 6.

The operator layer is modeled through the operators entity, along with the additional relations *operator_location_access* and *operator_machine_access*, thereby establishing multi-level access control. In this way, it is possible to precisely define which operator is allowed to work at particular locations and on particular machines. The *operator_sessions* entity enables the recording of logins, logouts, and session statuses, thereby introducing additional traceability of operator activities.

The process layer is modeled through the entities *presets, sequence_articles*, and *sequence_steps.* The presets entity includes predefined position presets, while the *sequence_articles* and *sequence_steps* entities model work sequences through logical grouping and the elaboration of individual steps. Such an organization enables process patterns to be stored, shared, reused, and adapted to different operators, machines, or locations. The *machine_parameters* entity is used to store machine configuration parameters, such as steps per millimeter, speeds, and accelerations, while *machine_state_snapshots* models current or periodically recorded machine states, including position, calibration status, reference mode, and end-limit states. In this way, the database stores not only static configuration information, but also dynamic operational data. The event and revision layer is modeled through the entities *machine_events* and *audit_logs*. The *machine_events* entity contains process and system events related to a specific machine and operator, while *audit_logs* provides a generic trace of administrative and entity-level changes in the system.

Long-Term Data Management and Database Scalability

Since the machine_state_snapshots table may

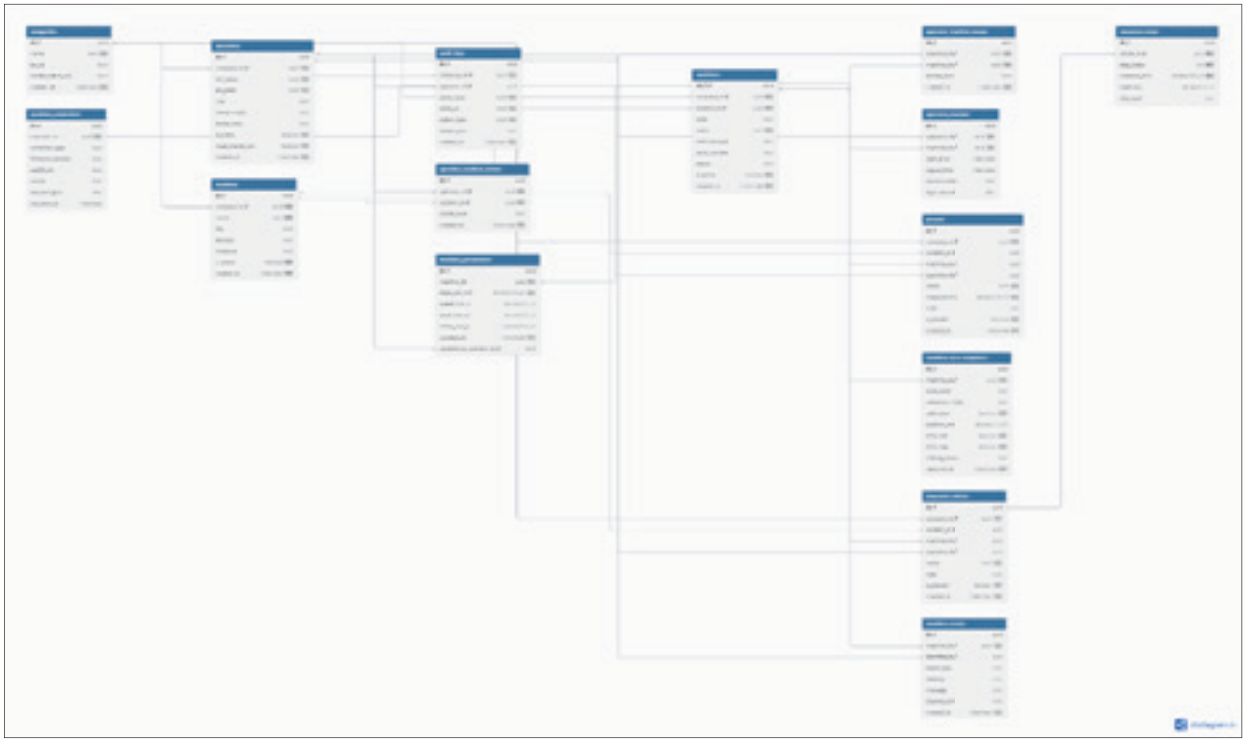


Figure 6. *Entity-relationship schema of the centralized relational database*

grow rapidly in a multi-machine environment, long-term data management must be considered as part of the database architecture. In the implemented model, current machine states and historical state records are conceptually separated. Recent records are used for supervision, diagnostics, and synchronization, while older records can be archived or aggregated depending on maintenance and reporting requirements.

For larger deployments, the table can be optimized using indexes over machine_id and timestamp fields, time-based partitioning, and retention policies that keep high-resolution state data only for a defined operational period. Older snapshot data may be transferred to historical tables or summarized into aggregated diagnostic records. In this way, the system preserves traceability without allowing high-frequency operational data to reduce database performance over time.

Integrity Constraints and Transaction Management

The quality of the model is reflected not only in the number of entities, but also in the way constraints are defined over the data. Primary keys ensure the unique identity of each record, while foreign keys prevent illogical associations between entities. Mandatory fields are defined as NOT NULL, and appropriate relations apply cascading deletion or NULL assignment rules depending on the semantics of the data. In this way, the database actively contributes to preserving system integrity.

In more complex operations, such as creating a work sequence with multiple steps, assigning access to an operator, or changing machine configuration, transactional consistency plays a key role. Such operations represent a set of mutually related writes that must be treated as a single logical unit. The PostgreSQL transaction mechanism ensures that such units are either fully committed or completely rolled back in the event of an error, thereby preventing partial writes and preserving the logical correctness of the system [9].

Multi-Client Access and Remote Availability

The centralized database allows an operator to retain access to personal settings, work measures, sequences, tasks, and work history regardless of the machine being used, provided that the machine belongs to the same system environment. In this way, a unified user and process context is achieved across multiple machines, which is particularly important in production systems where operators work on the same or similar machines distributed across multiple workstations or locations. Access to the database is organized through the application layer, which separates business logic from the query layer itself, thereby achieving clearer code organization, easier testing, and greater maintainability of the system. In the implemented architecture, desktop and mobile clients do not access the PostgreSQL database directly, instead, all interactions with centralized data are mediated through the application service layer within the Tailscale VPN environment, which enforces access control and isolates the database from direct client exposure.

The use of a centralized PostgreSQL solution, instead of a local database tied to a single workstation, naturally supports multi-client operation, central administration, and the availability of the same data across multiple devices.

The complete work process can be monitored and administered remotely using a dedicated Flutter mobile application [10] and desktop client, with access established through Tailscale VPN infrastructure. In the implemented solution, Tailscale enables secure connection of devices within a closed virtual network, without the need for public exposure of services or complex configuration of traditional VPN solutions. In this way, the mobile application, remote computer, and the server environment hosting the PostgreSQL database remain available through controlled, identity-based access, which supports secure supervision, centralized administration, and reliable access to operational and configuration data from different locations. The administrative interface for remote monitoring of service and database status is shown in Fig. 7.

Figure 7. *Administrative interface for remote monitoring of service and database status*

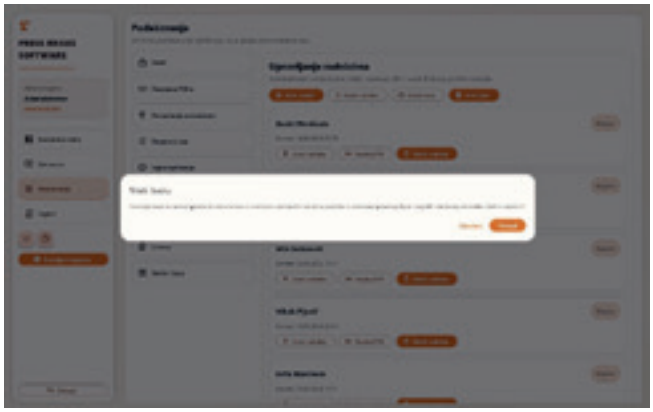


Figure 8. *Interface for restoring the database from a backup file*

The practical value of this architecture lies in the fact that all centrally defined data related to a specific worker are available through the VPN, including their settings, predefined measures, work tasks, and activity history. In this way, the system goes beyond the framework of a local control application tied to a single machine and becomes a centralized industrial software environment supporting multiple users, multiple devices, and multiple locations.

Data Portability and Security Mechanisms

An additional operational value of the system lies in its support for data export and import. The implemented approach enables backup creation, data migration, and restoration of the system state whenever servicing, infrastructure replacement, or transfer of configuration to another controlled environment is required. In this way, the system supports not only routine administrative procedures, but also recovery scenarios in which continuity of operation must be preserved. Export and import mechanisms [11] make it possible to transfer operator data, presets, work sequences, and related configuration elements without the need for repeated manual setup. This reduces downtime, simplifies maintenance procedures, and improves portability of the software solution across different workstations and deployment contexts. Consequently, these mechanisms do not represent merely auxiliary administrative functions, but an important element of system continuity, maintainability, and long-term operational reliability. An example of the interface used for database restoration from a backup file is shown in Fig. 8.

The security aspect of the system does not relate only to the protection of user access to the application, but also to the protection of the software solution itself against unauthorized use. For this reason, two mutually complementary protection mechanisms have been applied. The first relates to operator authentication, where login is performed using a PIN whose value is stored in the database in hashed form, thereby reducing the risk associated with direct exposure of authentication data and improving protection of user credentials. In addition, this approach supports personalized access to operator-specific settings, presets, and work sequences while ensuring that unauthorized users cannot access protected system functions. The operator login interface used for authentication and loading of personalized user settings is shown in Fig. 9.

Figure 9. *Operator login interface of the implemented industrial control application*

The second relates to software license protection using asymmetric cryptography [12]. In this model, the private key remains with the system author and is used for digitally signing the license, while the public key is embedded in the application and is used to verify its authenticity. Such an approach ensures that valid licenses can only be issued within the controlled service environment and that the verification process can be performed locally within the application itself. In this way, both user-level access protection and software-level authorization are addressed as integral parts of the overall system security model.

The implemented licensing subsystem also includes a dedicated internal License Manager application intended exclusively for the software author or authorized service personnel. The tool processes exported request files generated by the main application and enables license generation, renewal, and activation-code export. The licensing workflow supports both perpetual and time-limited licenses and is fully based on the same asymmetric cryptographic model, in which the private signing key is retained exclusively within the internal service environment. The administrative interface used for license verification and activation is shown in Fig. 10.

Figure 10. *Administrative interface for license verification and activation*

This architecture prevents local generation or modification of a valid license by the end user, while license verification remains possible without relying on a constant Internet connection or a remote license server.

Results

The performed validation confirmed that the implemented system provides stable and consistent operation of the centralized data model, distributed communication layer, and remote multi-client access architecture. The obtained results indicate that the proposed software-hardware integration satisfies both the functional requirements of precise industrial control and the organizational requirements of centralized data management in multi-machine environments.

At the application and database level, the centralized relational model enabled reliable management of operator accounts, position presets, work sequences, assigned tasks, and system events within a unified information environment. The average operator login time, including authentication, access-right verification, and loading of the user context, was 142 ms, while loading of position presets and work sequences required 31 ms and 56 ms, respectively. These values indicate that centralized persistence does not introduce delays that would negatively affect routine industrial use.

At the communication level, validation showed that the hierarchical organization based on the main control node and remote execution and sensor modules enables reliable exchange of commands and status information under industrial operating conditions. The average response time between the application layer and the main ESP32-S3 node over the USB-UART connection was 48 ms, while the average exchange time between the main node and remote modules over the CAN bus was 14 ms. No communication loss was observed during the test scenarios, and the functional separation between execution and sensing layers was preserved in all cases. These results confirm that the selected communication architecture provides both responsiveness and robustness in environments exposed to electromagnetic disturbances.

Remote access through the Tailscale VPN infrastructure was also successfully validated. Desktop and mobile clients were able to access centralized data and administrative functions with an average remote response time of 187 ms under normal operating conditions. This result confirms that the developed architecture supports centralized supervision and coordinated management of multiple users,

machines, and locations within a single software environment, without compromising practical usability.

At the database consistency level, the correctness of relations between the organizational, machine, operator, process, and revision layers was verified through complex transactional operations. Work-sequence modification, access-right assignment, and machine-parameter updates were executed with a 100% success rate in the validation scenarios, without integrity violations or partial writes. Export and import of configuration and operational data required 3.2 s and 4.7 s, respectively, confirming that backup, migration, and restoration procedures can be performed efficiently within maintenance and recovery workflows.

From the industrial perspective, the implemented solution improved operational consistency, reduced dependence on machine-local data, and increased the availability of process-related information. Centralized storage of operator settings, work measures, and task assignments enabled the same operator to continue work on different machines without repeated manual reconfiguration, thereby reducing setup effort and improving workflow continuity. In addition, the availability of logs, audit records, and machine-event history improved process traceability, diagnostics, supervision, and accountability in production conditions.

The obtained results also show that the distributed architecture contributes to maintainability and scalability of the overall system. By separating execution and sensing functions across dedicated remote modules and interconnecting them through the CAN bus, the implemented solution reduced sensitivity to long signal lines, improved communication robustness, and supported modular deployment across multiple machines and locations. In that sense, the proposed architecture provides not only a technically valid integration of software, database, and communication layers, but also measurable operational benefits relevant to industrial practice.

DISCUSSION

The obtained results confirm the justification for selecting a centralized relational data model for the analyzed type of industrial software system. The nature of the data within the system clearly corresponds to the relational approach, since the entities are predefined, their relationships are stable, and preserving the integrity of those relationships is essential for the correct operation of the system. Under such conditions, a relational database management system represents the most suitable architectural and functional solution.

The choice of PostgreSQL proved justified due to the centralized nature of the system. Unlike local databases tied to a single machine or workstation, a centralized RDBMS allows the same operator to access their data from multiple machines and different client platforms while preserving a unified operator and process context. Such an approach is particularly important in systems with multiple workstations, multiple users, and multiple locations, where local persistence would not be able to provide the required level of consistency, availability, and administrative transparency.

An additional value of the implemented solution lies in the breadth of the data model, which includes companies, locations, machines, controllers, access rights, sessions, machine parameters, states, events, and audit trails. In this way, the organizational, machine, operator, process, and revision layers of the system are integrated within a single database, enabling not only reliable persistence but also improved traceability, supervision, and process management.

From the communication perspective, the distributed architecture based on the main control node and remote execution and sensor modules proved suitable for the industrial environment. The choice of the CAN bus is particularly significant, since differential signal transmission provides greater resistance to electromagnetic interference typical of systems with electric motors, contactors, variable-frequency drives, and power lines. In this way, the communication layer does not represent only a transmission mechanism, but also an important element of the overall robustness and maintainability of the system. Remote access via the Tailscale VPN infrastructure further extends the functional scope of the solution by enabling centralized administration, log review, task assignment, and access to user settings from different locations.

The comparison with NoSQL approaches further confirms the appropriateness of the selected solution. A NoSQL approach may be useful in systems dominated by unstructured telemetry, high-volume

sensor streams, or document-oriented data storage. However, the analyzed industrial control system is based on clearly defined entities and stable relationships between operators, machines, locations, access rights, process parameters, work sequences, sessions, events, and audit records. In such a context, relational integrity is more important than flexible data representation.

For example, when an operator is granted access to a specific machine at a specific location, this relationship must remain consistent with the corresponding company, location, machine, session, and audit records. In a document-oriented NoSQL model, such information would often have to be duplicated or coordinated across multiple documents, which increases the risk of inconsistent access rights, outdated copies of security-related data, and more complex validation logic at the application level. In contrast, PostgreSQL enforces these relationships through foreign keys, constraints, transactions, and rollback mechanisms.

Therefore, NoSQL systems could be considered as an auxiliary layer for high-volume telemetry, analytical storage, or non-critical log aggregation, but they are less appropriate as the primary source of truth for this type of safety- and integrity-sensitive industrial application. For the implemented system, the centralized relational model provides clearer consistency rules, stronger auditability, and more reliable management of operator access, machine parameters, and work-sequence modifications.

Based on this, it can be concluded that the main contribution of the paper does not lie only in the application of individual technologies, but in their architectural integration. The centralized database layer, distributed communication architecture, multi-client access, and security mechanisms are integrated into a unified solution tailored to industrial precision-control systems. It is precisely in this integration that the key value of the proposed model can be recognized.

Conclusion

This paper examined the application of a centralized relational data model and distributed communication architecture in an industrial software system intended for high-precision machine control. The analysis showed that such systems require not only reliable control of process-related functions, but also structured management of operator accounts, position presets, work sequences, access rights, system events, and audit-related data. In that context, the relational model proved to be a suitable foundation for organizing and maintaining consistency of industrial operational data.

The proposed architecture combines a centralized PostgreSQL persistence layer, a distributed microcontroller-based communication structure, and multi-client remote access within a unified software environment. Such integration enables a clear separation between application, control, sensing, and data-management layers, while also supporting centralized supervision, improved traceability, and coordinated operation across multiple machines and locations. The obtained findings confirm that this approach is functionally, technically, and architecturally appropriate for industrial environments in which precision, consistency, and operational continuity are critical.

An additional contribution of the paper lies in showing that the selected architecture provides not only technical feasibility, but also practical industrial value through improved maintainability, centralized administration, controlled access, and support for backup, recovery, and licensing mechanisms. At the same time, the study indicates that the developed model provides a solid basis for further extension toward advanced supervision, analytics, production-resource coordination, and broader integration of multiple industrial machines into a shared information environment.

References

[5] H. Desamsetti, “Relational Database Management Systems in Business and Organization Strategies,” Global Disclosure of Economics and Business, vol. 9, no. 2, pp. 151–162, Jul. 2020, doi: 10.18034/gdeb.v9i2.700.

[6] N. M. Ahmed and G. M. Haji, “TRADITIONAL RDBMS TO NOSQL DATABASE: NEW ERA OF DATABASES FOR BIG DATA,” International Journal of Scientific & Technology Research, vol. 10, no. 9, pp. 101–106, Sep. 2021.

[7] Google, “Flutter architectural overview,” Flutter Documentation, official documentation. Available: https://docs.flutter.dev/resources/architectural-overview

[8] Tailscale Inc., “What is Tailscale?,” Tailscale Documentation, official documentation. Accessed: May 12, 2026. [Online]. Available: https://tailscale.com/docs/concepts/what-is-tailscale

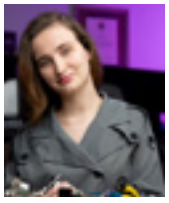
[9] D.-F. Hrițcan and D. Balan, “Using Tailscale and PfSense for Security and Anonymity of IoT Environments,” in 2024 International Conference on Development and Application Systems (DAS), 2024, doi: 10.1109/DAS61944.2024.10541192.

[10] Texas Instruments, "Introduction to the Controller Area Network (CAN)," Application Report, Rev. B.
[11] V. P. Oleksiuk and O. R. Oleksiuk, "The practice of developing the academic cloud using the Proxmox VE platform," Educational Technology Quarterly, vol. 2021, no. 4, pp. 605–616, Dec. 2021, doi: 10.55056/etq.36.
[12] Proxmox Server Solutions GmbH, "Proxmox VE Administration Guide," Proxmox Virtual Environment Documentation, official documentation. Accessed: May 12, 2026. [Online]. Available: https://pve.proxmox.com/pve-docs/pve-admin-guide.html
[13] The PostgreSQL Global Development Group, "Transactions," PostgreSQL Documentation, official documentation. Accessed: May 12, 2026. [Online]. Available: https://www.postgresql.org/docs/current/tutorial-transactions.html
[14] K. C. Panda, "Application Development Using Flutter and React Native: Cross Platform Development," Journal of Research in Science and Engineering, vol. 7, no. 1, pp. 6–8, Jan. 2025, doi: 10.53469/jrse.2025.07(01).02.
[15] D. Bordoloi, "Import and Export Database Management System," Mathematical Statistician and Engineering Applications, vol. 70, no. 1, pp. 182–189, Jan. 2021, doi: 10.17762/msea.v70i1.2298.
[1] C. Stohrer and T. Lugrin, "Asymmetric Encryption," in Trends in Data Protection and Encryption Technologies, V. Mulder et al., Eds. Cham, Switzerland: Springer, 2023, pp. 11–14, doi: 10.1007/978-3-031-33386-6_3.

Received: April 17, 2026
Accepted: May 5, 2026

ABOUT THE AUTHORS

Daniel Menićanin is a student at the Faculty of Information Technology, Pan-European University Apeiron, specializing in Programming and Software Engineering. His work is focused on robotics, industrial automation, and advanced software solutions for CNC and hydraulic systems. Throughout his academic career, he has participated in numerous innovation projects, including the development of adaptive gaming controllers and industrial control software. For his achievements, he received several awards at innovation and technology conferences in Bosnia and Herzegovina. His current research interests include embedded systems, industrial communication, and intelligent automation technologies.



Jelena Radanović is a Programming and Software Engineering student at Pan-European University Apeiron. She has been actively involved in software development and innovation projects related to industrial applications and modern programming technologies. Her work combines technical problem-solving with creative software design, particularly in automation and user-oriented systems. Jelena has received recognition for her contributions at regional innovation events and continues to expand her expertise in software engineering, application development, and emerging technologies.

Dražen Marinković received his M.Sc. degree in 2015 and Ph.D. degree in 2020 from the Faculty of Information Technology at Pan-European University Apeiron in Banja Luka. He currently works as an associate professor at the same institution. His academic and research activities are focused on computer networks, data science, distributed systems, and modern computing technologies. He has participated in various scientific and educational projects related to information technologies and software systems.

