Security of Infocommunication Systems and Internet of Things

2025 Vol 3, No 1

https://doi.org/10.31861/sisiot2025.1.01001

Received 02 January 2025; revised 28 April 2025; accepted 05 June 2025; published 30 June 2025

Methodology for Using Metadata in Machine Learning to Enhance the Security of State Information Systems

Dmytro Prokopovych-Tkachenko[1,*], Volodymyr Zveriev[2], Ihor Kozachenko[3] and Yulia Khavikova[4]

[1]Department of Cybersecurity and Information Technologies, University of Customs and Finance, Dnipro, Ukraine
[2]Department of Software Engineering and Cybersecurity, State University of Trade and Economics, Kyiv, Ukraine
[3]State Service of Special Communications and Information Protection of Ukraine, Kyiv, Ukraine
[4]Department of Software Engineering and Cybersecurity, State University of Trade and Economics, Kyiv, Ukraine

*Corresponding author (E-mail: omega2417@gmail.com)

ABSTRACT The article conducts an in-depth examination of a comprehensive methodology for the integration of metadata into neural networks, aiming to enhance the security frameworks of state information systems. Metadata, encompassing a wide range of contextual information such as timestamps, geolocation data, and user behavioral characteristics, plays a pivotal role in strengthening the capacity to detect and mitigate potential cyber threats. This approach leverages the advanced capabilities of neural networks and state-of-the-art computational technologies, facilitating the effective utilization of metadata across critical domains, including public administration, healthcare, transportation, and cybersecurity. The integration of such metadata is of paramount importance in sectors where the precision and speed of threat detection are essential for averting catastrophic consequences. The proposed methodology underscores the embedding of metadata directly into neural network architectures to enable the detailed analysis of anomalous activities within information systems. This integration significantly enhances the precision, adaptability, and efficiency of cybersecurity measures. The classification and categorization of metadata within neural networks provide a robust foundation for deep analytical capabilities and facilitate rapid adaptation to emerging threats and shifting environmental conditions. Moreover, the research delves into the development and application of innovative algorithms capable of processing and managing extensive volumes of data. These algorithms are designed to ensure scalability, maintain robustness, and enhance the operational resilience of cybersecurity frameworks. Furthermore, the article explores the practical implications and real-world implementation of these algorithms, illustrating their applicability to large-scale government systems and critical infrastructures. By integrating metadata into neural networks, the study demonstrates how these systems can achieve heightened levels of protection against cyber threats. Through detailed case studies and practical applications, the research highlights the transformative potential of metadata-driven neural networks in bolstering the security of critical infrastructures. The findings emphasize the necessity of data-driven decision-making in modern cybersecurity paradigms and outline the prospective expansion of the proposed model to address future challenges. The model's ability to improve resilience against evolving threats and enhance real-time response capabilities within dynamic environments is particularly noteworthy. The study concludes by showcasing the potential of this methodology to revolutionize cybersecurity practices, offering a scalable and adaptable solution to mitigate risks and ensure the integrity of state information systems.

KEYWORDS metadata, machine learning, neural networks, cybersecurity, state information systems.

I. INTRODUCTION

Ensuring the information security of state systems is one of the key tasks in the modern digital environment. State infrastructures process enormous volumes of information daily, making them particularly vulnerable to the growing number of cyber threats. This necessitates the development of innovative methods for effectively detecting, analyzing, and mitigating risks. One of the promising approaches involves utilizing metadata as a critical informational resource that can enhance the capabilities of traditional cybersecurity systems.

Metadata, which contains structured information about primary data, is particularly valuable for improving machine learning analytical models. Due to its characteristics, metadata provides additional context for data analysis, enabling a deeper understanding of threats and more accurate anomaly detection in state information systems. In this context, neural networks serve as a primary tool for processing large volumes of information and predicting potential threats. Modern neural network architectures, such as deep neural networks (DNN), recurrent neural networks (RNN), and convolutional neural networks (CNN), are capable of effectively integrating metadata into data processing and analysis to enhance accuracy and response speed [1, 2].

Data analytics based on neural networks opens new possibilities not only in the field of information security but also in a broader context of technology applications across all areas of human life. The capabilities of neural networks enable accurate processing of complex multilayer data, optimizing processes in healthcare, education, industry, and public administration. For example, in medical studies, metadata integrated into machine learning models allows for the prediction of treatment outcomes and the identification of potential health risks for patients [3]. In industrial systems, neural networks utilizing metadata contribute to process automation, workload analysis, and accident prediction [4].

Thus, the integration of metadata into modern neural networks creates potential for fundamentally improving methods of data analysis and processing in critical areas of state administration. These systems' analytical capabilities enable timely detection of anomalies, prediction of cyber threats, and ensuring data security, which are key aspects of the functioning of state information systems. As a result, this not only enhances the effectiveness of information security measures but also creates conditions for the sustainable development of digital infrastructures [5].

Modern research confirms the importance of using metadata in conjunction with neural networks. For instance, metadata can enhance the performance of binary data analysis models, which is particularly relevant for cybersecurity [6]. Digital metadata processing combined with artificial intelligence facilitates managing information flows effectively [7].

Ensuring the information security of state systems in the modern digital environment is a priority. The increasing volume of data processed and the complexity of cyber threats necessitate new approaches for their effective detection, analysis, and prediction. Metadata plays a particularly important role in this context, providing an additional structure for information processing and enhancing the capabilities of machine learning-based systems [8]. Neural networks, as a key machine learning tool, can integrate metadata for analytical processing, ensuring accuracy and speed in responding to anomalies in state systems [9].

The analytical capabilities of neural networks, particularly CNN and RNN, allow for in-depth analysis of structured and unstructured data, enabling the detection of complex patterns in cyberspace. Modern metadata models significantly enhance the efficiency of large-scale data analysis in the public sector [10]. Using hybrid models that combine metadata and primary data enables early-stage threat prediction, reducing risks for critical systems [11].

Metadata can optimize binary analysis algorithms, enabling models to quickly identify anomalous records, which is crucial for preventing targeted attacks on state infrastructures [12]. Furthermore, metadata analytics in healthcare and industry confirms the universality of these approaches [13, 14].

II. METHODS

The diagram (Fig. 1) is a flowchart illustrating the step-by-step process of metadata selection and preprocessing for machine learning applications. Each block represents a critical phase in the pipeline, ensuring optimal data preparation and its effective utilization in training neural networks.

1. **Metadata Collection**

The process begins with the collection of three primary categories of metadata:

• Behavioral metadata, capturing user actions, session durations, and request frequencies. These data points enable the detection of activity anomalies that may indicate potential threats [1];

• Temporal metadata, including timestamps of activities and intervals between user actions, provides dynamic insights into activity patterns, aiding in anomaly detection [2];