FOR CITATION

Daniel Menićanin, Jelena Radanović, Dražen Marinković, Centralized RDBMS and Distributed Communication in a Software System for High-Precision Industrial Machine Control, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:20-30, (UDC: 004.6:[004.9:004.451.9), (DOI: 10.7251/JIT2601020M), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Analysis of the expected effects of implementing ERP, GIS and DMS systems in a public enterprise

Sasa Ljubojevic[1], Branko Latinovic[2]
[1]*Public Forestry Enterprise "Forests of the Republic of Srpska", Banja Luka, Bosnia and Herzegovina, sasa.ljubojevic@sumers.org, ORCID ID: 0000-0001-7732-8088*
[2]*Pan-European University Apeiron, Faculty of Information Technology, Banja Luka, Bosnia and Herzegovina, branko.b.latinovic@apeiron-uni.eu, ORCID ID: 0009-0008-3766-3699*

Original scientific paper
https://doi.org/10.7251/JIT2601031LJ UDC: 004.655.3:004.652.4

Abstract: Integrated information systems represent a key element of the digital transformation of public enterprises, enabling data integration, improvement of business process control, and an increase in transparency. This paper analyzes the effects of implementing an integrated information system in the Public Forestry Enterprise "Forests of the Republic of Srpska" JSC Sokolac, which integrates SAP ERP, a Geographic Information System (GIS), and a Document Management System (DMS).
Preliminary implementation results indicate concrete operational improvements, including the elimination of a 24-hour data synchronization delay, reduction of the field data collection cycle from 48 hours to near-real-time, instant document retrieval through the DMS, and automatic invoice generation upon mobile application synchronization. Once fully operational, the system is expected to further improve efficiency, control, transparency, and analytical management.
Special focus is placed on the identification of effects in the areas of operational efficiency, work organization, control and transparency of business operations, as well as the improvement of analytical capacities through the application of BI and OLAP technologies.
The research results indicate that the implementation of an integrated information system creates the preconditions for data centralization, standardization of business processes, and the establishment of a unified information environment. The integration of SAP ERP and GIS enables more comprehensive monitoring of the forest timber assortment production process and the improvement of analytical capacities through the application of BI tools.
A particularly significant contribution of the system is reflected in the possibility of linking planned and realized activities, thereby creating the preconditions for more efficient business monitoring and decision-making based on reliable data.
The integrated information system represents an important tool for improving efficiency and data-driven decision-making, while the achieved implementation results indicate significant potential for further development through advanced analytical methods and digital resource management.

Keywords: integrated information system, SAP, ERP, GIS, DMS, public enterprises

Introduction

Modern business systems are characterized by a high level of complexity, a large amount of data, and the need for fast and reliable decision-making. In such an environment, information systems represent a key factor in the efficient management of organizations, especially in the public sector, where the requirements for transparency, control, and rational use of resources are particularly pronounced [1], [4]. The integrated information system in the Public Forestry Enterprise "Forests of the Republic of Srpska" is currently in the final phase of implementation, and the system is expected to achieve full operational capacity and provide complete support for business process management, control, transparency, and analytical decision-making.

Public enterprises that manage natural resources, such as the Public Forestry Enterprise "Forests of the Republic of Srpska," face specific challenges that include the management of spatially distributed resources, coordination of a large number of activities,

and processing of heterogeneous data. Under such conditions, traditional decentralized information systems often cannot provide an adequate level of integration and reliability of data, which leads to reduced business efficiency and more difficult decision-making [4], [5].

Modern forestry management is increasingly influenced by concepts of digital transformation and Industry 4.0, including the application of ERP systems, GIS technologies, mobile platforms, cloud infrastructure, and advanced analytical tools. In contemporary forestry systems, digital technologies enable more efficient resource management, improved monitoring of field activities, real-time data collection, and data-driven decision-making processes. The integration of spatial and business data through unified information systems represents one of the key directions in the modernization of forestry organizations and public enterprises managing natural resources [6], [13].

The earlier information system in the Public Forestry Enterprise "Forests of the Republic of Srpska" was based on a decentralized approach, whereby organizational units used different databases and their own coding systems. This approach led to data fragmentation, inconsistency of information, and limited possibilities for analysis. As a consequence, it was difficult to monitor business processes and establish a unified control system [8]–[10].

The introduction of an integrated information system represents a strategic step toward improving business operations and the digital transformation of the enterprise. The new system is based on the integration of an ERP system, a Geographic Information System (GIS), and a Document Management System (DMS), thereby creating conditions for the integration of business, spatial, and documentary data into a unified information environment [1], [4], [5]. Since the system is still in the final phase of implementation, the full effects of this integration are expected after its complete functional completion.

The aim of this paper is to examine the role of the integrated information system and its implementation in the Public Forestry Enterprise "Forests of the Republic of Srpska," with particular focus on improving the efficiency of business processes, increasing the level of control and transparency, and developing analytical capacities through the integration of different information components. Special attention is devoted to the role of the BI system as an analytical layer that enables the integration and interpretation of data from different sources [3], [12].

The paper is structured so that, after the introductory part and methodology, a description of the integrated information system follows, while the central part of the paper analyzes the effects of its implementation. The final part of the paper includes conclusions and recommendations for the further development of the system.

Research Methodology

The methodological framework of this paper is based on the application of the case study method, which enables a detailed analysis of the implementation of an integrated information system in a real organizational environment. This approach is particularly suitable for the study of complex information systems, as it enables consideration of the technical, organizational, and process aspects of their application in practice [1], [4].

The subject of the research is the integrated information system implemented in the Public Forestry Enterprise "Forests of the Republic of Srpska", which consists of three basic components: the ERP system, GIS, and DMS. Together, these components form a unified information framework that enables the integration of data and business processes at the level of the entire enterprise [8]–[10].

A combined methodological approach was applied in the paper, including qualitative and descriptive-analytical analysis. The qualitative aspect of the research is reflected in the analysis of the system implementation process, organizational changes, and the challenges that arise during the transition from a decentralized to a centralized system. The descriptive-analytical approach is used to identify and interpret the effects of system implementation on the enterprise's operations.

The research methodology includes three key phases. The first phase relates to the analysis of the situation before the implementation of the integrated information system, during which the basic problems of the decentralized approach were identified, including data fragmentation, inconsistency of coding systems, and different working methods among organizational units. The second phase includes the analysis of the system implementation process, with

particular focus on the integration of the SAP ERP, GIS, and DMS components, as well as on challenges in the area of data migration and standardization. The third phase relates to the evaluation of the effects of system implementation, during which changes in business efficiency, the level of control and transparency, and the analytical capacities of the organization are analyzed [8]–[10].

For the purposes of the research, internal enterprise documents, project documentation, and operational data generated during system implementation were used. In addition, an important source of information consists of practical experience gained through work on the implementation and use of the system, which enables a realistic and detailed insight into the functioning of the integrated information system.

A special aspect of the methodology relates to the analysis of the integration of different information components. The SAP ERP system is observed as the central operational system that generates and processes business data, GIS as the system that provides the spatial context of those data, while DMS enables document management and records of business processes. The analysis of their mutual integration enables an understanding of the way in which the integrated information system contributes to business improvement [1], [5].

For the purpose of a comprehensive evaluation of the effects of implementation, the results were analyzed through several dimensions, including operational efficiency, data quality, the level of process standardization, and the possibilities of analytical data processing [3], [12]. Such an approach enables an objective assessment of the impact of the integrated information system on the enterprise's operations.

The research was conducted during the implementation phase of the integrated information system between 2020 and 2025. The analysis included data and observations collected throughout different implementation stages, including system design, data migration, integration, testing, and operational deployment.

Data collection methods included:

- analysis of internal project documentation,
- analysis of conceptual designs and user manuals,
- operational reports and system records,
- direct observation during implementation activities,
- consultations and interviews with employees involved in implementation and operational processes.

Such an approach enabled a more comprehensive understanding of both technical and organizational aspects of the implementation process.

The methodological approach applied in this paper enables the identification of the key advantages and limitations of the implemented system and represents a basis for further research in the field of the application of integrated information systems in the public sector.

Integrated Information System Architecture

The integrated information system in the Public Forestry Enterprise "Forests of the Republic of Srpska" represents a complex and integrated information framework that includes three key components: the SAP ERP system as the central business system, the Geographic Information System (GIS) for the management of spatial data, and the Document Management System (DMS) for document management. Together, these components enable the integration of data and business processes at the level of the entire enterprise, thereby achieving a high level of integration, control, and efficiency [8]–[10].

The introduction of the integrated information system represents a significant step forward compared to the previous decentralized model of work, in which organizational units used different databases and their own coding systems. Such an approach led to data fragmentation, inconsistency in work, and more difficult monitoring of business operations. By implementing the new system, a unified information framework was established, enabling data centralization and standardization of business processes (Figure 1).

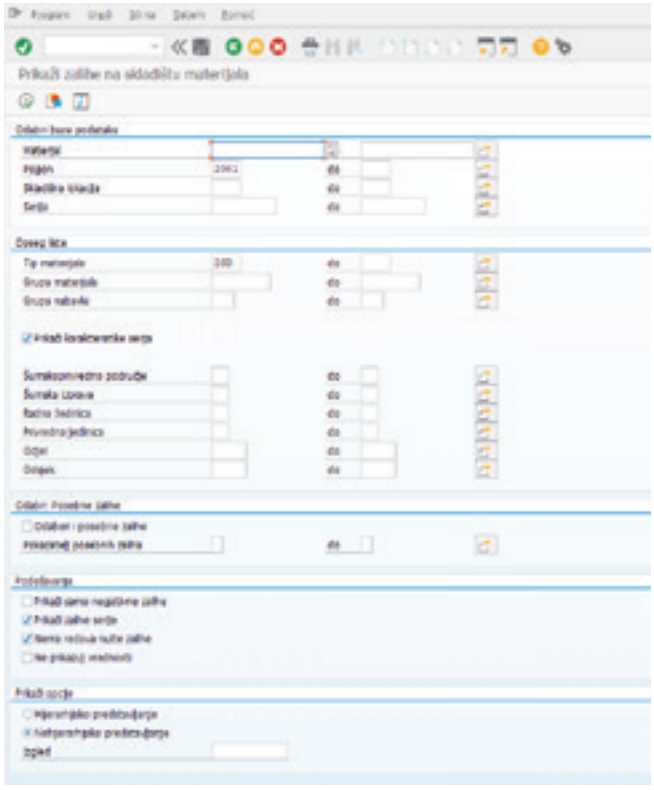


Figure 1. *User interface of the SAP ERP system*

The central role in the system is played by the ERP platform, which represents the basis for managing all key business functions. Through the implementation of the ERP system, the integration of procurement, sales, finance, payroll calculation, and project management processes was enabled. The MM (Materials Management) module is used for the procurement of goods and services, including specific processes in the field of forestry, while the SD (Sales and Distribution) module supports the sales processes of forest timber assortments. The FICO module enables the management of financial flows and controlling, while FICO/PY covers payroll and labor cost calculation [1], [7].

Of particular importance are PS (Project System), that is, WBS elements, which enable the planning, monitoring, and control of activities through projects, including silvicultural works, exploitation, and other operational processes. The PM (Plant Maintenance) module is used for fleet management and equipment maintenance, thereby further improving operational efficiency and resource availability [1], [7].

The DMS component of the system is used for electronic recording and document management, including electronic protocol records, electronic document exchange, and records of court cases (Figure 2). By introducing the DMS system, the digitization of business documentation was enabled, thereby improving data availability, reducing the use of paper documentation, and increasing efficiency in the management of administrative processes [8]–[10].

The GIS component of the system represents a key element for the management of spatial data and

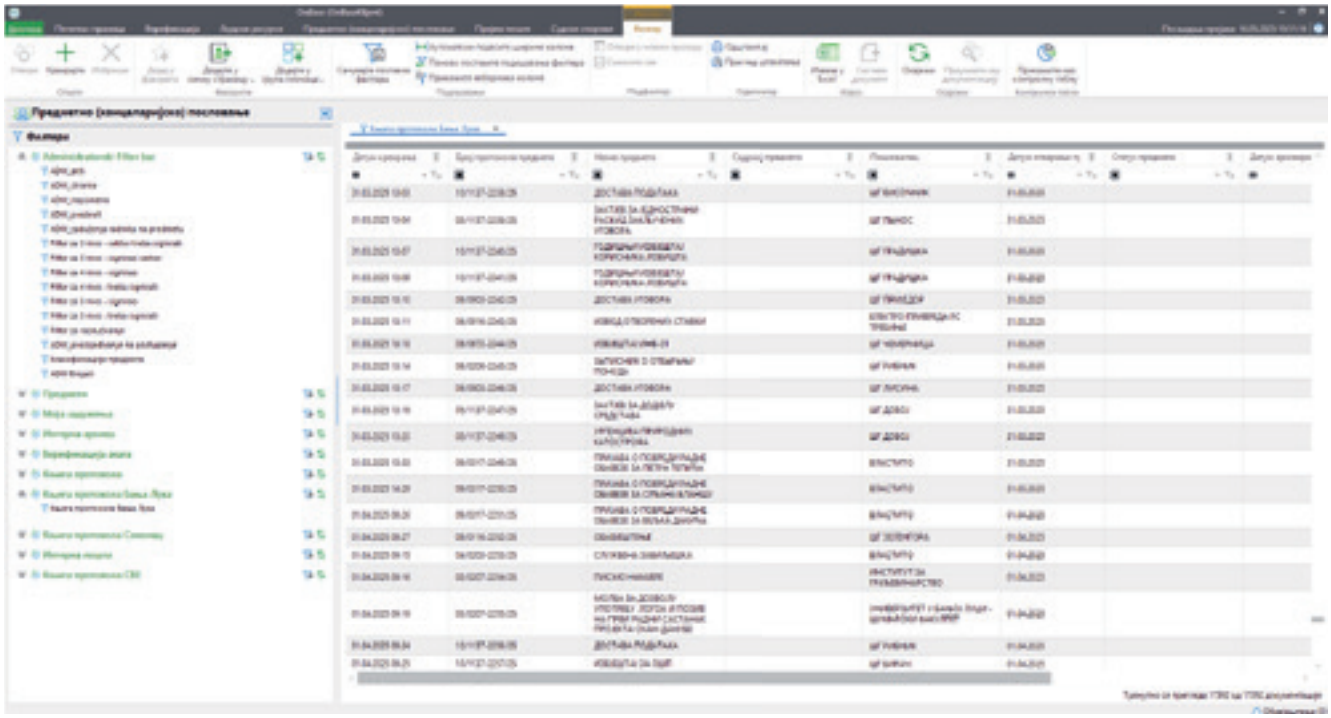


Figure 2. *DMS system interface*

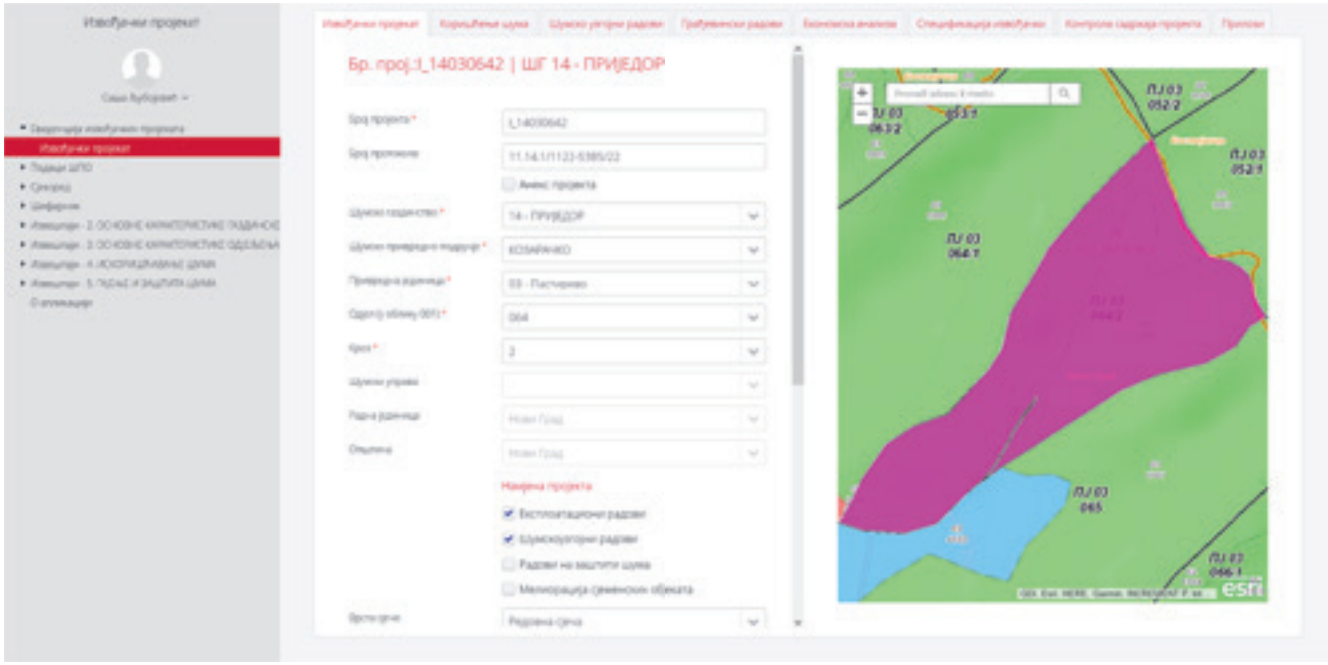


Figure 3. *GIS system interface*

their integration with business processes (Figure 3). Through the establishment of a unified GIS database, all spatial data that had previously been distributed across several different databases were integrated. This process was accompanied by significant challenges, especially with regard to overlapping polygons, inconsistencies in boundaries, and different ways of naming spatial objects, which required detailed data standardization and harmonization [5], [8]–[10].

The GIS system includes several thematic units, including forest management plans, operational projects, forest cadastre and felling schedules, forest communications, private forests, and forest protection. In this way, detailed planning and monitoring of field activities are enabled with a high level of precision and data reliability.

A particularly significant role within the integrated information system is played by mobile applications, which represent the link between field activities and the central information system. Two key mobile applications have been implemented: the GIS mobile application and the SAP mobile application.

The GIS mobile application enables spatial and attribute data to be collected directly in the field. Through this application, it is envisaged that data will be collected for the preparation of forest management plans (forest inventory), as well as data for operational projects, that is, the marking of trees. In addition, the GIS mobile application is also used in the forest protection segment, where it enables the recording of locations of illegal logging, pheromone traps, and polygons of burned areas [13].

An important aspect of this application is the possibility of electronic completion of forms that were previously in paper form. In this way, automatic centralization and availability of data are enabled, whereas previously such data remained in written form in archives and were not easily available for analysis and control.

Figure 4. *SAP mobile application integrated with a portable printer*

The SAP mobile application serves to record the movement of forest timber assortments through all

phases of the production process. This application enables the recording of felled trees, their extraction to truck roads, and the dispatch of forest timber assortments. This process envisages the recording of detailed data on each tree, as well as information on the type of felling and the method of carrying out the work (Figure 4).

The system enables precise recording of whether the works were performed by the enterprise's own workforce, in which case data on the executors are recorded (felling and extraction groups), or whether external contractors were engaged, thereby establishing a direct link with work performance contracts. In the dispatch phase, a connection is established with the buyer and sales contracts, whereby the system automatically checks whether all contractual and financial obligations have been fulfilled. If the buyer has outstanding obligations, it is not possible to issue a dispatch order, which provides an additional level of control.

One of the key characteristics of the system is the complete monitoring of all activities. Each phase of the process records data on completed works, including information on who performed a given activity, when, and in what manner. In this way, a clear and reliable trace is provided for all phases of work, which will significantly contribute to transparency and control of operations.

The integration of ERP, GIS, and DMS systems, supported by mobile applications, enables the consolidation of operational, spatial, and documentary data into a unified whole. Such an approach provides the basis for improving efficiency, increasing transparency, and developing advanced analytical capabilities, thereby positioning the integrated information system as a key tool for managing enterprise operations.

An important segment of the integrated information system is the BI (Business Intelligence) component, which enables the integration and analysis of data from ERP, GIS, and DMS systems. The BI system takes over data from the GIS component related to felling plans and planned activities and links them with data from the ERP system covering the realization of felling, extraction, and dispatch of forest timber assortments. In this way, the integration of planned and realization data is enabled, thereby providing a complete insight into production and realization flows [2], [3], [12].

Through such an integrated approach, the BI system enables the generation of reporting documentation that links planned and realized activities, thereby enabling continuous monitoring of deviations and the efficiency of plan realization. A particularly important aspect of this approach is reflected in the possibility of analyzing the entire process, from planning, through work execution, to final realization and sale.

In addition, the BI system enables the generation of a large number of standard and customized reports, including reports on work monitoring, stock levels, movement of forest timber assortments, realization of contracts, as well as monitoring of planning documents. These reports enable a detailed insight into the enterprise's operations through different dimensions, thereby significantly improving control and support to the decision-making process [2], [3], [12].

By integrating data from different sources and analyzing them through the BI system, a unified and reliable source of information is provided, which represents the basis for efficient management of business processes and strategic planning.

RESULTS AND DISCUSSION

Preliminary implementation results and operational observations indicate several important organizational and technological improvements. The implementation of the integrated information system in the Public Forestry Enterprise "Forests of the Republic of Srpska" has already led to certain changes in the way the enterprise functions. The effects of implementation can be observed through several interrelated dimensions, including operational efficiency, work organization, business control, and the development of analytical capacities.

One of the most significant observed and expected effects of system implementation is reflected in the improvement of operational efficiency. By introducing the ERP system and integrating it with GIS and DMS components, the automation of a large number of business processes has been enabled. This reduces the need for multiple data entries, decreases the possibility of errors, and accelerates information processing. Mobile applications contribute to improved data timeliness, as they enable data entry directly in the field and its availability in the central system [8]–[10].

At the organizational level, the implementation of the integrated information system has contributed

to the standardization of business processes and the unification of work methods across organizational units. The previously decentralized system, in which different parts of the enterprise used their own procedures and coding systems, is gradually being replaced by a unified work model based on centralized data and standardized rules. In this way, conditions have been created for greater consistency of work, better coordination, and a clearer distribution of responsibilities [1], [4].

A particularly important expected effect of system implementation is reflected in the increase in the level of control and transparency of business operations(Table 1). Through the integration of ERP and GIS systems, more complete monitoring of timber assortments is enabled through all phases of the process, from planning and felling to transport and sale. Each activity is recorded with relevant data on time, location, and executor, thereby creating conditions for more efficient supervision and more reliable business control.

Data centralization represents one of the key expected effects of system implementation. The previous system was characterized by data fragmentation and the existence of different material coding systems, which made integration and analysis more difficult. By introducing a unified database and standardizing coding systems, data consistency, easier accessibility, and a stronger basis for reporting and management have been established [8]–[11].

A significant expected effect of the implementation of the integrated information system relates to the improvement of the organization's analytical capacities. The BI component of the system enables the integration of data from the ERP and GIS systems and their analysis through different dimensions. Particularly important is the linking of planning data from the GIS system with realization data from the ERP system, thereby creating the preconditions for more precise monitoring of plan implementation and the identification of deviations [2], [3], [12].

Through the BI system, it is possible to generate a large number of reports that cover different aspects of operations, including work monitoring, stock levels, movement of forest timber assortments, realization of contracts, and execution of planning documents. These reports enable a more detailed insight into operations and represent the basis for more efficient resource management and decision-making.

The special value of the BI system is reflected in the possibility of integration with the GIS component, thereby enabling spatial data analysis. This approach enables business activities to be viewed in their geographical context, which is of particular importance in forestry. In this way, it would be possible to identify the spatial distribution of activities more precisely, monitor the execution of works, and analyze resource utilization [5].

Several concrete operational improvements have already been observed during the implementation process. In the area of data availability, the previous system relied on nightly synchronization between decentralized organizational units, which meant that all data in the system were at least 24 hours old at

Table 1. *Preliminary indicators and expected operational improvements*

Business Process	Previous State	Current / Expected State	Expected Improvement
Document processing	Paper-based workflow	Digital DMS workflow	Faster document availability
Field data collection	Manual paper forms	Mobile GIS data collection	Real-time data availability (48h reduction)
Timber tracking	Partial traceability	Integrated SAP-GIS tracking	Increased transparency
Reporting generation	Manual report preparation	Automated BI reporting	Reduced reporting time
Data storage	Decentralized databases	Centralized database	Improved data consistency (24h nightly sync eliminated)
Monitoring of contracts	Manual verification	ERP-based control mechanisms	Improved operational control
Planning and realization analysis	Separate systems	Integrated BI analysis	Better decision-making support
Document retrieval (DMS)	Physical archive search (minutes to hours)	Instant electronic search by protocol no., date, or sender	Near-instant retrieval (search time eliminated)
Invoice generation	Handwritten dispatch note → manual re-entry (next day) → invoice: min. 24h delay	Mobile app sync → automatic invoice generation and delivery to buyer	24h reduction in billing cycle

any given point. The new integrated system eliminates this delay, providing near-real-time data availability across all organizational units. In the case of field data collection, the previous workflow required field workers to complete handwritten forms on-site, manually re-enter the data into the system the following day, and then wait for an additional synchronization cycle — resulting in a total data delay of up to 48 hours. The integrated mobile application eliminates this entirely by capturing data directly in the field and synchronizing it with the central system immediately. In the domain of document management, the DMS enables instant retrieval of any document by protocol number, date, or sender, replacing a process that previously required physical archive searches lasting from several minutes to several hours. A particularly significant operational improvement relates to the invoicing process: previously, dispatch notes for timber assortments were handwritten in the field, manually re-entered into the system the following day, and only then could an invoice be generated — resulting in a minimum 24-hour delay in the billing cycle. The new system generates and delivers invoices to buyers automatically upon mobile application synchronization, eliminating this delay entirely and contributing to faster revenue collection. Finally, whereas year-over-year data comparisons previously required manual extraction from separate annual databases, the integrated system now supports cumulative multi-year analysis directly through the BI component.

In addition to the expected positive effects, the implementation of the integrated information system has also been accompanied by certain challenges. The most significant challenges relate to data migration, standardization of coding systems, and adaptation of employees to a new way of working. Particularly complex is the process of integrating spatial and business data, as well as the need to harmonize different work practices across organizational units [8]–[10].

Viewed as a whole, the implementation of the integrated information system has a potentially transformational impact on the enterprise's operations. By integrating operational, spatial, and analytical components, a unified information framework is being established that enables more efficient resource management, increases transparency, and provides a better basis for strategic and operational decision-making, especially after the system begins to operate at full capacity.

LIMITATIONS OF THE RESEARCH

The integrated information system analyzed in this paper is still in the final phase of implementation, which represents one of the main limitations of the research. As a result, certain effects described in the paper are based on preliminary operational observations and expected long-term benefits rather than fully measurable empirical indicators.

In addition, some organizational units are still undergoing the process of adaptation to standardized workflows and centralized data management. Therefore, a complete quantitative evaluation of all system effects will be possible only after full operational stabilization and long-term use of all implemented functionalities.

Future research should include detailed quantitative analysis of operational indicators, comparative efficiency measurements, and long-term evaluation of the impact of integrated information systems on forestry management processes.

CONCLUSION

The results of the analysis indicate that the implementation of the integrated information system in the Public Forestry Enterprise "Forests of the Republic of Srpska" represents a significant step toward the modernization and digital transformation of business operations. Since the system is still in the final phase of implementation, it is possible to speak primarily of the establishment of a unified information framework that should integrate business, spatial, and documentary data, while the full effects are expected after its complete functional completion.

One of the key expected effects of system implementation is reflected in the increase in operational efficiency, through process automation and the reduction of the need for manual data entry. The introduction of mobile applications further improves the accuracy and timeliness of data, enabling their collection directly at the point of origin.

At the organizational level, the system contributes to the unification of business processes and the increase in employee responsibility, while at the management level it should enable a higher degree of control and transparency. A particularly important aspect is that the system enables more complete monitoring of timber assortments through all phases of the process, which contributes to more efficient su-

pervision and more rational use of resources.

The development of the BI component of the system enables the improvement of analytical capacities through the integration of data from ERP and GIS systems, as well as their processing through different analytical models. In this way, data-driven decision-making should be enabled, which represents a key element of modern organizational management [2], [3], [12].

Although the implementation of the system has been accompanied by significant challenges, particularly in the areas of data migration and standardization, as well as the adaptation of employees to a new way of working, the results achieved so far and the functionalities established indicate the justification of investing in the integrated information system and its potential benefits.

The further development of the system can be directed toward the completion of all planned functionalities, the improvement of analytical capabilities, greater integration with advanced technologies, including artificial intelligence, as well as the continuous improvement of data quality and business processes. This would further increase management efficiency and enable the development of a modern model of digital resource management.

References

[1] T. H. Davenport, "Putting the Enterprise into the Enterprise System," *Harvard Business Review*, vol. 76, no. 4, pp. 121–131, 1998.

[2] W. H. Inmon, *Building the Data Warehouse*, 4th ed. Wiley, 2005.

[3] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Wiley, 2013.

[4] K. C. Laudon and J. P. Laudon, *Management Information Systems: Managing the Digital Firm*, 16th ed. Pearson, 2020.

[5] P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, *Geographic Information Systems and Science*, 4th ed. Wiley, 2015.

[6] I. Mergel, N. Edelmann, and N. Haug, "Defining Digital Transformation: Results from Expert Interviews," Government Information Quarterly, vol. 36, no. 4, p. 101385, 2019.

[7] E. Monk and B. Wagner, *Concepts in Enterprise Resource Planning*, 4th ed. Cengage Learning, 2012.

[8] Public Forestry Enterprise "Forests of the Republic of Srpska", "Project Manual for the Integrated Information System", internal project documentation, 2020.

[9] Public Forestry Enterprise "Forests of the Republic of Srpska", "Conceptual Design of the Integrated Information System", internal project documentation, 2021.

[10] Public Forestry Enterprise "Forests of the Republic of Srpska", "User Manuals for SAP, GIS and DMS Modules", internal project documentation, 2022.

[11] A. Shollo and R. D. Galliers, "Towards an Understanding of the Role of Business Intelligence Systems in Organisational Knowing," Information Systems Journal, vol. 26, no. 4, pp. 339–367, 2016.

[12] E. Turban, R. Sharda, and D. Delen, *Business Intelligence and Analytics: Systems for Decision Support*, 10th ed. Pearson, 2014.

[13] R. Venanzi, F. Latterini, V. Civitarese, and R. Picchio, "Recent Applications of Smart Technologies for Monitoring the Sustainability of Forest Operations," Forests, vol. 14, no. 7, p. 1503, 2023.

Received: April 21, 2026
Accepted: May 2, 2026

About the authors

Saša Ljubojević is the Head of IT Department at the Public Forest Enterprise "Forests of Republic of Srpska" and a doctoral candidate in the field of Geographic Information Systems at Pan-European University Apeiron, Banja Luka. His research interests include the application of Artificial Intelligence and GIS technologies in forest fire risk assessment, disaster risk management, cybersecurity, and digital transformation in the public sector. He has experience in the implementation and management of integrated information systems, including GIS, ERP, and document management systems.

Branko Latinović was born on April 28, 1956 in Prijedor, Bosnia and Herzegovina. He graduated from the Faculty of Economics in Banja Luka in 1980. At the same faculty he enrolls a master's degree that ends in 1994, and in 1997 successfully defended his doctoral dissertation. He worked as a dean of the Dean of the Faculty of Information Technologies of Pan-European University "Apeiron" in Banja Luka.

For citation

Sasa Ljubojevic, Branko Latinovic, Analysis of the expected effects of implementing ERP, GIS and DMS systems in a public enterprise, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:31-39, (UDC: 004.655.3:004.652.4), (DOI: 10.7251/JIT2601031LJ), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

SENTIMENT ANALYSIS ON SOCIAL MEDIA CONTENT

Boris Borovčanin

Department of Information Technologies, Faculty of Engineering, Natural and Medical Sciences, International Burch University, Sarajevo, Bosnia and Herzegovina, boris.borovcanin@stu.ibu.edu.ba, ORCID ID: 0009-0002-7993-0544

Original scientific paper

https://doi.org/10.7251/JIT2601040B UDC: 004.738.5:[658.8:659.1

Abstract: Following research evaluated conventional machine learning and deep learning algorithms used for the purpose of binary text classification, in accordance with previous research demonstrating advantages in supervised learning models such as Naive Bayes, Logistic Regression, and LSTM networks. Models that were subject of implementation are: Logistic Regression, Naive Bayes, Support Vector Machine (SVM), Random Forest, and LSTM. Responses from nonprofit organizations have been cleaned, tokenized, and preprocessed implementing either TF-IDF vectorization or sequence trimming determined by the model that was chosen. The majority of the models were performed using 50,000 samples because of computational capacity limitations, whereas the LSTM was executed only with 5,000 samples. LinearSVC is implemented for the purpose of accelerating training of the SVM model, as well as Random Forest parameters optimization for algorithmic efficiency. On the other hand the LSTM model provided an embedding component and a single LSTM unit for maintaining the sequence information. The performance of the models was evaluated according to the accuracy, precision, recall, and F1 score metrics. The findings are indicating that fundamental models perform effectively and consistently, however the LSTM model demands more computational capacity to provide context for classification.

Keywords: sentiment analysis, twitter data, machine learning, performance metrics

INTRODUCTION

Sentiment analysis or opinion mining is a widely studied application of Natural Language Processing (NLP) that involves identifying and classifying opinions contained in text data-specifically if the sentiment of a statement is positive, negative, or neutral. With user-generated content ballooning exponentially on Twitter, Facebook, Reddit, and Instagram, sentiment analysis has been rendered crucial in understanding public opinion, monitoring brand reputation, detecting misinformation, and even predicting political leanings.

In the age of the internet, people tend to share opinions online. Those unstructured bits of data have plenty of information but are hard to analyze at large manually. The overall interest in sentiment analysis is due to its applications across fields—businesses use it to improve user experience, governments monitor it for indicators of unrest, and researchers use it for studies of mental illness or consumer behavior.

The problem solved by this work is the analysis of social media sentiment data through the automatic classification of social media postings (such as tweets) to identify whether they have a positive or negative sentiment. In this research, posting classification used five different machine learning models to classify postings into binary (two classes) or multiclass (three classes) classifications. For this purpose, I have used the Sentiment140: Twitter Sentiment Analysis Dataset **[1]**, which contains 1.6 million annotated tweets labeled for sentiment classification.

The Sentiment140 dataset contains tweets assigned polarity sentiment classifications. An example of this is the positive tweet "Lyx is cool" that was created by user robotickilldozr on May 16, 2009, and is given a sentiment score of 4 (positive). The purpose of this example is to show how the dataset encodes the public's expression of feelings toward something into values that can be used for analyzing sentiment.

RELATED WORKS ON SENTIMENT ANALYSIS

Since there is a lot of user-generated content for expressing the general attitude of society at large it is crucial to highlight that social media sentiment anal-

ysis and statistics, especially data related to Twitter posts, have been a primary focus of interest in natural language processing (NLP). It is important to consider different strategies that have been utilized with a wide range of models including rule-based and conventional machine learning models on the one side, as well as transformer and deep learning models on the other. Although preliminary studies were focused on techniques for remote monitoring in large-scale data processing, research presented in **[1]**, it is important to emphasize that recent initiatives have included additional domain features and neural architectures to enhance sentiment classification efficiency. Even though significant advances have been made, concerns including flexibility of the domain, as well as the informal text, and absence of suitable labeled data in languages with limited resources remain in the focus of the corresponding studies.

Observing the approach established in **[1]** the foundation utilizing Sentiment140 data by leveraging emoticons as distant supervision labels (":\)" and ":\(") allowing scalable gathering of sentiment-labeled tweets. Corresponding study utilizes the same dataset as utilized in our project. Their methodology encouraged most modern social media sentiment processing pipelines.

Research presented in **[2]** talks about deep learning models (CNN, RNN, LSTM) used in sentiment analysis and highlights their advantages over traditional ML methods. This survey is relevant to the topic of my research since it provides theoretical grounding for your use of LSTM and justifies your choice of deep learning for modeling sentiment on unstructured text.

The main idea of work proposed in **[3]** paper surveys sentiment analysis techniques over Twitter data and suggests challenges such as short text, sarcasm, and colloquialisms. Paper is relevant to this topic since it provides assistance to establish the unique nature of Twitter sentiment analysis and helps the need to preprocess noisy tweets before classification.

When it comes to the idea introduced in **[4]** article provides a clear explanation of sentiment analysis algorithms (lexicon-based, ML-based) and areas of application like politics, business, and healthcare. Considering that article helps in defining a project's social impact and positions your model in the broader scope of application.

Although CNN-focused, the work discussed in **[5]** introduced new methods for using deep learning for text classification that set the stage for many subsequent sentiment models. Demonstrates how deep structures like CNNs (and later, LSTMs) can achieve better results than classic methods - comparative justification, relating it to the topic of my research.

Dataset preparation

In the manner of this study in the foreground is a comparison of five supervised learning models for binary text classification, including Logistic Regression, Naive Bayes, Support Vector Machine (SVM), Random Forest, and Long Short-Term Memory (LSTM) networks. Following that the pipeline began with a preprocessing stage which included tokenization, contaminating the text, and encoding the labels. Considering the traditional machine learning model, the feature extraction process was performed using TF-IDF vectorization.

In light of computer limitations and to accelerate selected models, only 50,000 samples were used to train all the models, except the LSTM model which was limited to 5,000 samples. LSTM stands out from other models because it could not handle textual data as a matrix of fragmented features. Keras's Tokenizer and pad_sequences were utilized in order to convert the text into sequences of augmented integers to be used by LSTM since it takes sequential inputs. Taking it into consideration all models were trained and tested according to different performance measures including: accuracy, precision, recall, and F1 score.

Data Preprocessing

Dataset has been broken down into training and testing sets for the purpose of performance evaluation on data that is not part of the sample. Each component of textual data encountered basic the preparation process, including:

- Stopword removal and punctuation
- Lowercase all tokens
- Stemming or lemmatization

The dataset was already in binary classification format, consequently there was no need for one-hot encoding or manual label encoding.

Text Vectorization

Three vectorization approaches have been addressed:

- TF-IDF (Term Frequency–Inverse Document Frequency): The majority of models supported this approach, which involved transforming text to a minimal matrix of normalized word frequencies.
- CountVectorizer: Converts a collection of text documents into a token counts matrix.
- HashingVectorizer: This method represents a more efficient, as well as memory-friendly method of vectorizing text that implements the hashing trick.

Considering how it impacts the memory and performance, TF-IDF has been chosen with `max_features=1000`, as the default for all the classical models.

RESEARCH DESIGN

Exploratory Data Analysis (EDA)

In advance of doing the analysis, the initial step was Exploratory Data Analysis (EDA), for the purpose of better dataset comprehension. This has been accomplished by implementing text lengths analysis, while discovering the most frequent words, and visualizing the class distributions. Additionally, the preprocessing stage included a wide range of tasks, such as identifying missing values, stopwords, as well as text normalization that has been applied in the same time frame. The most commonly used phrases in each class are presented as well using semantic clouds and bar charts, which enhanced the feature selection and allowed observations, which could be useful in understanding the data itself. Furthermore, EDA has provided assistance for easier decision making when it comes to tokenization and vectorizer implementation used afterwards in the stage related to development of the model.

Model Development

Five different algorithms were trained and implemented for binary classification of text. Each algorithm was chosen based on its applicability to the task, scalability, and ability to work with sparse or sequential data.

Logistic Regression was used as a baseline, linear classifier algorithm due to its speed, simplicity, and interpretability (Fig. 1). It had been trained with TF-IDF vectorizer features in order to preserve importance of the term while still being computationally efficient. The model would choose features that had a sufficient amount of importance while still disregarding terms of lower importance. Training the Logistic Regression model directly to a limited, high-dimensional matrix could result in relatively inefficient convergence. Furthermore, the maximum number of iterations completed was increased to ensure convergence while training.

```
# Logistic Regression
lr = LogisticRegression(max_iter=1000)
lr.fit(X_train_vec, y_train)
metrics['Logistic Regression'] = evaluate_
model("Logistic Regression", y_test,
lr.predict(X_test_vec))
```

Figure 1. *Logistic regression*

Naive Bayes algorithm (MultinomialNB specifically) was chosen for its stochastic structure, while resulting in outstanding performance with previous research in binary classifications of text (Fig. 2). Naive Bayes assumes independence among features, which is a potential simplification of reality, but effectively works in classification tasks where the features can be substantially represented in their states, either with word frequencies or terms method of TF-IDF. Since Naive Bayes is a lightweight architecture and performs very well with sparse representations of data, it was expected to be one of the most efficient models in the experiment.

```
# Naive Bayes
nb = MultinomialNB()
nb.fit(X_train_vec, y_train)
metrics['Naive Bayes'] = evaluate_
model("Naive Bayes", y_test, nb.predict(X_
test_vec))
```

Figure 2. *Naïve Bayes*

Support Vector Machine (SVM) was implemented using the SVC() classifier at first, but because it had slow training times and convergence issues with large, high-dimensional datasets, it was topologically replaced with LinearSVC. The linear version, while different theoretically, is more suitable for sparse and large-scale text classification tasks by providing more

scalability and implementation alternatives without significantly compromising robustness, such as accuracy. (Fig. 3)

```
# SVM
svm = LinearSVC(max_iter=1000)
svm.fit(X_train_vec, y_train)
metrics['SVM'] = evaluate_model("Linear SVM", y_test, svm.predict(X_test_vec))
```

Figure 3. *SVM*

Random Forest was also implemented because it is an ensemble learning method, therefore combining multiple decision trees to improve performance and robustness. There were considerable challenges with resources and an extensive training period since the problem was how it was used in the implementation. Following that, the model was tuned (n_estimators=10, max_depth=5, n_jobs=-1) to limit the tree number, tree depth, and enable parallel computations just to train a model efficiently while still capturing important feature interactions. (Fig. 4)

```
# Random Forest
rf = RandomForestClassifier(n_estimators=10, max_depth=10, random_state=42)
rf.fit(X_train_vec, y_train)
metrics['Random Forest'] = evaluate_model("Random Forest", y_test, rf.predict(X_test_vec))
```

Figure 4 *Random forest*

Long Short Term Memory (LSTM) neural network was used to explore deep learning-based sequential modeling. The model was built using TensorFlow/Keras. The text input was tokenized and represented as augmented integer sequences, which were converted to be input to the embedding layer and LSTM. In general, the architecture of the LSTM is done in a structure of first an embedding layer that takes tokens and converts them into dense vectors, followed by a single LSTM layer that captures temporal dependencies in sequences of words. Additionally, two layers were built including the regularization dropout layer as well as the final dense layer for binary classification with a sigmoid activation function. Because of limited resources, the maximum size of the training set was limited to 5000 text samples, one training epoch, a batch size of 64, embedding dimension of 32. Despite the reduced efficiency of the LSTM algorithm, the LSTM provided clear evidence referring to potential and capabilities of deep learning models in the field of text classification. (Fig.5)

```
# LSTM
sample_size = 5000
X_train_sample = X_train[:sample_size]
y_train_sample = y_train[:sample_size]
```

Figure 5 *LSTM*

Evaluation

The corresponding models were evaluated according to four performance metrics including accuracy, precision, recall, and F1 score. Accuracy is an overall metric of how frequently predictions were correct, while precision is the proportion of true positives based on positive predictions. On the other hand recall represents the proportion of true positives found from all true positives, and F1 score is the representation of balance between precision and recall. The metrics were calculated using scikit-learn's evaluation measuring functions. Afterwards, all the results gathered were stored as a dictionary, and the results were able to be visualized in a bar graph, making it relatively straightforward to observe all models on the same chart. The evaluation pipeline ensured consistency and equity between models.

The data set used in the current study consisted of 1,583,691 Twitter tweets, about evenly split between sentiment classes (50.1% positive and 49.9% negative), and hence well-suited for objective model testing. A comprehensive data quality check revealed no missing values, an extremely low rate of duplication of about 0.14%, and significant variability in the lengths of tweets (mean = 74.47 characters, std = 36.2), consistent with previous reports that Twitter data is short-lengthed and noisy **[3]**. Preprocessing involved token cleaning and noise reduction, as used in earlier studies to improve performance for social media text in an informal setting **[4]**.

Five classifiers were trained and tested: Logistic Regression, Naive Bayes, Linear Support Vector Machine (SVM), Random Forest, and a deep learning

LSTM-based model. Their performance, in terms of accuracy, precision, recall, and F1 score.(Table 1)

Table 1. *Performance Comparison of Machine Learning Models on Text Classification Task*

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.7586	0.7484	0.7833	0.7654
Naive Bayes	0.7453	0.7436	0.7531	0.7483
Linear SVM	0.7579	0.7463	0.7854	0.7653
Random Forest	0.6773	0.6353	0.8411	0.7238
LSTM	0.6099	0.5826	0.7905	0.6708

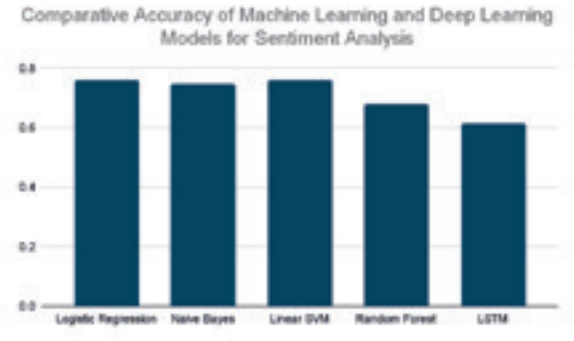


Figure 6: *Comparative Accuracy of Machine Learning and Deep Learning Models for Sentiment Analysis*

The accuracy examination demonstrates that simple machine learning models can be significantly more accurate than deep learning models under controlled conditions. Logistic Regression (75.86%) and Linear SVM (75.79%) are reaching around 75.8% accuracy rate (75.86% for Logistic Regression and 75.79% for Linear SVM) **as shown in Figure 6,** while supporting the effectiveness of the corresponding method when it comes to handling high-dimensional, insufficient data such as tweet texts. This finding is in line with that of **[3],** who highlighted that linear classifier, and SVMs in particular, are optimal for short-text sentiment analysis since they manage to monitor optimum separating hyperplanes in sparse feature spaces. In addition, accuracy results are supported by **[1]** as one of the previous studies as well, while indicating that standard models perform well on Twitter data. It is important to emphasize that Naive Bayes was close behind these results with accuracy equal to 74.53% while Random Forest did well, with 67.73% there was a significant decline in performance. On the other hand, achieving relatively low accuracy of 0.6099 **as shown in Figure 6.** while demonstrating a wide range of limitations when it comes to the interpretation of the LSTM model to the brief texts with minimal context without depending on more complex architectures or pretrained incorporated data. This opposed the estimated outcome of the deep learning model achievements from **[2]**, expanding on the statement that neural networks most commonly require considerable tuning alongside with other semantic resources before overtaking simpler models in the tasks related to the classification of the texts.

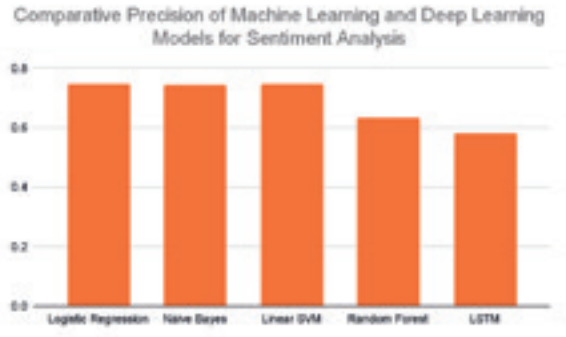


Figure 7: *Comparative Precision of Machine Learning and Deep Learning Models for Sentiment Analysis*

Precision results highlight that the traditional methods are effective for Twitter sentiment analysis. Logistic Regression has the highest precision at 74.8%, followed closely by Linear SVM at 74.6% has a marginally lower precision rate **shown in Figure 7.** within the same variation range, demonstrating identical effectiveness. Taking it into account there is an additional reinforcement of SVM model determined chronology of performance in corresponding NLP tasks, especially in cases when comprehension takes a secondary function prioritizing performance. However, Naive Bayes had 74.4% when it comes to precision rate, as we can see from **Figure 7.** reflecting marginal variation when it comes to overall performance of models which were subject of examination. Following that, results of precision metric are consistent with previous work done by **[1]**, where the development of corresponding model is also reported to be a reliable alternative on emoticon-labeled data similar to the Sentiment140 dataset. It is important to emphasize that Random Forest has reached 63.5%, while LSTM achieved 58.3% precision rate as **shown in Figure 7.** as well, indicating far lower results in the field of positive sentiment identification, revealing

several types of constraints in terms of implementing the LSTM model to brief texts, as mentioned earlier.

One of the reasons for the performance inconsistencies could be that traditional models, such as TF-IDF, are able to perform more effectively with incomplete information. However, the LSTM method developed in the current research involved hyperparameter adjustment, contextual embeddings, and large datasets, while weak performance had been expected given the absence of pretrained embeddings, along with concurrently fragmented and misleading format of tweets.

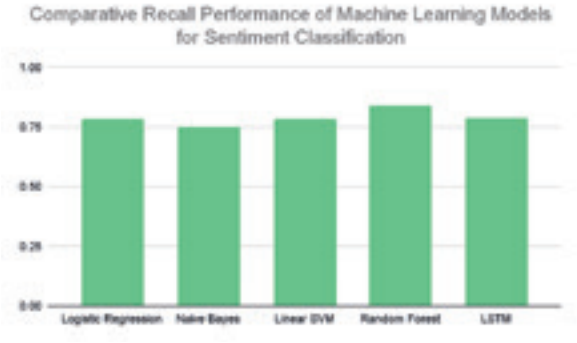


Figure 8: *Comparative Recall Performance of Machine Learning Models for Sentiment Classification*

Recall performance provides insight into the different strengths of the models in identifying true positive sentiments. Random Forest had the highest recall performance with 84.1%, while LSTM achieved 79.1% and Linear SVM with a very close percentage of 78.5% **as is represented in Figure 8.** Logistic Regression has reached 78.3%, although Naive Bayes has the lowest recall percentage among recall performance results which equals to 75.3% **as shown in Figure 8 as well.** The difference between the samples tends to be related to the capacity of various models to identify more positive sentiment instances (in this case positive sentiment), with the more responsive deep learning and ensemble frameworks achieving more efficient performance then more resistant traditional methods. Taking into consideration ensemble structure and capacity to compromise precision in favor of identifying more positives, Random Forest most certainly achieves high levels of recall. Indicating the tendency to overpredict positive sentiment while building up a rate of false positives. Despite the accomplishments of models based on trees on the corresponding textual classification tasks, responsiveness to noise on high-dimensional data was certainly the reason for these results as well, which is in line with **[4]**.

Corresponding results demonstrate a wide range of limitations, when it comes to implementing the LSTM model to brief phrases with minimum context, while it is not depending on more advanced architectures or preconditioned data. This opposed the estimated outcome of the deep learning model achievements from **[2]**, expanding on the statement that neural networks most commonly require considerable tuning alongside with other semantic resources before overtaking simpler models in the tasks related to the classification of the texts. In addition the results of this analysis also align with the findings of **[1]**, since their learning was supported through using the Sentiment140 dataset, which had emoticon labelled tweets, and was based on an extensive repository of labelled data that provided communicative labels for comprehension, which implements recall-based learning.

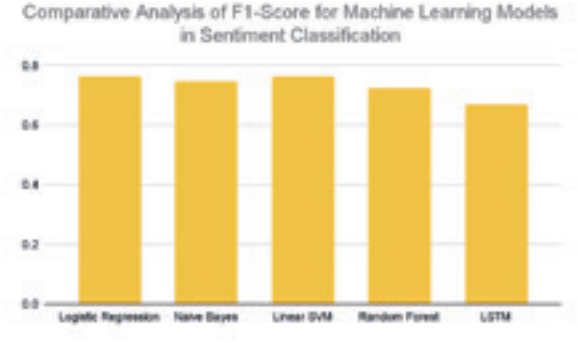


Figure 9: *Comparative Analysis of F1-Score for Machine Learning Models in Sentiment Classification*

Furthermore, in the context of Twitter sentiment analysis, the F1 evaluations indicate the manner in which conventional models perform while maintaining an even balance between precision and recall. The most outstanding values of F1 Score were reported by Logistic Regression and Linear SVM, achieving 76.5% regarding both two models respectively **such as Figure 9. represents**. On the other hand, Naive Bayes followed closely behind with a recall percentage of 74.8% **as shown in Figure 4.** The convenience of Naive Bayes is an important benchmark because of its responsiveness and decent performance, particularly in systems that are running in real time and when computational resources are limited as it is explained by **[1]**. Each of these models discussed above

achieved a balance between evaluating real positives and decreasing false positive results, which is crucial when handling simple terms, as well as fragmented text such as tweets. Random Forest has achieved 72.4%, while LSTM generated a lower percentage of 67.1% when it comes to F1 Score **as it is visible from Figure 9.**, which additionally reflects the reliability of their results, to indicate nothing of the outstanding recall they reported in certain instances. Traditional models perform better in terms of F1 performance, which is in accordance with **[1]**, while indicating that various methods including Logistic Regression consistently performed effectively while trained on the **Sentiment140 dataset**. The results demonstrate that traditional machine learning models still have a competitive advantage, especially in contexts with well-preprocessed and balanced data, as well as reduced tweet lengths. This also corresponds with **[3]** and **[4]**, addressing that Logistic Regression and SVM will outperform deep models in short-text conditions without significant semantic complexity.

The main reason for the corresponding performance was the capacity of conventional models to manage fragmented text representation (TF-IDF). Taking it into consideration Logistic Regression could be fractionally implemented on social media information, since it can be modified to informal language and have reduced capacity in order to prevent overfitting. The limited dataset size could have been responsible for low performance of the LSTM in corresponding research, while being determined in the field of consistent data that reflects social media content and insufficient contextual embeddings. On top of that, low performance results achieved by the LSTM model regardless its ability to process sequential context are in the line with **[1]** and **[5]**, while domain-sensitive limitations, such as acronyms, informal language, and sarcasm, are required to be responded to and solved more effectively, which involves hybrid or attention models.

Future research should be focused on combining contextualized embeddings (BERT, RoBERTa) and investigation of the ensemble methods that provide balance precision and recall. Sentiment misclassifications could also be assisted with comprehension evaluations, uncovering feature importance or using attention mechanisms to detect limited textual indicators.

The corresponding research is important for sentiment analysis because it established a comprehensive benchmark comparing classical machine learning methods - Logistic Regression, Naive Bayes, SVM, and Random Forest - with a deep learning model, LSTM, and provided relevant information related to the differences between accuracy, scalability, and computational effectiveness through an objective evaluation of the five methodologies.

This study operates under real-world conditions, including limited hardware and reduced dataset size (limiting the LSTM training, for instance, to 5000 samples) in order to to make the results more relevant for working in real-time, and potentially useful in implementing related research, opposing several studies that perform in ideal environments with a wide range of resources. Additionally, through evaluating all models using accuracy, precision, recall, and F1 Score metrics, the research presented a balanced, and comprehensive assessment of the models in terms of the advantages and limitations of each model, especially around imbalanced or complex sentiment datasets.

This research presents a responsive and uniform pipeline with an accurate and consistent evaluation of a range of models: whether they be commonly used linear classifiers, decision tree approaches, or recent deep learning architectures such as LSTM. The use of common processing and evaluation methods enhances reproducibility across models, providing an important baseline for future work and extensibility. Furthermore, this work includes an effective LSTM implementation which demonstrates how significant performance can be achieved even with limited computational capacity - an important contradiction that is most frequently ignored in sentiment analysis work. Moreover, this project has a balanced approach towards performance and interpretability. Even though deep learning models provide more contextual information, it is important to emphasize efficiency and transparency for the less complex models, such as Logistic Regression. This work is useful for implementation across different fields where comprehensibility is nevertheless taken into account, even if it is not as critical as accuracy.

CONCLUSION

This study investigated sentiment analysis on social media data, while evaluating the performance of five machine learning models (Logistic Regression, Naive Bayes, Linear SVM, Random Forest, and LSTM),

trained against the Sentiment140 dataset and vectorized using TF-IDF. Research found that out of the five models evaluated, traditional classifiers (Logistic Regression, Linear SVM) produced the most balanced results with F1-scores of 76.5% and 76.5%, respectively, demonstrating that the classical models are more relevant when working with highly limited and dimensional text data that are typical for Twitter. Analysis of precision scores further supported that classical methods were more efficient, while recall scores indicated Random Forest and LSTM performed well when it comes to identification of a wider range related to positive sentiments, even though compromising the precision.

Corresponding research supports findings of **[1]** indicating that distant supervision model and development of a scalable methodology for classifying sentiment while incorporating data labeled using emoticons. Research also supports the outcomes of **[4]** that traditional machine learning models are effective methods for sentiment analysis datasets, particularly because traditional approaches are simple to interpret and require low resources, limiting input and complexity. Furthermore, the restrictions observed in current research with the deep-learning models (LSTM), as discussed by **[5]** , indicated that deep learning models need to be effectively optimized with large, appropriately interpreted datasets, and context sensitive embeddings to successfully outperform their alternatives.

Future research could incorporate contextual word embeddings (eBERT or GloVe) and transformer-based architectures to improve semantic understanding for tweets, preprocessing specific to a domain, hyperparameter tuning, and ensemble methods. There is also potential for extending the application to multilingual sentiment datasets, real-time analysis scenarios, and other useful applications.

The approach used in this paper creates a modular, extensible framework for exploring different approaches to text classification. By allowing for a direct comparison of traditional machine learning techniques and modern deep learning techniques, this approach can suggest practical findings regarding strengths and weaknesses among both categories of techniques. In this study, we were able to optimize both the models and the sampling strategy to make the study feasible, but one advantageous next step would be to increase the dataset size, play with different neural models such as transformer models, along with hyperparameter tuning or cross-validation, to achieve more generalization. This study serves as a basis for future work to create scalable, well-performing text classification systems.

References

[1] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," Stanford University, 2009. [Online]. Available: https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf

[2] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1253, 2018, doi: 10.1002/widm.1253.

[3] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–41, 2016, doi: 10.1145/2938640.

[4] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[5] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014. [Online]. Available: https://arxiv.org/abs/1408.5882

Received: April 28, 2026
Accepted: May 4, 2026

About the authors

Boris Borovčanin is an engineering graduate and MSc student at International Burch University in Sarajevo, Bosnia and Herzegovina. He has academic and practical experience in data analysis and information systems, including work at Raiffeisen Bank dd BIH. His research interests include data science, network security, and machine learning applications.

For citation

Boris Borovčanin, Sentiment Analysis on Social Media Content, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:40-47, (UDC: 004.738.5:[658.8:659.1), (DOI: 10.7251/JIT2601040B), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

INTENT-DRIVEN PAYMENTS: A PROPOSED FRAMEWORK FOR USING LARGE LANGUAGE MODELS TO TRANSLATE NATURAL LANGUAGE INTO STRUCTURED PAYMENT INSTRUCTIONS

Vijay Narayanan
Independent Researcher, USA, vijaynarayanan.blr@gmail.com

Preliminary communication

https://doi.org/10.7251/JIT2601048N UDC: 811.111'373.612.2:7.038.53

Abstract: The modern payment infrastructure assumes well-defined inputs by users, involving specifying parameters such as the amount of the transaction, payee identity, and timing of payment execution. This mode of operation causes procedural overhead and constrains the possibility of abstraction, especially with the ongoing evolution of financial products towards greater user-friendliness. This article presents a conceptual framework for intent-driven payments, in which users formulate their financial intentions in natural language and large language models (LLMs) are used to generate corresponding structured payment workflows. While the term has been used informally in prior industry commentary, this work offers a structured architectural treatment of the paradigm. The framework described here is proposed and has not yet been empirically implemented. The intent-to-payment system presented is built as a multistep pipeline, incorporating such stages as intent extraction, entity recognition, constraint validation, and orchestration. One significant novelty of this work concerns the development of the financial intent compiler, which is designed to enforce that the outputs generated by the system are deterministic, transparent, and aligned with applicable regulatory constraints. This article touches upon a number of topics relating to the design of systems, such as latency-related problems, handling ambiguity, considerations of security, and verifications of computations made by people.

Keywords: Intent-Driven Payments; Large Language Models; Natural Language Processing; Financial Automation; Semantic Parsing

INTRODUCTION

Modern payment systems are designed mostly on the basis of an interface approach. For a payment to be made, a user has to submit certain information in an orderly manner, which is required for starting a transaction. Although the approach is very solid and deterministic, it does not match the way humans describe financial purposes. To convert their thought about making a particular payment into filling out form fields is quite problematic.

Advances in large language models (LLMs) have substantially expanded the frontier of natural language processing, enabling systems to produce outputs that approximate inference over ambiguous, underspecified, or contextually rich inputs with considerable surface accuracy [1, 2]. We note that, although such outputs may resemble reasoning, LLMs operate fundamentally as statistical predictors over token sequences rather than as deliberative reasoning agents; any apparent reasoning is an emergent property of pattern completion rather than logical inference [1].

These capabilities present a compelling opportunity to reconceptualize the human-payment system interface: rather than constraining users to the vocabulary of software forms, a language-centric model would allow users to express financial goals in natural terms, with the system assuming responsibility for resolving intent into actionable, compliant instructions.

This article proposes a conceptual framework for intent-driven payments. For the purposes of this work, we define intent-driven payments as a payment interaction paradigm in which the user expresses a financial goal in unconstrained natural language, and a system layer is responsible for parsing that ex-

pression, resolving its referents, validating it against schema and regulatory constraints, and translating it into a structured instruction that an existing payment rail can execute. The term has appeared in prior industry discussion, but it has not been formalized as a system-design construct in the academic literature; this article offers such a formalization. This article is intended as a conceptual framework and research agenda rather than an empirical systems paper. The system described here is a proposed framework and has not yet been implemented or empirically evaluated; no prototype, simulation, or benchmark study is reported in this article, and the contribution is architectural and conceptual. The empirical study of an implementation against the evaluation framework presented in Section 7 — including measured tradeoffs and quantitative performance results — is left to future work. Accordingly, the claims made in this article are limited to architectural feasibility and research framing; the detailed limitations of this position are enumerated in Section 8.1.

The central thesis is that LLMs can serve as the interpretive layer between unstructured natural language input and the structured, schema-governed workflows required by payment infrastructure. By abstracting the mechanics of transaction specification, such a system lowers the barrier to financial participation and accommodates a broader spectrum of user literacy and interaction styles. The approach is positioned not as a replacement for deterministic execution infrastructure but as an intelligent front-end layer that mediates between human intent and machine-executable instruction. Concretely, this article makes three contributions: (i) it defines intent-driven payments as a constrained natural-language-to-payment-instruction mapping problem; (ii) it proposes a layered architecture that separates probabilistic intent interpretation from deterministic validation and execution; and (iii) it defines an evaluation framework for future empirical assessment of accuracy, latency, compliance, ambiguity handling, and adversarial robustness.

The remainder of this article is organized as follows. Section 2 reviews related work in semantic parsing and LLM-based dialogue systems as they pertain to financial applications. Section 3 formalizes the problem as a constrained mapping task. Section 4 describes the proposed multi-layered system architecture. Section 5 details the learning framework and hybrid design approach. Section 6 examines key system design considerations. Section 7 presents the evaluation framework. Section 8 discusses implications and limitations. Sections 9 and 10 provide details on future directions and conclusions, respectively.

Background and Related Work

This section reviews the three bodies of work on which the proposed framework draws: advances in large language models and their capabilities for structured output generation; the field of semantic parsing, which provides the theoretical basis for mapping natural language to formal representations; and the technical landscape of modern payment infrastructure, against which the gap addressed by this article is defined.

Large Language Models and Natural Language Understanding

The advent of transformers has completely changed the field of natural language processing. Transformers trained on massive amounts of data possess significant in-context learning skills, which allows them to perform complicated tasks in a few shots without any fine-tuning for task specificity [1]. Tasks ranging from parsing to multi-step inference have been achieved by LLMs, indicating that they might be capable of performing semantic tasks like intent analysis in constrained domains such as finance. Importantly, the ability to generalize in few shots that has been demonstrated for foundation models [1] suggests that LLMs may be capable of adapting to payment intent comprehension without requiring domain-specific fine-tuned corpora.

Further advances have expanded the range of possibilities even more, allowing for increased capability to follow instructions, generate structured outputs, and perform multistep dialogues [2]. Maintaining consistent context during a conversation is especially important for the field of finance, where the user's intention may require information from previous interaction or account status.

Semantic Parsing and Structured Query Generation

Semantic parsing, which involves translating natural language into formal meanings, serves as the

theoretical basis for building intent-driven payment systems [7]. Prior work in task-oriented dialogue systems has established the feasibility of converting user utterances into structured slot-fill representations, particularly in domains such as flight booking, restaurant reservation, and database querying. The Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics document significant advances in this area, including neural approaches to compositional generalization that are directly applicable to financial instruction formation [7].

Nonetheless, there exist certain qualitative differences in the constraints imposed on finance-specific areas that do not occur in traditional semantic parsing testbeds. The directives for payments should not only be semantically correct and reflect what users wish to do, but they should also be syntactically and compliance-wise correct in terms of their relationship to schemas and regulatory and risk frameworks, respectively.

Recent work on schema-constrained generation and function calling has shown that LLMs can emit structured, API-compatible outputs when supplied with explicit schemas and representative examples, leveraging the few-shot learning capabilities established for foundation models [1, 2]. Related work in task-oriented dialogue and slot filling has long demonstrated the feasibility of converting unstructured or semi-structured natural language inputs into bounded record representations in domains such as flight booking, restaurant reservation, and database querying [7]. Adjacent applications in document information extraction (for example, parsing invoices and receipts into structured fields) and web-form completion further indicate that schema-bounded extraction from natural language is technically feasible across a range of production contexts.

What distinguishes the payment setting from these adjacent applications is the combination of three constraints simultaneously: irrevocability of execution, multi-jurisdictional regulatory binding, and adversarial exposure [4, 5, 6]. To the authors' knowledge, no published work has examined the integration of these three constraints into a single architecture targeting fast-payment rails, which is the gap this article addresses.

Payment Infrastructure and the Gap Addressed

Modern fast payment systems, as surveyed in reports by the Bank for International Settlements, are characterized by near-real-time settlement, high availability, and stringent security requirements [5]. The FedNow Service, as documented in its operating procedures, exemplifies an execution-layer infrastructure optimized for structured, validated transaction messages rather than natural language input [6]. These systems represent the execution substrate upon which an intent-driven layer must operate, and their design constraints — latency, schema compliance, and irrevocability — directly shape the requirements for any upstream intent processing architecture.

The gap between the expressive richness of natural language and the structural rigidity of payment rails constitutes the central engineering challenge this work addresses. Existing literature has not systematically examined how LLM-based semantic parsing can be integrated with payment orchestration systems in a manner that preserves the compliance and auditability properties expected of financial infrastructure. This article contributes a structured framework for such integration. Table 1, compiled by the authors from prior surveys of payment infrastructure and LLM application architectures [4, 5, 6], summarizes the principal differences between rule-based and intent-driven payment system architectures along eight dimensions. (Table 1)

Table 1: *Comparative analysis of rule-based and intent-driven payment system architecture [4, 5, 6].*

Dimension	Rule-Based Payment Systems	Intent-Driven Payment Systems
Input Model	Structured, form-based user input	Natural language expressions of user intent
Interaction Abstraction	Low-user must specify all parameters	High - system resolves ambiguity and infers context
Flexibility	Rigid; breaks on unanticipated input patterns	Adaptive; handles varied phrasing and edge cases
Compliance Enforcement	Hard-coded rule sets	Dynamic validation layer with regulatory constraint checking
Auditability	Deterministic logs; straightforward traceability	Requires explicit financial intent compiler for auditability

Security Profile	Well-understood attack surface	Additional exposure to prompt injection and adversarial inputs
Latency	Low; optimized deterministic paths	Potentially higher; inference overhead requires optimization
User Experience	High friction; requires domain literacy	Low friction; aligns with natural cognitive patterns

Problem Formulation

Natural language parsing to generate payment instructions can be viewed as a problem of mapping from the unstructured input domain to the structured output domain. Let I denote the space of all possible natural language utterances pertaining to financial actions, and let O denote the space of valid payment instruction schemas recognized by the execution layer. The intent-driven payment system must learn a function $f: I \rightarrow O$ that is semantically faithful, schema-compliant, and constraint-satisfying.

This mapping is non-trivial for several reasons. The first challenge is that natural language inputs can be highly variable lexically and syntactically; the exact same financial intent may be conveyed using very different surface expressions, including everything from terse imperative sentences to complex conditional clauses. The second issue is that inputs may be under-specified, lacking information that is implied by the situation. Third, the output space is strictly constrained: an instruction that violates schema requirements or regulatory rules is not merely suboptimal but categorically invalid and must not be forwarded to the execution layer.

Ambiguity management constitutes a particularly important design dimension. In cases where a given input admits multiple plausible interpretations, for example, when a payee identifier is ambiguous or when a temporal expression is underspecified, the system must either resolve the ambiguity through contextual inference or surface the ambiguity to the user through a targeted clarification request. The latter represents a human-in-the-loop mechanism that trades latency for accuracy, and its calibration involves a design trade-off between automation rate and error rate that must be empirically tuned for the deployment context [8].

System Architecture

The proposed system is organized as a multi-layered architecture comprising three primary functional strata: an intent processing layer, a validation layer, and an execution layer. This stratification reflects both the logical sequence of operations required to transform natural language into a payment transaction and the distinct technical disciplines — language modeling, rule-based reasoning, and transactional engineering — that govern each stage. The overall flow is illustrated in Figure 1. A simplified pipeline-flow notation has been chosen here in preference to a strict UML activity diagram because the architecture at this conceptual level is a linear sequence of processing stages with a single feedback edge, and does not contain the branching decision nodes, parallel forks, or swimlanes that UML activity notation is designed to express. A streamlined notation therefore communicates the pipeline structure more directly to the intended readership of system architects and payments engineers; a UML activity diagram would become appropriate in future work that elaborates the decision logic of the validation layer, where multiple branching outcomes and parallel checks need to be represented explicitly. In the diagram, rounded rectangles denote the principal processing stages of the pipeline, with each stage's name shown in the upper portion of the box and its constituent techniques listed below; solid arrows indicate the forward control flow from the user input through to the confirmed transaction, and the dashed arrow on the left denotes the feedback loop through which post-execution outcomes inform model refinement at the intent-extraction stage.

The intent processing layer takes as input the user utterance received through the user interface, which it processes with the help of an LLM to convert it to an intermediate form. The output contains the action type, resolved entities such as the payee, amount, currency, and time, as well as conditions if any. An LLM follows constrained decoding during inference in order to ensure that the output adheres to a particular schema for the intent and does not violate downstream requirements [1, 2].

The validation layer takes as input the structured output from the intent processing phase. The validation layer can be seen as a financial intent compiler because it does not change the semantics of the instruction but simply verifies that all of its components fall within acceptable ranges and that it complies with the pertinent regulations and risk thresholds. Instructions that fail validation are either returned to

the user for clarification or rejected with an explanatory rationale.

The execution layer interfaces with the underlying payment infrastructure, such as fast payment rails, to complete the validated transaction [6]. This abstraction hides details of the payment system itself and provides a standardized interface to the validation layer irrespective of how the system actually settles its payments. Feedback from execution to the interpretation of intent improves the performance of the model for edge cases.

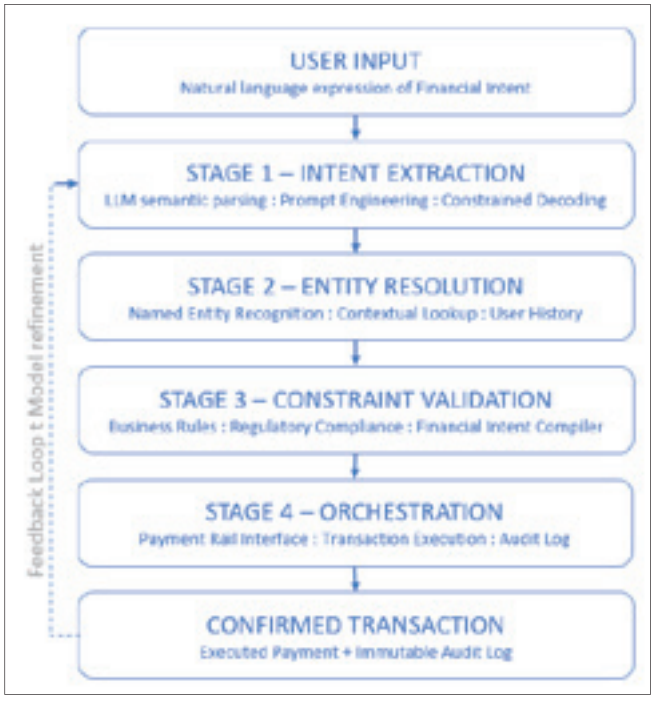


Figure 1. *Multi-stage pipeline architecture for intent-driven payments, showing the four processing stages, the user-input and confirmed-transaction endpoints, and the feedback loop from execution outcomes back to intent extraction. Diagram constructed by the authors; pipeline stages informed by the FedNow operating model [6].*

LEARNING FRAMEWORK

Scope of methodology. The methodological contribution of this article is architectural synthesis rather than experimental implementation. The section that follows specifies the learning framework, hybrid design, and decoding strategy that an implementation should adopt; it does not report training runs, hyperparameter searches, or measured performance figures. This scoping is deliberate and consistent with the conceptual-framework framing established in Section 1: the purpose of the article is to define a viable architecture and a research agenda, and the empirical realization of the architecture is left to future work as described in Section 8.1.

The learning model that forms the basis of the intent processing layer is a combination of two different models that leverage the strengths of generative learning model and rule-based learning model. Although generative learning models provide maximum flexibility, they do not by themselves enforce schema conformance or regulatory compliance. Meanwhile, rule-based models fail to manage the intricacies involved in dealing with natural language inputs. The hybrid model leverages the strengths of both learning models by allocating tasks to the appropriate component.

Prompt engineering plays a central role in guiding LLM behavior toward the desired output format. By carefully constructing system prompts that specify the output schema, enumerate constraint classes, and provide representative few-shot examples, system designers can substantially improve the reliability and precision of intent extraction without requiring fine-tuning [1, 2]. Where few-shot prompting proves insufficient for edge cases or domain-specific vocabulary, targeted fine-tuning on curated financial utterance datasets can further improve accuracy.

There are constrained decoding methods which serve as another method to enforce output structure during the generation stage. Unlike in the case of the model reproducing only trained patterns, with constrained decoding, the tokens available during the generation process will only be those that conform to the intended schema, thus ensuring that the output generated by the model remains syntactically correct even in situations where its semantic meaning is ambiguous. This is especially useful in a payments system, where failure to adhere to a schema has implications.

The contextual representations obtained from user activities give more representation material to the intent processor layer than just plain text. By encoding patterns of past financial behavior into the model's context, the system can resolve ambiguities that would be intractable from the raw utterance, thereby improving both accuracy and the naturalness of human-AI interaction [7, 8].

System Design Considerations

Translating the architecture described in earlier sections into a deployable system surfaces several engineering concerns that cut across the three layers of the pipeline. Four of these — latency, security, regulatory compliance, and human oversight — are particularly consequential for intent-driven payment systems and are examined in turn below.

Latency and Performance

Payment systems have extremely high latency constraints, where the user side has to see the response within seconds while settlement finality is determined based on the operating procedures of the payment network [6]. For meeting the aforementioned latency constraints, we need to efficiently handle the overhead of the LLM computations. Architectural strategies include model quantization, speculative decoding, caching of common intent patterns, and selective invocation of the LLM only for inputs that cannot be resolved by a lightweight rule-based prefilter. The design trade-off between model capability and inference latency is a central empirical concern in production deployments of this architecture.

Security and Adversarial Robustness

LLMs integrated into financial transactions present new forms of security threats that do not exist in rule-based systems. In particular, prompt injection attacks involve sending malicious inputs to the LLMs to force them to violate their instructions or perform unauthorized actions such as making payments [4]. Prompt injection is the number one issue on the OWASP Top 10 for LLM applications, and mitigating it requires architectural defenses including input sanitization, schema-based output validation, and mandatory human approval for high-value transactions [4]. However, defense-in-depth dictates that no architectural layer alone can provide absolute protection against attacks.

Regulatory Compliance and Explainability

There exist multi-layered regulations on financial systems, which include anti-money laundering measures, know-your-customer rules, and transaction reporting, among other rules. The compliance validation component through the financial intent compiler has to implement these obligations into computational constraints, which have to be tested against any proposed payment instruction before being passed to the execution layer. The NIST AI Risk Management Framework presents an approach to manage risks associated with AI technology, and this framework has informed the design of the compliance validation component [3].

It is equally necessary to design an explainable system. When the system refuses a payment instruction or when further clarification is required from the user, then the reason behind the decision needs to be clearly explained both to the end-user and the regulator. It is for this reason that it is necessary to design the intent compiler separately from the LLM: while the language understanding layer utilizes probabilistic algorithms, the compliance validation layer ensures that decisions are taken based on a deterministic algorithmic process.

Human-in-the-Loop Validation

The human-in-the-loop approach acts as an essential defense against potential mistakes made by machine learning models and also against any sort of malicious attack. There is a threshold incorporated within the architecture that mandates human approval when a transaction exceeds certain predefined value thresholds, involves new parties, or shows signs of anomalies. Human-AI interaction studies have found that a balance of trust is important in these kinds of cases. Users should not overly trust their automation processes to the extent that they ignore any kind of incorrect output, nor should they mistrust them too much to make the automation useless [8].

Evaluation Framework

Because the framework presented in this article is conceptual and has not yet been implemented, this section does not report empirical results. Instead, it specifies the evaluation framework under which a future implementation should be assessed. A rigorous evaluation of an intent-driven payment system requires a measurement methodology that is multivariate and accounts for the performance of the complete processing workflow, from how accurately the system interprets intent at the language-understanding level through to its ability to correctly execute the transaction [8]. The dimensions proposed below are intended as a comprehensive baseline for system as-

Table 2: *Multi-stage pipeline [1, 2].*

Pipeline Stage	Function	Key Techniques	Output
Intent Extraction	Parses natural language input into structured intent representation	Prompt engineering, semantic parsing, contextual embeddings	Intent JSON schema
Entity Resolution	Resolves ambiguous entities (payee, amount, timing) from context	Named entity recognition, user history, contextual lookup	Resolved entity map
Constraint Validation	Applies regulatory rules, business logic, and risk thresholds	Rule engine, compliance ontology, NIST AI RMF alignment	Validated instruction set
Orchestration	Sequences and executes validated payment instructions on rails	Workflow engine, FedNow / fast payment APIs	Executed transaction
Feedback Loop	Learns from interaction outcomes to refine future intent parsing	Fine-tuning, reinforcement from user correction	Updated model weights / logs

sessment, and they define the empirical agenda for follow-on work. Table 2, compiled by the authors based on foundational LLM literature [1, 2], decomposes the proposed pipeline into its constituent stages and identifies the function, key techniques, and output of each. Table 3, compiled by the authors with reference to the NIST AI Risk Management Framework [3] and prior work on human-AI interaction evaluation [8], summarizes the proposed evaluation metrics and methods.

The metrics shown in Table 3 are proposed for use by a future empirical study; no measured values are reported in this article. Intent accuracy is defined as the proportion of natural language inputs that are correctly mapped to the intended payment action, to be assessed against a labeled ground-truth corpus. Execution correctness evaluates whether validated instructions produce the correct payment outcomes in end-to-end simulation. Ambiguity resolution rate captures the system's capacity to handle underspecified inputs without requiring repeated user clarification, which directly affects the user experience quality of the system. The latency, calculated at the 95th percentile to account for tail effects, needs to be evaluated relative to the response time demands of the target environment [6].

The comparative analysis with rule-based systems offers the main baseline for experiments in order to evaluate the gain in capabilities offered by the use of the language model. The adversarial analysis, in which adversarial inputs are constructed to provoke ambiguity, break schemas, or perform prompt injection attacks, evaluates the robustness of the system in face of malicious usage of the system [4].

DISCUSSION

The intent-driven payment paradigm represents a substantive reconceptualization of the user's relationship to financial infrastructure. By interposing an LLM-based interpretation layer between user intent and transaction execution, the architecture absorbs a category of cognitive labor that has historically been delegated to the user, thereby reducing friction and broadening access to financial services for populations with limited familiarity with conventional banking interfaces. Because this discussion precedes empirical validation of the framework, the trade-offs identified here are derived from the architectural analysis and

Table 3: *Proposed evaluation rubric for future empirical assessment of intent-driven payment systems [3, 8].*

Metric	Definition	Evaluation Method
Intent Accuracy	Proportion of inputs correctly mapped to intended payment action	Simulated test sets with labelled ground-truth intents
Execution Correctness	Rate at which validated instructions produce correct payment outcomes	End-to-end simulation against rule-based baseline
Ambiguity Resolution Rate	Frequency with which the system successfully resolves underspecified inputs without user re-prompting	Adversarial and edge-case input batteries
Latency (P95)	95th-percentile end-to-end response time from input to validated instruction	Load testing under realistic transaction volumes
User Satisfaction Score	Subjective assessment of interaction quality and trust	Post-interaction surveys and human-AI interaction studies
Compliance Pass Rate	Proportion of validated instructions that satisfy all applicable regulatory constraints	Automated rule-checking against compliance ontology

from analogous evidence in adjacent domains (form-filling, tool use, and task-oriented dialogue). Empirical trade-off analysis based on a running implementation is explicitly outside the scope of this conceptual framework paper; quantifying these trade-offs against a deployed system is part of the empirical agenda outlined in Section 8.1 and Section 9.

Nevertheless, the adoption of this paradigm entails risks that must be systematically addressed rather than minimized in scope. The probabilistic nature of LLM inference means that the intent processing layer will, with nonzero frequency, produce incorrect or partially correct interpretations. In a financial context, such errors carry consequences that differ in kind from those associated with incorrect answers in informational domains: a misinterpreted payment instruction may result in irreversible financial harm. The architecture therefore places a premium on the correctness of the validation layer and the design of human-in-the-loop mechanisms, treating the LLM not as an infallible oracle but as a high-accuracy draft generator subject to mandatory downstream verification [3, 8].

Trust and reliability constitute the defining governance challenges for systems of this type. Users must develop accurate mental models of system capabilities and limitations to interact with it appropriately; overconfidence in system accuracy may lead to inadequate scrutiny of confirmation prompts, while excessive skepticism may erode the usability benefits that motivate the architecture. Ongoing research in human-AI interaction provides theoretical and empirical resources for navigating this challenge [8], and its findings should inform both interface design and user onboarding protocols.

The security dimension warrants particular emphasis. The consequences of a successful prompt injection attack in a payment context, potentially including the unauthorized initiation of high-value transactions, are severe enough that adversarial robustness must be treated as a primary design constraint rather than a secondary hardening concern [4]. Architectural controls, monitoring infrastructure, and incident response protocols must be designed in anticipation of adversarial exploitation attempts.

Limitations

This article presents a conceptual framework and is subject to several limitations that bound the strength of its claims. First, the framework has not been implemented; the trade-offs surfaced in the architectural analysis have not been quantified against a running system, and the performance, accuracy, and robustness characteristics of the pipeline remain to be established empirically. Second, the evaluation framework presented in Section 7 specifies what should be measured but does not itself supply measurements; absolute thresholds for intent accuracy, latency, and compliance pass rate that would qualify a deployment as production-ready cannot be set without implementation data. Third, the regulatory analysis is conducted at the level of regime category (AML, KYC, transaction reporting) rather than against the specific statute set of any single jurisdiction, and porting the framework into a concrete jurisdiction will surface compliance details that the current treatment abstracts away. Fourth, the security analysis identifies prompt injection as the dominant adversarial concern but does not enumerate or stress-test specific attack vectors, which an implementation would need to address through red-teaming. Resolving these limitations is the empirical agenda implied by this article.

Future Work

Several productive directions for future research emerge from the framework presented in this article. The integration of intent-driven payment capabilities within autonomous financial agent architectures, wherein an LLM-based agent manages a portfolio of financial tasks across multiple platforms and accounts, represents a natural extension of the single-transaction paradigm described here. Such multi-turn, multi-objective settings introduce additional challenges in context management, authorization scope, and error recovery that warrant dedicated investigation.

Cross-platform interoperability constitutes a further research frontier. The present framework assumes a relatively homogeneous execution substrate, but real-world deployments would require the orchestration layer to interface with multiple payment networks, each with distinct message formats, settlement terminologies, and compliance regimes. Designing abstraction layers that retain the intent-driven interactive nature of this paradigm despite this heterogeneity is an important problem for engineering research.

Improvements in structured task decomposition, uncertainty estimation, and formal verification of model outputs can be expected to contribute to the effectiveness of the intent processing module in a substantial manner. Specifically, advances in these areas may reduce reliance on the rule-based validation layer over time, thereby allowing better handling of complex payments [2, 3].

Conclusion

This article describes an intent-driven payment solution framework that uses a large language model as an interpretation layer between natural language user commands and payment workflows. This multi-stage process of identifying intent, resolving entities, validating constraints, and orchestrating workflows provides solutions to the fundamental issues of semantics, schema, regulatory compliance, and security associated with LLM-based financial interfaces before their application in the industry environment.

The financial intent compiler introduced in this work is designed to enforce that LLM outputs are deterministic, auditable, and aligned with applicable regulatory frameworks, thereby aiming to support the correctness properties that financial infrastructure demands while accommodating the expressiveness and accessibility benefits of natural language interaction. As mentioned in Section 7, this approach is an effective basis for performing empirical evaluations on the systems based on this architecture.

The idea of an intent-driven payment model appears quite promising, as it allows for significant reduction of the transaction friction, more convenient access to finance services for different users, and smart financial automation. However, these challenges require more efforts within a number of fields – from natural language processing through financial systems design and regulation to human-computer interaction.

References

[1] Tom B. Brown et al., "Language Models are Few-Shot Learners," arXiv, 2020. Available: https://arxiv.org/pdf/2005.14165

[2] OpenAI, "GPT-4 Technical Report," arXiv, 2023. Available: https://cdn.openai.com/papers/gpt-4.pdf

[3] National Institute of Standards and Technology, "AI Risk Management Framework," NIST, 2023. Available: https://www.nist.gov/itl/ai-risk-management-framework

[4] OWASP, "OWASP top 10 for large language model applications," OWASP. Available: https://owasp.org/www-project-top-10-for-large-language-model-applications/

[5] Bank for International Settlements, "Fast payments: design and adoption," BIS Quarterly Review, 2024. Available: https://www.bis.org/publ/qtrpdf/r_qt2403c.pdf

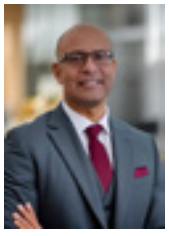
[6] Federal Reserve Financial Services, "FedNow service operating procedures," Federal Reserve Financial Services, 2025. Available: https://www.frbservices.org/binaries/content/assets/crsocms/resources/rules-regulations/062425-fednow-service-operating-procedures-redline.pdf

[7] Ruisheng Cao et al., "Semantic Parsing with Dual Learning," Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. Available: https://aclanthology.org/P19-1007.pdf

[8] Shalini De Mello, "AI-Mediated Human Interaction," Proceedings of the 33rd ACM International Conference on Multimedia, 2025. Available: https://dl.acm.org/doi/10.1145/3746027.3758123

Received: April 28, 2026
Accepted: May 4, 2026

About the authors



Vijay Narayanan is a Software Engineering Director at U.S. Bank, where he leads innovation initiatives in AI, automation, and digital banking platforms. He has over two decades of experience in software engineering, enterprise architecture, fintech modernization, and large-scale banking transformation programs across the United States, Europe, and India. His expertise includes artificial intelligence, mobile banking, distributed systems, cloud-native platforms, fraud detection, and financial technology solutions. Prior to joining U.S. Bank, he held leadership roles with JPMorgan Chase and Citibank. Vijay is the inventor and co-inventor of multiple U.S. patents in AI-powered mobile check deposit and OCR technologies. He holds a Bachelor of Engineering degree from Nagpur University and a PGP in AI/ML from the University of Texas at Austin. He can be contacted at vijaynarayanan.blr@gmail.com.

For citation

Vijay Narayanan, Intent-Driven Payments: A Proposed Framework for Using Large Language Models to Translate Natural Language into Structured Payment Instructions, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:48-56, (UDC: 811.111'373.612.2:7.038.53), (DOI: 10.7251/JIT2601048N), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Security Analysis of the S-DES Cryptographic System

Dragana Božilović Đokić[1], Vladimir Đokić[2], Lazar Stošić[3], Željko Stanković[4], Olja Krčadinac[5]

[1]*University Union Nikola Tesla, draganadjokic@unionnikolatesla.edu.rs, ORCID ID: 0000-0002-9206-2764*
[2]*University Union Nikola Tesla, vladimirdjokic@unionnikolatesla.edu.rs, ORCID ID: 0009-0004-9678-6999*
[3]*Univerzitet Union Nikola Tesla, lstosic@unt.edu.rs, ORCID ID: 0000-0003-0039-7370*
[4]*University Union Nikola Tesla, stanz@medianis.net, ORCID ID: 0000-0002-9893-9088*
[5]*Universoty Union Nikola Tesla, okrcadinac@unionnikolatesla.edu.rs, ORCID ID: 0000-0002-6299-371X*

Review article

https://doi.org/10.7251/JIT2601057DJ UDC: 351.817:336.746

Abstract: S-DES (Simplified Data Encryption Standard) is a pedagogically oriented, reduced-complexity variant of the full DES algorithm. It operates on 8-bit plaintext blocks with a 10-bit secret key, providing a tractable environment for studying fundamental cryptanalytic methods — linear cryptanalysis, differential cryptanalysis, and combined linear-differential cryptanalysis. This paper analyzes the architectural design, subkey generation mechanism, permutation logic, and security properties of S-DES. The cipher's vulnerability to brute-force exhaustive search and differential cryptanalytic attacks is examined in detail, and its potential use as an image encryption primitive — enhanced through chaotic key generation — is evaluated experimentally. The study concludes that, while S-DES is cryptographically inadequate for practical deployment, it constitutes a highly effective educational tool for illustrating the core principles of symmetric block cipher design.

Keywords: S-DES, DES, cryptography, differential cryptanalysis, symmetric encryption, image encryption

Introduction

Cryptography is a scientific discipline concerned with methods for securing information against unauthorized access. The word derives from the ancient Greek kryptos (hidden, secret) and graphos (writing). A cryptographic algorithm transforms human-readable plaintext into an unintelligible ciphertext; cryptanalysis is the complementary science of recovering plaintext or secret keys from ciphertext without prior knowledge of the decryption secret. In any well-formed cryptosystem, the decryption function must be the exact mathematical inverse of the encryption function, though the converse need not hold [1].

S-DES is a scaled-down variant of the Data Encryption Standard (DES). It retains the essential structural properties of DES while operating on considerably smaller parameters: 8-bit plaintext blocks encrypted with a 10-bit key. This reduced scale makes S-DES a practical teaching cipher for introducing students to contemporary cryptanalytic techniques such as linear cryptanalysis, differential cryptanalysis, and combined linear-differential cryptanalysis [2].

This paper presents an analysis of S-DES, treating it as a representative simplified block cipher that makes core cryptographic concepts tractable for study. The investigation covers three areas: (i) the internal architecture of S-DES, including its data flow, subkey generation, and permutation logic; (ii) the security posture of S-DES with respect to its structural simplicity and inherent key-space limitations; and (iii) the applicability of exhaustive-search and differential cryptanalytic attacks against the cipher. Additionally, the paper evaluates S-DES as an image encryption primitive when combined with chaotic key generation.

S-DES has been widely used in academic laboratory settings for cryptography and information theory courses, where it serves as a vehicle for visualizing key scheduling, substitution functions, and permutation operations. The central research question is how S-DES processes an 8-bit data block with a 10-bit key in the context of standard DES design principles, and

how its security constraints compare with those of more robust symmetric ciphers.

Three research hypotheses were formulated on the basis of the reviewed literature:

(H1): S-DES achieves faster encryption and decryption than full DES for both single and multiple data blocks, owing to its reduced input and key sizes.

(H2): Using S-DES in educational contexts provides a clearer understanding of cryptographic operations — such as permutations and key selection — with lower conceptual complexity than traditional DES.

(H3): Under experimental conditions, S-DES is capable of encrypting a color image when augmented with a chaotic key generation strategy.

The motivation for this research is to demonstrate, through a concrete practical example, the cryptographic properties and educational value of the S-DES algorithm.

CRYPTANALYSIS OF S-DES

In S-DES, the same master key is used for both encryption and decryption; however, the subkeys are applied in reverse order during decryption, making the decryption procedure the structural mirror image of encryption. Each plaintext block is subjected to an initial permutation (IP), followed by two rounds of key-dependent computation, and concluded by the application of the inverse initial permutation (IP^{-1}).

Keys

S-DES uses a 10-bit master key to generate two distinct 8-bit subkeys, K1 and K2, each applied in a dedicated cipher round. The key schedule is denoted KS [3].

Key generation begins by applying Permuted Choice 1 (PC1) to the master key, yielding two 5-bit halves: C0 and D0. Key bits are indexed 0 through 9. The PC1 bit-selection pattern is given in Table 1.

Table 1. *PC1 Permutation [3]*

9	**7**	**3**	**8**	**0**
2	6	5	1	4

The upper row of Table 1 defines the bit positions constituting C0; the lower row defines D0. Each half is then independently cycled one position to the left, producing C1 and D1. Subkey K1 is formed by concatenating C1 and D1 and applying Permuted Choice 2 (PC2), which selects 8 of the 10 available bits, as shown in Table 2.

Table 2. *PC2 Permutation [3]*

3	**1**	**7**	**5**	**0**	**6**	**4**	**2**

To obtain K2, the halves C1 and D1 are each shifted two further positions to the left, yielding C2 and D2. These are concatenated and subjected to the same PC2 mapping, producing the second 8-bit subkey K2. Figure 1 shows the complete key schedule.

[Figure 1. Subkey generation schedule for K1 and K2 [3].]

Encryption

The encryption function is expressed as follows:

$C = E(P, K) = IP^{-1}(f_2(f_1(IP(P), K1), K2))$ (1) [3]

The 8-bit plaintext block P is first subjected to IP. The resulting byte is split into two 4-bit halves, L0 and R0, according to Table 3.

Table 3. *Initial Permutation IP [4].*

7	**6**	**4**	**0**
2	5	1	3

L0 contains the bits at positions 7, 6, 4, and 0 of the input; R0 contains bits at positions 2, 5, 1, and 3. Round 1 applies the Feistel transformation [4]:

$L1 = R0, R1 = L0 \oplus f(R0, K1)$ (2)

Round 2 uses the outputs of Round 1 [4]:

$L2 = R1, R2 = L1 \oplus f(R1, K2)$ (3)

After Round 2, R2 and L2 are concatenated in that order to form R2L2, which is then passed through IP^{-1} to produce the final ciphertext. Figure 2 illustrates the complete encryption dataflow.

[Figure 2. Encryption dataflow diagram [4].]

Round Function f

The round function f takes a 4-bit half-block and an 8-bit subkey as inputs. The first operation is the expansion function E, which maps 4 bits to 8 bits according to the bit-selection pattern in Table 4 [5].

Table 4. *E-Bit Expansion [5]*

3	**0**	**1**	**2**	**1**	**2**	**3**	**0**

The 8-bit result E(R) is XORed with the current

round subkey. The resulting byte is partitioned into two 4-bit groups: B1 (the four most significant bits) and B2 (the four least significant bits). B1 and B2 are passed to substitution boxes S0 and S1, respectively. Each S-box accepts a 4-bit input and produces a 2-bit output [5].

Table 5. *S0 Substitution Box [5]*

	Col 0	Col 1	Col 2	Col 3
Row 0	1	0	2	3
Row 1	3	1	0	2
Row 2	2	0	3	1
Row 3	1	3	2	0

Table 6. *S1 Substitution Box [5]*

	Col 0	Col 1	Col 2	Col 3
Row 0	0	3	1	2
Row 1	3	2	0	1
Row 2	1	0	3	2
Row 3	2	1	3	0

To illustrate S-box addressing, consider S0 with input 1101. The first and last bits form the binary value 11 (decimal 3), which selects row 3. The two middle bits form the value 10 (decimal 2), which selects column 2. Reading S0 at row 3, column 2 yields 2, encoded in binary as 10. The 2-bit outputs from S0 and S1 are concatenated into a 4-bit string and passed through the permutation P (Table 7), producing the round function output [5].

Table 7. *Permutation P [5]*

1	**0**	**3**	**2**

Decryption follows the identical procedure, with the sole modification that K2 is applied in Round 1 and K1 in Round 2 — i.e., the subkey order is reversed.

Brute-Force Attack

Brute-force cryptanalysis systematically enumerates all possible keys and tests each against a known plaintext-ciphertext pair. The practical feasibility of this approach depends entirely on the size of the key space. A key space of up to 2^{56} candidates (approximately 7.2×10^{16} possible keys) is considered tractable given sufficient computational resources [6]. Systems employing longer keys — such as AES with its 128-bit key — are effectively immune to exhaustive search on general-purpose hardware [7].

With a 10-bit key, S-DES has a search space of only $2^{10} = 1024$ candidates. A single known plaintext-ciphertext pair is therefore sufficient to recover the master key through exhaustive search in negligible time, making a brute-force attack entirely practical against S-DES [8].

Differential Cryptanalysis

Differential cryptanalysis is a chosen-plaintext attack originally developed to target DES-like block ciphers [9]. The adversary selects pairs of plaintexts with a known input difference and studies how that difference propagates through the cipher structure to produce an output difference. Statistical biases in the resulting difference distributions are exploited to recover partial subkey information. The technique has been extensively studied against DES and has had a lasting influence on the design of modern ciphers, which are typically engineered to provide explicit resistance to differential attacks [10].

To understand why this is non-trivial, consider a purely linear cipher $C = P \oplus K$. In this case, the XOR difference of any ciphertext pair is identical to the XOR difference of the corresponding plaintext pair, revealing nothing about the key. S-DES is not a linear cipher: its S-box components introduce nonlinearity, causing ciphertext differences to depend on the secret key. This dependence creates exploitable statistical structure that can be used to recover key bits [11].

S-Box Difference Distribution

For substitution boxes S0 and S1, let X denote the S-box input and Y the output. A difference pair is expressed as $(\Delta X, \Delta Y)$, where $\Delta X = X' \oplus X''$. For a fixed ΔX, the corresponding second input is $X'' = X' \oplus \Delta X$, and the resulting ΔY is tabulated in the difference distribution table.

As an example, for S0 with $\Delta X = 8$ and $\Delta Y = 1$, the distribution table shows that exactly two input pairs satisfy this differential. Since $\Delta Y = 1$, the output values are {1, 3}, and the only input pair that simultaneously yields these outputs while satisfying $\Delta X = 8$ is {9, 1}.

This information enables partial key recovery. Assume $X' = 2$ and $X'' = 8$ for S0 ($\Delta X = 10$). If the corresponding outputs are $Y' = 0$ and $Y'' = 2$ ($\Delta Y = 2$), the actual S-box inputs after XOR with the subkey are $I' = X' \oplus K$ and $I'' = X'' \oplus K$. Since the subkey is added identically to both inputs, it cancels in the difference:

$\Delta I = \Delta X = 10$.

From the distribution table, the input pairs satisfying $\Delta X = 10$ and $\Delta Y = 2$ are {7, 13}. The subkey candidates K are recovered as [11]:

$K = I' \oplus X' = 7 \oplus 2 = 5$ (4)

$K = I'' \oplus X' = 13 \oplus 2 = 15$ (5)

The results are summarized in Table 8.

Table 8. *S-Box Key Candidates for X' = 2, X'' = 8 [11]*

X'	X''	Possible I pair	Key candidates K
2	8	{7, 13}	{5, 15}

METHOD

Several complementary scientific methods were applied to analyze the characteristics and security aspects of the S-DES cryptosystem.

A comparative method was used to contrast the individual processing stages of S-DES against the general principles governing symmetric encryption systems. This approach enabled the identification of structural parallels and substantive differences between the simplified cipher and its theoretical cryptographic foundations, and facilitated an assessment of S-DES security properties in relation to DES.

A case study approach formed the empirical core of the research. The complete encryption and decryption pipeline was traced step by step — covering subkey derivation, initial and inverse permutations, S-box substitutions, and round-level Feistel transformations — allowing a detailed observation of how subkeys K1 and K2 govern the transformation of plaintext to ciphertext.

Descriptive analysis was used to present and interpret the results systematically. The study relied on secondary data sources, primarily peer-reviewed journal articles, academic textbooks, and conference proceedings in the fields of cryptography and information security.

The research proceeded through four sequential phases: (1) collection and critical review of relevant literature; (2) examination of the theoretical foundations of the S-DES algorithm; (3) comparative analysis of algorithmic components; and (4) interpretation of results in the context of symmetric cipher design principles.

FINDINGS

A 10-bit master key is used to generate two distinct 8-bit subkeys, K1 and K2, each applied in a specific cipher round. Key scheduling begins with PC1, which selects and reorders 10 key bits into two 5-bit halves, C0 and D0 (Table 9).

Table 9. *PC1 Mapping [11]*

9	7	3	8	0
2	6	5	1	4

A cyclic left shift is applied independently to C0 and D0, producing C1 and D1. Subkey K1 is formed by concatenating C1 and D1 and applying PC2 (Table 10), which reduces the 10 bits to 8 bits for use as the round subkey.

Table 10. *PC2 Mapping [11]*

3	1	7	5	0	6	4	2

The 4-bit half-block R is expanded to 8 bits by the expansion function E, whose bit-selection pattern is given in Table 11.

Table 11. *E-Bit Selection [11]*

3	0	1	2	1	2	3	0

The expanded block is XORed with the active subkey. The resulting 8-bit value is split into B1 and B2, which are fed to S0 and S1 respectively (Table 12).

Table 12. *S1 Lookup Table [11]*

	Col 0	Col 1	Col 2	Col 3
Row 0	0	3	1	2
Row 1	3	2	0	1
Row 2	1	0	3	2
Row 3	2	1	3	0

In S-box addressing, the first and last bits of the 4-bit input select the row, and the two middle bits select the column. For S0 with input 1101: the outer bits select row 3, and the inner bits select column 2, yielding the value 2 (binary: 10). The 4-bit concatenated output of S0 and S1 is then permuted by P (Table 13).

Table 13. *Permutation P [11]*

1	0	3	2

For the image encryption experiment, the first iterate of a logistic chaotic map applied to the Lena benchmark image is used to derive an S-DES encryption key. The chaotic transformation produces a pseudo-random binary matrix, which is binarized and used to drive pixel-block encryption. This approach leverages the high sensitivity and apparent randomness of chaotic orbits to overcome S-DES's inherently limited key space. The original Lena image and its first-iteration chaotic counterpart are shown in Figures 3 and 4, together with their respective RGB histograms.

[Figure 3. (a) Original Lena image; (b) First-iteration chaotic image [9].]

[Figure 4. (a) RGB histogram of original image; (b) RGB histogram of encrypted image [9].]

The binarized chaotic image serves as the S-DES key. Because the limited key space of S-DES creates significant security risks under repeated use, the high entropy of chaotic key material substantially improves the statistical unpredictability of the encrypted output compared with static key selection.

Discussion

The analysis reveals a fundamental tension between the pedagogical utility and the cryptographic adequacy of S-DES.

The principal advantages of S-DES are: (1) its considerably simpler structure relative to full DES, making it far more accessible for study and implementation; (2) its operation on smaller plaintext blocks with a shorter key, demanding fewer computational resources; and (3) its substantially faster execution compared with DES on equivalent hardware.

Its limitations are equally clear: (1) the 10-bit key is critically insufficient for any real-world security application; (2) the resulting limited key space leaves S-DES vulnerable to trivial exhaustive-search attacks; and (3) when applied to large data payloads, such as high-resolution images, S-DES cannot satisfy practical encryption security requirements.

The integration of chaotic mapping as a dynamic key-generation mechanism substantially compensates for the limited key space of S-DES. Chaotic sequences yield high-entropy key material that improves the statistical unpredictability of ciphertext, demonstrating that appropriate key management strategies can partially offset the weaknesses of a simplified cipher in controlled educational contexts.

Conclusion

Cryptanalysis is a scientific discipline dedicated to deciphering the content of encrypted information without access to the secret decryption key. This paper has shown that even elementary attacks — such as differential cryptanalysis — are highly effective against ciphers with reduced structural complexity.

The differential cryptanalysis case study demonstrated that the full 8-bit second-round subkey of S-DES can be fully recovered through differential cryptanalysis. Given these bits, the remaining 2 unknown bits of the 10-bit master key can be determined by testing only four additional candidates, making complete key recovery computationally trivial.

Despite its cryptographic weakness by modern standards, S-DES holds significant historical and pedagogical importance. Its compact, transparent structure effectively illustrates the core principles of symmetric block cipher design — permutations, substitutions, key scheduling, and Feistel rounds — and the study of its cryptanalytic weaknesses builds intuition for the structural vulnerabilities that motivated the development of stronger standards such as AES.

The image encryption scheme proposed in this paper combines S-DES with a logistic chaotic map. The chaotic key generation strategy compensates for the cipher's limited key space and yields encrypted output with strong statistical randomness properties. These results support hypothesis H3 and suggest that, when augmented with appropriate key generation, S-DES remains a valuable instructional tool for exploring more advanced encryption paradigms.

References

[1] H. Kim et al., "Deep-Learning-Based Cryptanalysis of Lightweight Block Ciphers Revisited," Entropy, vol. 25, no. 7, p. 986, 2023.

[2] Z. Hou, J. Ren, and S. Chen, "Improved Machine Learning-Aided Linear Cryptanalysis: Application to DES," Cybersecurity, Springer, 2025.

[3] K. Dworak and U. Boryczka, "Breaking Data Encryption Standard with a Reduced Number of Rounds Using Metaheuristics Differential Cryptanalysis," Entropy, vol. 23, no. 12, p. 1697, 2021.

[4] S. Sikdar, S. Dutta, and M. Kule, "On Cryptanalysis of 3-DES using Nature-Inspired Algorithms," International Journal of Computer Network and Information Security, pp. 54-71, 2025.

[5] D. Shah et al., "A Novel Approach for Security Enhancement of Data Encryption Standard," Computers, Materials & Continua, vol. 75, no. 3, pp. 5073-5086, 2023.

[6] M. Zheng and H. Kang, "Lattice-Based Cryptanalysis of

RSA-Type Cryptosystems: A Bibliometric Analysis," Cybersecurity, Springer, 2024.
[7] D. Gerault et al., "SoK: 6 Years of Neural Differential Cryptanalysis," Cryptology ePrint Archive, 2024.
[8] A. Jain, V. Kohli, and G. Mishra, "Deep Learning Based Differential Distinguisher for Lightweight Block Ciphers," 2021.
[9] L. Zhang and Z. Wang, "Improving Differential-Neural Distinguisher Model for DES, Chaskey and PRESENT," 2022.
[10] R. Kumari, J. G. Pandey, and A. Karmakar, "An RTL Implementation of the Data Encryption Standard (DES)," 2023.
[11] A. Ullah, M. Khan, and S. Ali, "Lightweight Block Ciphers for Resource-Constrained Environments: A Comprehensive Survey," Future Generation Computer Systems, 2024.

Received: April 21, 2026
Accepted: May 2, 2026

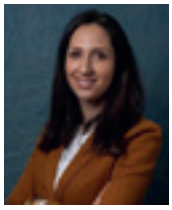
ABOUT THE AUTHORS

Dragana Đokić is a teaching assistant "Union-Nikola Tesla" University, Faculty of Informatics and Computer Science, Belgrade, Republic of Serbia. Finished Master of Science in Mehanical Engineering (M.Sc. MEI.) University of Belgrade. Her current research interests include the fields of computer networks, security, high-perfor-mance systems (HPC), Internet of Things (IoT), software development and testing.

Vladimir Đokić is professor at "Union – Nikola Tesla" University - Faculty of Informatics and Computer Science, Belgrade. He holds a PhD in Information Systems and is actively engaged in teaching and research in the field of information and communication technologies. He is the author and co-author of numerous scientific papers published in international peer-reviewed journals indexed in major scientific databases. His research work is interdisciplinary, combining information systems and computer science with applications in biomedicine, pharmacology, and engineering sciences.

Lazar Stošić is a university professor at the Faculty of Informatics and Computer Science, University Union—Nikola Tesla, Belgrade, Serbia and the President of the Association for the Development of Science, Engineering and Education, in Serbia. He is also a leading researcher at the Center for Scientific Competence of DSTU, Department of Scientific and Technical Information and Scientific Publications Don State Technical University, Russia. His expertise includes computer science, IKT, editorial workflow management, conference organization, web technologies, web design, indexing, XML production, SEO, digital marketing, and new media technologies.

Zeljko Stankovic received his higher education in Cleveland, Ohio, USA, where he graduated in 1981. The topic of the thesis was "Reversible sound in halls". He defended his master's thesis ("Learning control system (LMS) based on ADL SCORM specifications") in 2006 at the University of Novi Sad, Faculty of Science, Department of Informatics. He defended his doctoral dissertation (Laser perception of defined objects and encapsulation of control and logic elements for an autonomous robotic teaching tool) at Singidunum University, Belgrade, in 2010. He has been programming since 1984, creating programs for his first Commodore 64 computer.



Olja Krcadinac (Latinovic, maiden name) is assistant professor at "Union – Nikola Tesla" University - Faculty of Informatics and Computer Science. She earned her Ph.D. in biometric field from University of Belgrade – Faculty of Organizational science, where she conducted groundbreaking research on speaker recognition. In addition to her teaching responsibilities, Olja has authored numerous impactful publications in peer-reviewed journals, contributing valuable insights to the scientific community. Her research focuses on biometric, sensors, IoT and AI, addressing critical issues in AI and making significant contributions to the academic community.

FOR CITATION

Dragana Božilović Đokić, Vladimir Đokić, Lazar Stošić, Željko Stanković, Olja Krčadinac, Security Analysis of the S-DES Cryptographic System, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:57-62, (UDC: 351.817:336.746), (DOI: 10.7251/JIT2601057DJ), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Integrated Approaches in the Development of Intelligent Information Systems: A Comprehensive Review of Cloud, IoT, Big Data, Machine Learning, and Information Forensics Challenges

Olja Krčadinac, Lazar Stošić
"Union – Nikola Tesla" University, Faculty of Informatics and Computer science, Belgrade, Serbia
okrcadinac@unionnikolatesla.edu.rs, ORCID ID: 0000-0002-6299-371X
lstosic@unt.edu.rs, ORCID ID: 0000-0003-0039-7370

Review article
https://doi.org/10.7251/JIT2601063K UDC: 004.7.056.5:004.056

Abstract: The rapid evolution of Integrated Information Systems (IIS) has led to a complex convergence of Cloud Computing, Internet of Things (IoT), and Machine Learning (ML). While this synergy enhances computational efficiency, it introduces significant challenges in information forensics and system security. This paper explores the multidimensional security landscape of unified ecosystems, focusing on the vulnerabilities inherent in distributed resources. We analyze the necessity of "Forensic-by-Design" principles and the role of robust biometric solutions in securing e-commerce and integrated environments. Special attention is given to the impact of user interaction variability on speaker recognition performance, as well as the potential of modern IT tools in assessing and optimizing system integrity. By synthesizing recent advancements in MLOps and cloud-native architectures with empirical findings on digital literacy and security technologies, this study provides a strategic framework for developing resilient and accountable intelligent systems. The findings emphasize that technical excellence must be balanced with rigorous forensic standards to mitigate risks in increasingly dynamic, cloud-based infrastructures.

Keywords: Intelligent Information Systems, Cloud-IoT Security, Biometric Recognition, Information Forensics, IT Tool Assessment

Introduction

The rapid evolution of digital ecosystems has led to a paradigm shift where information systems are no longer mere data storage entities but have evolved into Intelligent Information Systems (IIS). These systems operate at the complex intersection of several transformative technologies that are increasingly becoming inseparable. The proliferation of Internet of Things (IoT) devices across industrial and urban environments generates unprecedented volumes of unstructured data, necessitating robust Cloud Computing infrastructures for scalable storage and high-performance processing [1]. This surge in data volume, variety, and velocity—commonly defined as Big Data—provides the foundational "fuel" for Machine Learning (ML) algorithms to extract actionable insights and enable autonomous decision-making [2].

However, the seamless integration of these heterogeneous technologies remains a significant challenge for both researchers and practitioners. As systems transition toward cloud-native architectures and edge-based processing, the traditional boundaries of software engineering are being redefined [3]. The deployment of intelligent layers requires not only advanced analytical models but also a rigorous operational framework that encompasses DevOps and MLOps practices to ensure system reliability. Furthermore, this increased complexity and distribution of resources significantly expand the cyber-attack surface. In such a landscape, the role of Information Forensics becomes critical. Traditional forensic methods often fail in dynamic, multi-tenant cloud environments and fragmented IoT networks, highlighting a pressing need for "forensic readiness" to be integrat-

ed directly into the system's architectural design [4].

Despite the extensive literature available on individual components such as cloud security or IoT data analytics, there is a notable scarcity of research that addresses the holistic integration of these five domains. Most existing studies focus on isolated optimizations, often overlooking the forensic and security implications of a fully integrated intelligent pipeline [5]. This paper aims to bridge this gap by providing a comprehensive review of the state-of-the-art integrated approaches in IIS development. By synthesizing current research trends and identifying critical challenges across cloud, IoT, big data, machine learning, and digital forensics, this study provides a strategic roadmap for developing resilient, scalable, and forensically sound intelligent systems.

RESEARCH METHODOLOGY

To ensure methodological rigor and reproducibility, this study follows a structured systematic literature review (SLR) approach, focusing on the convergence patterns between distributed cloud resources, edge computing, and forensic accountability, a task that inherently necessitates a multi-disciplinary perspective [1]. The research process was systematically divided into three distinct phases: database querying, screening, and qualitative synthesis.

The academic search was conducted across three leading databases: IEEE Xplore, Scopus, and Google Scholar, targeting peer-reviewed literature published between 2020 and 2026. This specific timeframe was selected to capture the most recent advancements in cloud-native architectures, real-time edge analytics, and automated MLOps workflows [2]. The search strings were designed using Boolean operators to target the precise terminology intersecting intelligent system deployment and forensic readiness, with a particular emphasis on technical standards defined by the National Institute of Standards and Technology [4]. The detailed execution of the search strategy is structured in Table 2.

Strict inclusion and exclusion criteria were applied during the screening phase. To be included in the final corpus, studies had to satisfy the following rigorous conditions: (1) treat security and forensics not as isolated components, but as integral design elements, adhering to the "Forensic-by-Design" principle [5], and (2) address the holistic integration challenges across at least two intersecting domains of the core pipeline (Cloud, IoT, Big Data, Machine Learning, and Information Forensics). Exclusion criteria removed non-English publications, white papers lacking peer review, and studies focusing solely on isolated component optimizations without system-wide integrative relevance.

Furthermore, international standards for cloud security [6] and fresh frameworks for edge logging transparency [7] [8] were factored in to ensure the methodology aligns with the contemporary standards of the natural-mathematical field. Through this rigorous multi-stage filtering process, the initial pool of 365 records was systematically distilled into a highly relevant core corpus of 35 papers for deep architectural synthesis.(Table 1)

Table 1. *Systematic Literature Search Strategy Matrix*

Database	Search Query / Keywords	Initial Results	After Title/ Abstract Screening	Final Selection
IEEE Xplore	("Intelligent Information Systems" OR "Cloud-IoT") AND ("Forensic-by-Design" OR "Forensic Readiness")	145	42	15
Scopus	("MLOps" OR "Big Data Architecture") AND ("Digital Forensics" OR "Immutable Logging")	122	31	12
Google Scholar	("Intelligent Systems" AND "Cloud Security Standards" AND "NIST forensic")	98	24	8
Total		365	97	35

CONCEPTUAL FRAMEWORK AND ARCHITECTURAL INTEGRATION

The The development of Intelligent Information Systems (IIS) requires a multi-layered architectural approach that transcends traditional client-server models, moving towards a more fluid and distributed paradigm [1]. At the heart of this integration is the cloud-IoT continuum, which bridges the gap between the physical perception layer and centralized processing power. Unlike static systems, a modern IIS must manage "data gravity" by strategically deploying Edge and Fog computing layers to process information as close to the source as possible [9] [3]. This reduces

Table 2. *Comparative Analysis of Integrated IIS Frameworks and Strategic Novelty Positioning*

Framework / Study	Architectural Layers Covered	Primary Security / Forensic Mechanism	Cross-Layer Audit Trails	Edge Resource Optimization
Chang (2021)	Cloud Core	Reactive Cloud Forensics	No	No
Zhao et al. (2023)	IoT, Edge, Cloud	Standard Encryption	No	Yes (Data Gravity focus)
Boutros & Shah (2024)	IoT, Edge	Local Edge-Inference Logging	Partial	Yes (Decentralized)
Al-Mansoori et al. (2025)	ML Layers, Cloud Core	Dynamic Concept Drift Alerts	No	No
Proposed IIS Framework	IoT, Edge, Big Data, ML, Cloud	Forensic-by-Design & Immutable Hash Chains	Yes (Full Continuum)	Yes (Hierarchical Filtering)

latency and prevents network congestion, allowing the system to perform real-time filtering and initial inference before any data reaches the core cloud infrastructure. This hierarchical processing is essential for maintaining the responsiveness required in industrial or urban smart environments [10].

Transitioning from data acquisition to management, the integration challenges shift toward handling the sheer velocity and variety of Big Data. Relational databases, while robust for transactional integrity, are increasingly supplemented by polyglot persistence strategies [11]. By utilizing a combination of NoSQL systems for unstructured telemetry and Data Lakehouse architectures for analytical processing, integrated systems can provide the necessary "fuel" for advanced intelligence[12] [2].

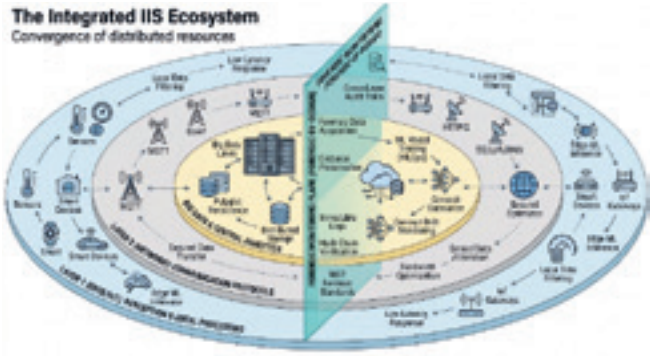


Figure 1. *The Integrated IIS Ecosystem – Convergence of distributed resources and forensic monitoring.*

This unified data layer is what allows Machine Learning (ML) to move from a research experiment into an operational reality. The integration of ML requires a transition from traditional software development cycles to MLOps frameworks, where model training, deployment, and monitoring for concept drift are treated as continuous, automated processes [13] [14]. In this context, the intelligent layer is not a standalone component but a dynamic service that evolves in tandem with the incoming data streams, ensuring that the system's decision-making capabilities remain accurate even in volatile environments [2] [15].

The architecture visualized in Figure 1 is structured as a series of concentric operational layers, emphasizing that no component operates in isolation within a modern intelligent ecosystem. The outermost layer represents the Edge/IoT perception zone, where sensors and smart devices perform local data filtering and initial machine learning inference. This decentralized processing is critical for reducing the "data gravity" effect, ensuring that only relevant, high-value information is transmitted through the network layer via optimized protocols like MQTT and CoAP, which must be strictly secured against ephemeral data loss [16] [3].

As data penetrates the inner layers, it reaches the Cloud Core and Big Data analytics zone, where polyglot persistence and data lakehouse structures provide the necessary scalability for complex model training and automated MLOps cycles [12] [14].

However, the defining feature of this integrated model, as illustrated in Figure 1, is the vertical "Forensic Monitoring Plane" that physically intersects every single operational layer. This cross-cutting component represents the "Forensic-by-Design" principle, asserting that forensic readiness is not an afterthought but a fundamental, structural architectural requirement [5]. By embedding immutable logging, cryptographic hash chain verification, and NIST-compliant forensic data acquisition points across the entire continuum—spanning from the IoT edge gateways to the central cloud database—the system ensures that all autonomous decisions, state transitions, and data transfers remain fully verifiable, transparent, and forensically sound [8] [4] [17].

This holistic visualization underscores the primary argument of this study: the long-term reliability and legal-technical accountability of an intelligent system are directly proportional to the seamless

structural integration of its analytical, operational, and forensic domains [10] [18].

Information Forensics and Security Challenges in Integrated Systems

The convergence of distributed technologies inherently amplifies the attack surface of Intelligent Information Systems (IIS), presenting multi-faceted forensic and security challenges that span the entire cloud-edge continuum. At the infrastructure level, multi-tenant cloud environments introduce complex isolation anomalies, making traditional, reactive digital forensics practically obsolete. Ensuring forensic readiness in contemporary systems requires shifting away from post-incident data gathering toward proactive, containerized tracking mechanisms capable of automated evidence acquisition without violating cross-layer data privacy standards [19] [20].

Furthermore, the integrity of the analytical pipeline introduces a new frontier of vulnerabilities within continuous deployment environments. The operationalization of machine learning through MLOps frameworks creates specific blind spots, particularly regarding data poisoning, model inversion, and adversarial manipulations [13]. When an autonomous model undergoes continuous retraining based on dynamic telemetry, traditional audit trails fail to capture the subtle metadata shifts associated with data drift. To establish legally and technically binding accountability, the security architecture must deploy continuous, dynamic drift monitoring and real-time cryptographic hash chains capable of validating the evolutionary lifecycle of the deployed models [21] [15]. Table 3 shows key IIS challenges and forensic impacts.

Table 3. *Key IIS Challenges and Forensic Impacts*

Domain & Layer	Integration Challenge	Forensic Implication
IoT / Edge	Latency & Heterogeneity	Volatile evidence; short log retention
Big Data	Velocity & Variety	Integrity of massive, fluid datasets
ML Layers	Model Drift & MLOps	Difficulty in verifying training inputs
Cloud Core	Multi-tenancy	Data remanence; isolation of traces
System-wide	Interoperability	Fragmented audit trails across nodes

Finally, as specialized security tools become deeply integrated into organizational workflows, the human factor remains a critical, volatile vector. The technical implementation of a "Forensic-by-Design" architecture is heavily dependent on the digital literacy, operational security compliance, and socio-technical adaptation of the personnel executing the protocols [22]. Consequently, bridging the gap between sophisticated technical integration and the practical human-system interaction paradigm represents one of the most significant, ongoing challenges in engineering resilient and forensically sound intelligent systems.

Technical Challenges and Framework Limitations

While the proposed integrated framework offers a holistic paradigm for secure and intelligent system deployment, several systemic challenges and inherent limitations must be addressed to translate this conceptual model into operational reality. First and foremost, implementing a continuous "Forensic-by-Design" plane requires constant telemetry capture and the real-time generation of cryptographic hash chains. When deployed on low-power IoT edge devices operating via lightweight protocols such as MQTT or CoAP, this cryptographic verification introduces significant computational friction and energy overhead [16]. Striking a sustainable operational balance between high-throughput machine learning inference and dense, immutable logging creates a critical infrastructure bottleneck. Consequently, overcoming this limitation requires advanced hardware optimization and the development of dynamic resource-allocation protocols that prevent telemetry from degrading edge processing performance [7].

Furthermore, as the system ingests heterogeneous data streams across the edge-cloud continuum, the underlying architecture must navigate the severe challenges associated with "data gravity" [9]. Retaining forensically sound, multi-layer audit trails over extended regulatory retention periods inherently leads to massive, exponential storage requirements within the Big Data lakehouse infrastructure [12]. Without specialized data-pruning mechanisms that can selectively archive records while fully preserving metadata integrity, the framework risks escalating cloud infrastructure costs to unsustainable levels. This architectural vulnerability is tightly coupled with the risks present

in the analytical layer, where autonomous decision-making processes remain susceptible to adversarial manipulation and structural data poisoning [21]. While real-time drift monitoring can alert administrators to macro-level anomalies [15], retroactively isolating the exact point of data corruption in a continuous retraining pipeline remains highly problematic due to the inherently non-transparent, "black-box" nature of deep neural networks [13].

Ultimately, even the most resilient automated architecture remains bound to the unpredictability of human operational profiles. The practical viability of deploying forensically ready information systems relies on the digital literacy, security awareness, and strict adherence to technical protocols of the organizational personnel managing the system [22]. Because user interaction variability introduces an unpredictable socio-technical variable, technical architectures can only mitigate these behavioral inconsistencies through automated constraints, but can never entirely eliminate them from the security equation. Navigating these interconnected trade-offs between edge capability, storage efficiency, model explainability, and human compliance represents the next crucial phase in the evolution of unified intelligent systems.

Conclusion

The rapid and continuous evolution of Intelligent Information Systems demands a fundamental paradigm shift from isolated component optimization toward holistic architectural integration. This study has systematically reviewed and synthesized the complex convergence patterns within the cloud-IoT continuum, demonstrating that the long-term viability, analytical precision, and security of modern ecosystems are entirely dependent on how seamlessly their cloud, edge, Big Data, and Machine Learning layers are interconnected. By adopting a rigorous, reproducible systematic literature review methodology spanning a comprehensive corpus of recent literature, this paper has mapped out the foundational standards and contemporary milestones that define resilient distributed architectures.

The primary structural contribution of this research lies in the conceptual validation of the "Forensic-by-Design" principle as a non-negotiable architectural requirement. Rather than treating security and digital forensics as reactive, post-incident measures, the unified framework presented herein demonstrates that establishing permanent, cross-layer audit trails and immutable logging mechanisms is essential for securing autonomous environments. This integrated approach effectively bridges the gap between high-throughput intelligent processing (MLOps) and strict legal-technical accountability, ensuring that system state transitions and machine learning decisions remain fully transparent and verifiable.

Nevertheless, translating this holistic conceptual model into widespread industrial practice reveals several critical technical trade-offs and inherent limitations. As demonstrated throughout the analytical synthesis, balancing real-time edge inference with the computational and energy overhead of cryptographic hash chain logging remains a substantial challenge for resource-constrained IoT gateways. Furthermore, managing data gravity within expanding lakehouse infrastructures without compromising metadata integrity introduces severe storage scalability issues, while the "black-box" nature of complex neural networks continues to obfuscate absolute forensic clarity during concept drift events. Ultimately, the unpredictable nature of human interaction variability and organizational digital literacy underscores that technical resilience must always co-evolve with socio-technical adaptation.

Future research trajectories must therefore focus on mitigating these specific architectural bottlenecks. Priority should be given to developing ultra-lightweight, decentralized cryptographic logging protocols optimized for edge deployment, as well as designing explainable AI (XAI) frameworks that can be natively audited within automated MLOps pipelines. Additionally, empirical validation through real-world deployment scenarios will be crucial for refining the hierarchical filtering mechanisms proposed in this model. By continuously addressing these technical frontiers, the academic and professional community can advance toward a future where intelligent systems are not only highly efficient but inherently accountable, transparent, and structurally secure.

Acknowledgment

This research was carried out within the framework of the UNT-Lab – Artificial Intelligence Center at the Faculty of Informatics and Computing, University „Union – Nikola Tesla", Belgrade. The authors gratefully acknowledge the support of the Center's infrastructure and the internal research and development programs of the University. This work contributes to the ongoing initiatives

of the UNTLab AI Center in the field of distributed and intelligent information systems.

REFERENCES

[1] Armbrust, M., et al. (2022). "A Decade of Cloud Computing: Lessons and Future Directions." Communications of the ACM, 65(4), pp. 48-56.

[2] Kreuzer, G. (2025). "Adaptive MLOps: Managing Model Integrity in Real-Time Data Streams." International Conference on Software Engineering.

[3] Zhao, Z., et al. (2023). "Cloud-Native Architectures and Mobile Edge Computing: A Survey." IEEE Access, 11, pp. 12450-12475.

[4] NIST (2020). "Special Publication 800-101 Revision 1: Guide to Integrating Forensic Techniques into Incident Response." National Institute of Standards and Technology.

[5] Chang, V. (2021). "A proposed framework for Cloud Computing Forensics." Journal of Cloud Computing, 10(1).

[6] ISO/IEC (2020). "ISO/IEC 27017:2015 — Information technology — Security techniques — Code of practice for information security controls."

[7] Boutros, F., & Shah, J. (2024). "Continuous Forensic Readiness in Decentralized IoT Systems using Edge-Inference Logging." Journal of Digital Forensic Practice, 12(1), pp. 78-92.

[8] Gomez, L., et al. (2025). "Forensic-by-Design Protocols for Smart City IoT Edge Gateways." Computers & Security, 148, 103650.

[9] Martinez, J. (2024). "Data Gravity Challenges in Cloud-Fog Architectures: A Forensic Perspective." IEEE Cloud Computing, 11(2), pp. 34-41.

[10] Castell, N., et al. (2024). "Next-generation spatial networks for urban intelligence." IEEE Internet of Things Journal, 11(2), pp. 1102-1115.

[11] Wang, J., et al. (2024). "Polyglot Persistence in Big Data Ecosystems: Challenges and Solutions." ACM Computing Surveys, 56(3).

[12] Ibrahim, O. (2025). "Architecting Forensic Readiness in Big Data Lakehouses." Data & Knowledge Engineering, 151, 102310.

[13] Hassan, M., & Tariq, S. (2024). "MLOps Vulnerabilities: A Systematic Survey of Security and Accountability in Autonomous Systems." ACM Computing Surveys, 56(8), pp. 112-135.

[14] Luceri, A., et al. (2024). "Adversarial Robustness in Distributed MLOps Pipelines." Journal of Big Data, 11(2).

[15] Ranganathan, S. (2026). "Real-Time Drift Monitoring and Forensic Accountability in Intelligent Systems." Journal of Software Engineering and Applications, 19(2), pp. 88-104.

[16] Patel, H., et al. (2025). "Securing MQTT and CoAP Protocols against Ephemeral Data Loss at the Perception Layer." Internet of Things Journal, 18, pp. 301-315.

[17] Nguyen, T., et al. (2025). "Immutable Logging for Distributed Ledger Architectures in Edge-Cloud Continnums." Journal of Network and Computer Applications, 240, 103890.

[18] White, D., et al. (2025). "Cross-Layer Audit Trails for Intelligent Information Systems: Integrating DevOps, MLOps, and SecOps." Journal of Systems and Software, 215, 112045.

[19] Fernandez, R., & Kumar, A. (2024). "NIST-Compliant Incident Response Frameworks for Multi-Tenant Cloud Environments." Cybersecurity Review, 17(3), pp. 210-225.

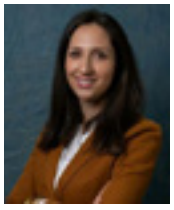
[20] Kim, J., & Park, Y. (2026). "Automated Evidence Acquisition Protocols in Containerized Cloud Environments." Future Generation Computer Systems, 174, pp. 89-101.

[21] Al-Mansoori, S., et al. (2025). "Securing Cloud-Native MLOps Pipelines against Data Poisoning and Concept Drift." IEEE Transactions on Dependable and Secure Computing, 22(2), pp. 145-159.

[22] Silva, E., & Santos, M. (2024). "Digital Literacy and the Adoption of Forensically Ready Information Systems: A Case Study Approach." Technology in Society, 76, 102432.

Received: April 29, 2026
Accepted: May 4, 2026

ABOUT THE AUTHORS



Olja Krcadinac (Latinovic, maiden name) is assistant professor at "Union – Nikola Tesla" University - Faculty of Informatics and Computer Science. She earned her Ph.D. in biometric field from University of Belgrade – Faculty of Organizational science, where she conducted groundbreaking research on speaker recognition. In addition to her teaching responsibilities, Olja has authored numerous impactful publications in peer-reviewed journals, contributing valuable insights to the scientific community. Her research focuses on biometric, sensors, IoT and AI, addressing critical issues in AI and making significant contributions to the academic community.

Lazar Stošić is a university professor at the Faculty of Informatics and Computer Science, University Union—Nikola Tesla, Belgrade, Serbia and the President of the Association for the Development of Science, Engineering and Education, in Serbia. He is also a leading researcher at the Center for Scientific Competence of DSTU, Department of Scientific and Technical Information and Scientific Publications Don State Technical University, Russia. His expertise includes computer science, IKT, editorial workflow management, conference organization, web technologies, web design, indexing, XML production, SEO, digital marketing, and new media technologies.

FOR CITATION

Olja Krčadinac, Lazar Stošić, Integrated Approaches in the Development of Intelligent Information Systems: A Comprehensive Review of Cloud, IoT, Big Data, Machine Learning, and Information Forensics Challenges, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:63-68, (UDC: 004.7.056.5:004.056), (DOI: 10.7251/JIT2601063K), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Implementation of Cowrie Honeypot System and Improvement of Log Analysis

Milan Panić[1], Nemanja Maček[2]

[1] *Pan-European University "APEIRON", Faculty of Information Technology, Banja Luka, Bosnia and Herzegovina, milan.v.panic@apeiron-edu.eu, ORCID ID: 0009-0006-9270-7354*

[2] *University Business Academy in Novi Sad, Faculty of Social Sciences, Belgrade, Republic of Serbia, maceknemanja@gmail.com, ORCID ID: 0000-0002-3465-7524*

Professional paper

https://doi.org/10.7251/JIT2601069P UDC: 37.016::003-028.3

Abstract: This paper aims to explain how honeypots work, how they are implemented, and why they have become a key aspect of cybersecurity. Honeypots are capable of doing everything from detecting new attacks never seen before in their environment to tracking programmed credit card fraud and identity theft. The paper implements the Cowrie honeypot system in a controlled environment to simulate attacks on SSH and Telnet services. Special focus is placed on the analysis of generated JSON log records, the complex structure of which makes forensic processing difficult. As a contribution to the paper, a Python helper module has been developed to convert raw log files into a readable and structured text format, thus improving the efficiency of security event analysis.

Keywords: Cowrie, honeypot, SSH, Telnet, log

Introduction

High-tech crime represents one of the biggest security challenges in the world. The development of information technologies, in addition to the benefits that it has improved the operation of many systems, has also led to the development of new forms of crime. The previous decade has been marked by a constant fight against new forms of high-tech crime, an increase in crimes in this area, but also by the constant strengthening of the capacities of departments for combating high-tech crime.

In order to create a system that is interesting enough for hackers to carry out an attack, we must create a sufficiently tempting system. They will try to gain access by using security vulnerabilities in the system. By following hackers, we are not sure whether we will be the ones who will have control. It is necessary to determine the following:

- Does the hacker know that this is a real system or is it a honeypot?
- Is he aware of how great a tool it is for administrators to obtain information about security vulnerabilities in the system, but also about the attackers themselves?
- What is gained from hacking such a system?

A (high-level) hacker is often like a cat, he enters quietly, leaves small traces and makes the whole game an art of evasion. A digital forensics expert in this context is like a pathology technician, when a hacker makes a mess, the forensics expert comes in and looks at the situation from the beginning, reads the log files like fingerprints. Accordingly, it is necessary to mention the cybersecurity administrator who is not just a "gatekeeper", he already knows where the keys are hidden and how to quickly take them away from the attacker. When such parties meet, then it is no longer just a clash of knowledge, but a dance of roles. In some cases, the administrator becomes a hunter who does not shoot, but sets traps (honeypots). From this aspect, digital forensics plays the role of an archivist and judge, i.e. records either defeat or victory, while the administrator acts as a battlefield that the hacker must bypass. It is an eternal struggle in which every move quickly changes the rules of the game.

METHODS AND MATERIALS

Cowrie honeypot is an advanced system designed to simulate SSH (Secure Shell) and Telnet services, in order to attract hackers to enter the network and thus learn more about the attackers, in order to increase the level of protection. In this way, virus analysis can also be performed if a hacker releases a virus on the cowrie. The creators of this honeypot system are Upi Tamminen, who created Kippo - a tool written in the Python programming language, and it is a simulation of a real Linux system, which at first glance really works like a real system. In the background, the Twisted library for network programming is used, in order to properly simulate the above-mentioned protocols. So the basis of this system is Kippo together with the Twisted library. In addition, Dave Germiquet, who did unit tests and optimization, as well as Oliver Bilodeau and Ivan Korolev, who also made a huge contribution, also participated in the development. It is important to mention Florian Pelgrim, as well as Guilherme Borges, who contributed to keeping the code neat and contributing to Docker and Proxy solutions. [1] The implementation of the cowrie system has even been supported by Microsoft since October 4, 2024. When an attacker connects to this system, log files are recorded via json, but it is also possible to store these files in a sql database and send them to a remote server so that the administration team can analyze what exactly is happening. It is important to note that such a system is completely isolated from the main corporate network. [6] In this way, the methods and motives used by hackers can be learned, but also to potentially locate and catch hackers. As already mentioned, cowrie offers various possibilities, and one of the main features is that it is open-source, which means that the code is available via the Github platform, so it can be further improved, but it is also a free tool. There is also a commercial version introduced by Microsoft, where additional benefits can be obtained for a certain fee. Accordingly, it is important to note some additional powerful tools, such as Threat Intelligence. This provides insight into the growth trend of cyber attacks, as well as new techniques that are becoming more and more common every day. [2] By collecting information in this way, cybersecurity experts can significantly contribute to the protection of user systems. Based on collected data, with adequate documentation and quick response, it can significantly contribute to the protection of the system. Another important feature is the detection of previously unknown threats. This tool, when integrated with other security solutions, can catch new types of attacks before they attack the system. Also, companies can update their IDS (Intrusion Detection Systems) solutions, as well as firewall rules, based on the data collected from the cowrie system. [7]Another important advantage is the low operational risk associated with its implementation, because the system itself is completely isolated from the main one, and in this way it can be monitored more easily. [3]

Implementing Cowrie through VMWARE

The Cowrie honeypot will be installed on an Ubuntu virtual machine. As mentioned in the opening chapter, a ready-made virtual machine with Ubuntu configured will be downloaded. Using VMWare, the machine will be imported into the development environment, where cowrie will be configured. The image below shows how to import an Ubuntu virtual machine(Figure 1).

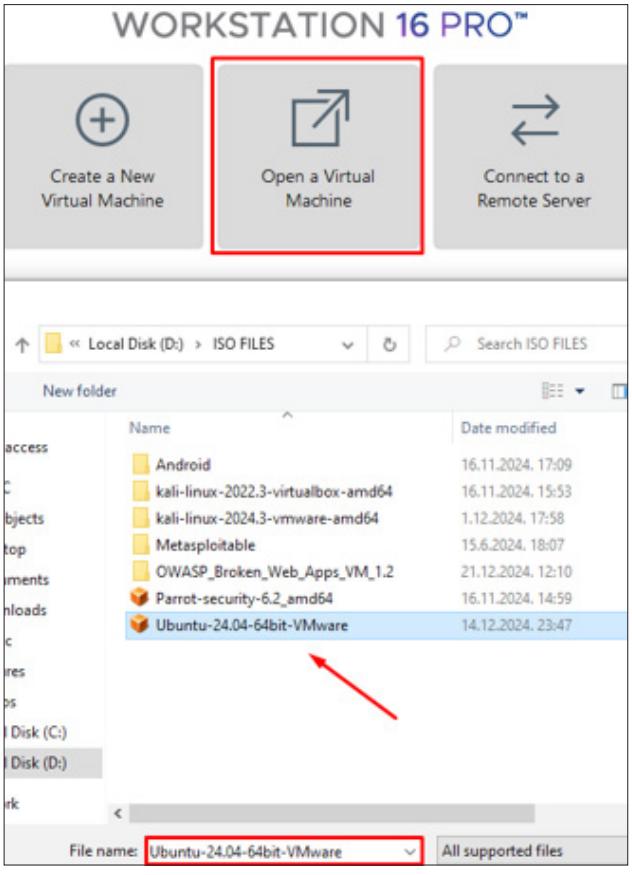


Figure 1. *Import Ubuntu virtual machine*

The name of the virtual machine within VMWare is CowrieUbuntu, for easier organization (this part is not directly visible if the machine is accessed "from the outside"). The network adapter is set to NAT network to make the virtual environment secure and isolated. Other settings such as CPU and memory are left at default, because Ubuntu can run smoothly with 2 GB of RAM and 2 VCPUs (Figure 2).

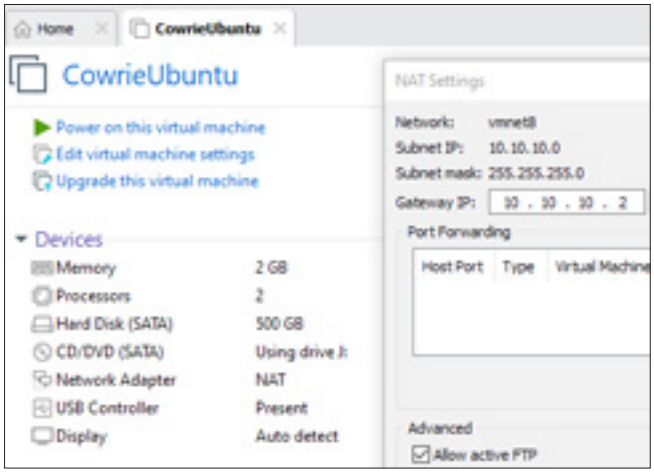


Figure 2. *Initial virtual machine setup*

Next, the virtual machine is started by clicking on "Power on this virtual machine". After updating with sudo apt update and sudo apt upgrade, it is necessary to install Git, with the command sudo apt install git, in order to be able to download cowrie from Github [4] (Figure 3)

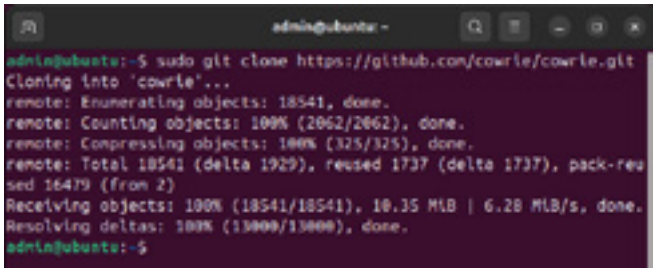


Figure 3. *Download Cowrie from Github*

The next step is to install the Python3 virtual environment. The Python virtual environment is very useful, because it provides Cowrie isolation, and thus does not affect the rest of the system. It also enables easy management and updating of the Cowrie package. Also, the security itself has been raised. The installation is done with the help of a simple command:
sudo apt install python 3.12-venv

Then, it is necessary to create a virtual environment. This is possible with the help of the following command:

python3 -m venv cowrie-env

The previously created virtual environment must then be activated. All packages installed using pip will be moved to that environment.

source cowrie-environment/bin/activate

In addition, it is necessary to install the previously mentioned pip (Python Package Installer):

sudo apt install python3-pip

Once this installation is complete, it is necessary to install everything required to run Cowrie, and the necessary installations are located within requirements.txt. [5] (Figure 4)

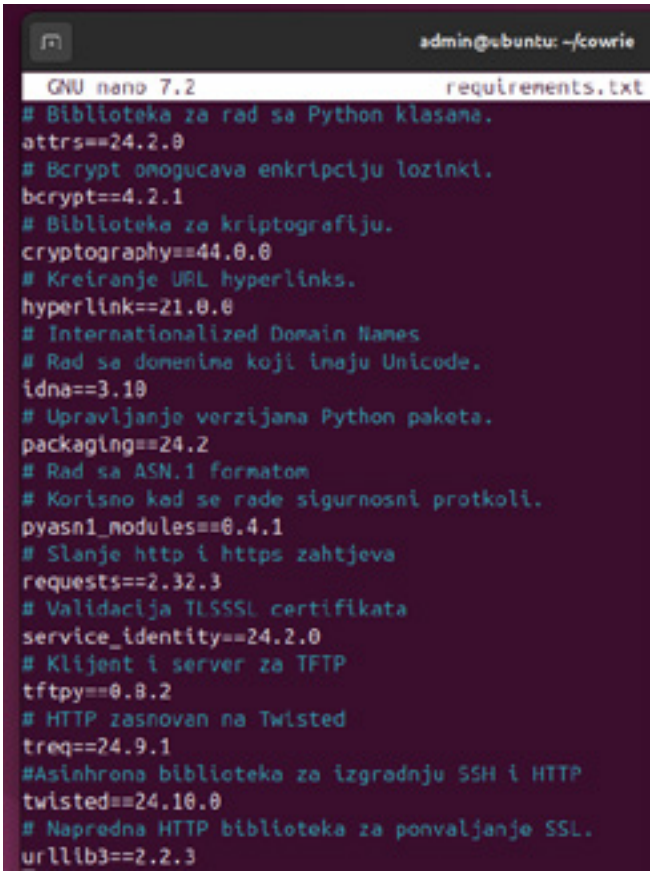


Figure 4. *The contents of the requirements.txt file*

Figure 5. *Installation of necessary tools from requirements.txt and display inside cowrie*

The Figure 5 shows the files inside cowrie. Inside honeyfs are all the necessary files that simulate a fake honeypot system. Inside etc is the main configuration file. Inside var are log files that are collected after someone tries to access the honeypot. Once this is set it is possible to start

cowrie with: *Bin/cowrie start*

Figure 6. *Running Cowrie on port 2222*

The Figure 6 shows that cowrie is properly started, and also with the help of a simple netstat command, the open port is displayed. Based on the picture, the honeypot is started, and now it is possible to monitor log files through cowrie/var/log/cowrie, and there are two types of log files stored there through .json and .log.

RESULTS

Now, as a conceptual example, a potential attack via SSH on the honeypot will be shown, using the Kali Linux machine. Kali Linux is also installed in a similar way to the Ubuntu honeypot, i.e. a ready-made virtual machine that can be downloaded online and easily implemented within VMWare Workstation. We will also take into account that the hacker has "accessed" the honeypot system, and that cowrie is located at the IP address 10.10.10.4 One of the first stages is to collect information and scan open ports, and this can be done with the help of a simple nmap tool, (Figure 7).

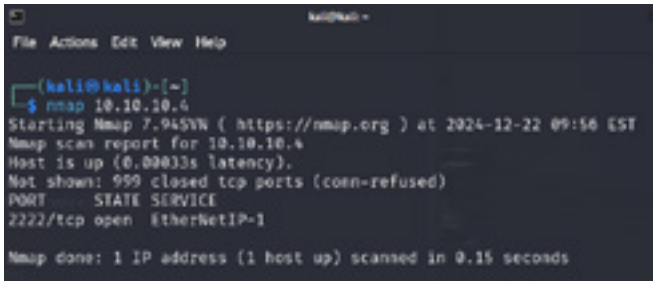


Figure 7. *Nmap scanning for open ports*

After this, the hacker will try to access using simple credentials that are already so-called well known passwords and usernames. The image below shows access to the server, which can look extremely tempting, because it really gives the impression of a real server. (Figure 8)

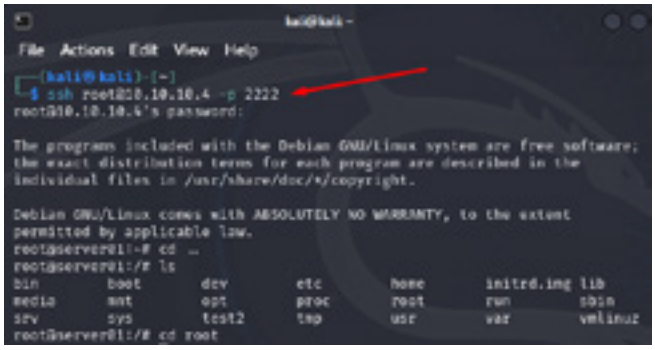


Figure 8. *Hacker access via SSH to cowrie honeypot*

The hacker is very likely not aware that he is caught in a cowrie honeypot, and that his IP address is stored inside the log file. Thus, in the log files it is possible to see the IP address, but also all the activities that the hacker tried to do, even if he tried to run a malicious script, usually of the ransomware type. (Figures 9 and 10)

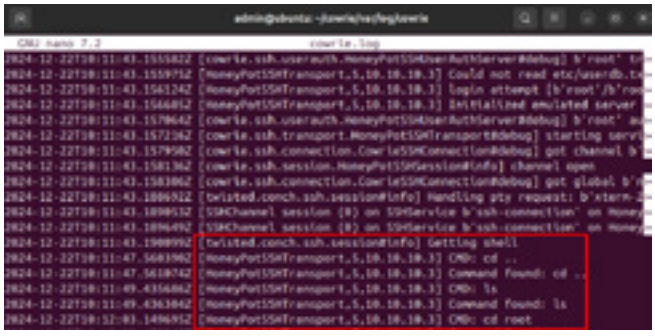


Figure 9. *View log files and all commands typed*

Figure 10. *View log files from cowrie.json file*

Given that these two types of log files exist within the cowrie directory, they are not very well struc-

tured, and are quite difficult to analyze, as can be seen in the two images above, therefore, as a contribution to this honeypot, as my contribution to this, I created a simple python script that converts the json log file into a readable output.txt file. The script and the readable log file will be shown in the Figures 11 and 12. There are various types of conversions also in the form of graphs, but the lack of an initial file, or the part that will perform the direct conversion, was noticed.

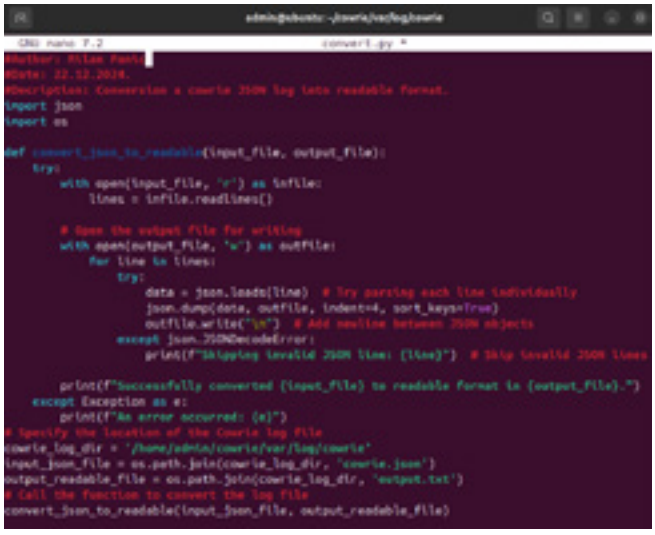


Figure 11. *Python script to convert JSON to txt*

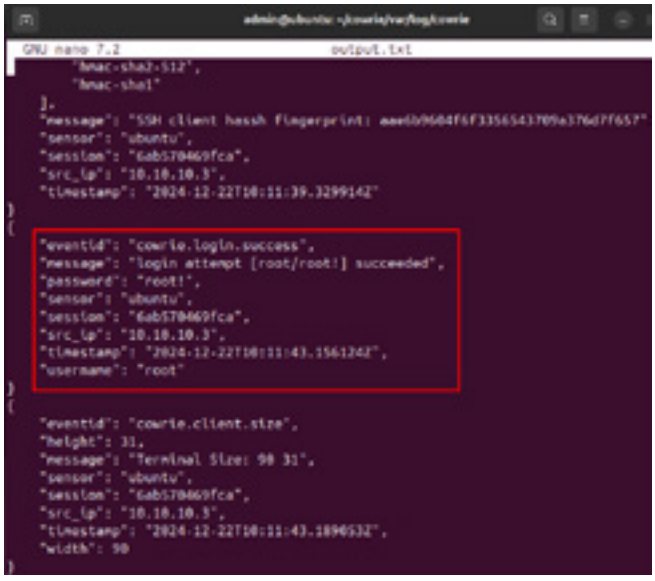


Figure 12. *Edited log file*

Discussion

The essence of the observation is a complete honeypot system, or honeynet (a system of multiple honeypots), implemented in the VMWare development environment by creating virtual machines, and can be used as an additional level of protection and defense, in order to catch and block a potential attacker in a timely manner. This research part should answer the question of whether a honeynet system can further raise the level of security and safety, in order to discover the identity of the attacker and thus disable access to the system. In addition to the functionality itself, another aspect is the material, technical and time frame that is necessary to implement such a system, whether it can also be upgraded, as well as what effect is achieved by using it. One of the main focuses is on the Cowrie honeypot system, which is designed to record Telnet and SSH connections (Figure 13). It also offers session recording capabilities. This type of honeypot is often connected to the Internet, in order to monitor additional tools, hosts and scripts that attackers use when trying to guess the password to access the system. The SSH attacker will conceptually connect to port 22, and will be redirected to our honeypot towards port 2222, as shown in the previous part of the paper.

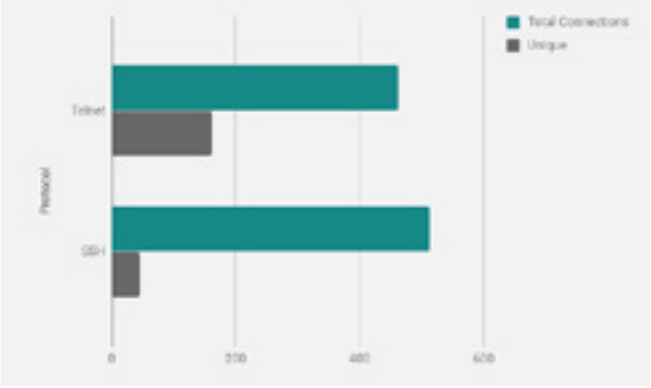


Figure 13. *Total number of attack attempts*
Source: *https://hackertarget.com/cowrie-honeypot-ubuntu/*

Conclusion

This paper explains the Cowrie honeypot system in detail, and demonstrates its functionality and effectiveness in real systems. The main goal was to understand the strategy of this interesting technique, in order to draw attackers' attention to the fake system. One limitation of this research is that the implemented Cowrie honeypot system was tested exclusively within an isolated VMware NAT environment. Although such an approach significantly improves security and prevents

unintended exposure of the host infrastructure, the obtained results may differ from those observed in publicly exposed production systems connected directly to the Internet. Real-world environments may include more complex attack patterns, larger traffic volumes, and interactions with additional security mechanisms such as IDS/IPS platforms and enterprise firewalls. Companies use honeypot systems to protect the network and servers of the entire organization, and the research part conducts academic experiments on the same at universities and schools. As is already known, with the rapid growth of artificial intelligence and cyber attacks, information security is very important for all systems, because any network that is not protected in the network can easily be compromised at any time. Based on this informations, important company data can be lost, which can represent a huge loss, and it can also be very dangerous for someone else to have access to our important personal data. Therefore, this paper provides insight into the security aspects of modern information systems, and one of the main goals was to examine whether honeypot systems are easy to hack, and to check whether they are truly isolated from the rest of the network. It is evident that traditional methods in the field of cybersecurity are increasingly giving way to modern and proactive solutions that rely on concepts such as honeypot systems. In this paper, the focus was on a specific type of honeypot model that was developed with the intention of identifying and protecting potential targets of high-tech crime. Given the rapid development of information technologies, especially in the field of artificial intelligence and adaptive security solutions, it is expected that such systems will become increasingly advanced, efficient.

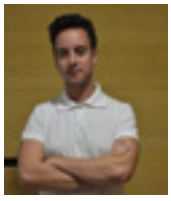
Acknowledgements
This study did not receive any external funding. The authors are fully responsible for the content of this article.

REFERENCES

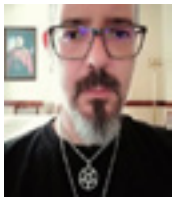
[1] A. Sardana, Honeypots - a new paradigm to information security, Science Enfield USA, 2012.
[2] S. Holder, Honeypots for Windowds, Dublin: Apress, 2014.
[3] M. Nawrocki, M. Wahlisch, T. C. Schmidt and J. Schonfelder, "A Survey on Honeypot Software and Data Analysis," *ACM Computing Surveys,* vol. 56, no. 2, 2023.
[4] O. M. Youssef and A. Almulhem, "Advanced Honeypot Architectures for Modern Cybersecurity," *IEEE Access,* vol. 12, pp. 11245-11260, 2024.
[5] "https://github.com/cowrie/cowrie," [Online]. [Accessed 25 2 2026].
[6] "https://hackertarget.com/cowrie-honeypot-ubuntu/," [Online]. [Accessed 20 2 2026].
[7] H. Jahankhani, Cyber Criminology, London: Springer Nature, 2018.

Received: March 17, 2026
Accepted: April 30, 2026

ABOUT THE AUTHORS



Milan Panić was born in Gradiška in 1999. He earned the title of Bachelor of Computer Science and Informatics Engineer in the field of computer security and information protection in July 2023 at the Pan-European University "Apeiron" Banja Luka as a top student with an average of 9.48. He completed his Master's studies in December 2025 with a grade point average of 10.00. Since 2021, he has been employed at the Pan-European University Apeiron in Banja Luka as a system administrator, and since 2024 as an assistant in the courses of computer forensics and mobile forensic. He holds a certificate from ISC2 (CC), which confirms his understanding of key principles in the field of cybersecurity, including network security, data protection and risk management.



Nemanja D. Maček is a senior lecturer at the School of Electrical and Computer Engineering and the full professor at the Faculty of Social Sciences. He also works, as a contractor, for the SECIT Security Consulting and as a senior AI researcher and scientific advisor for several companies. Nemanja graduated from University of Novi Sad in 2006 and received Advanced Security Systems PhD from Singidunum University in 2013. The research of Prof. Maček involves machine learning, pattern recognition and natu- ral language processing applied to information security, as well as biometric systems, cryptology and designing novel security mechanisms.

FOR CITATION

Milan Panić, Nemanja Maček, Implementation of Cowrie Honeypot System and Improvement of Log Analysis, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 16(2026)1:69-74, (UDC: 37.016::003-028.31), (DOI: 10.7251/JIT2601069P), Volume 16, Number 1, Banja Luka, June (1-76), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

JITA

Journal of Information Technology and Applications
Pan-European University APEIRON Banja Luka
https://www.jita-au.com
jita@apeiron-edu.eu

PUBLISHER: **Pan-European University APEIRON**, Banja Luka
College of Information Technology Banja Luka, Republic of Srpska, BiH
www.apeiron-uni.eu
Darko Uremović, Person Responsible for the Publisher
Aleksandra Vidović, PhD, Editor of University Publications

ISSN 2232-9625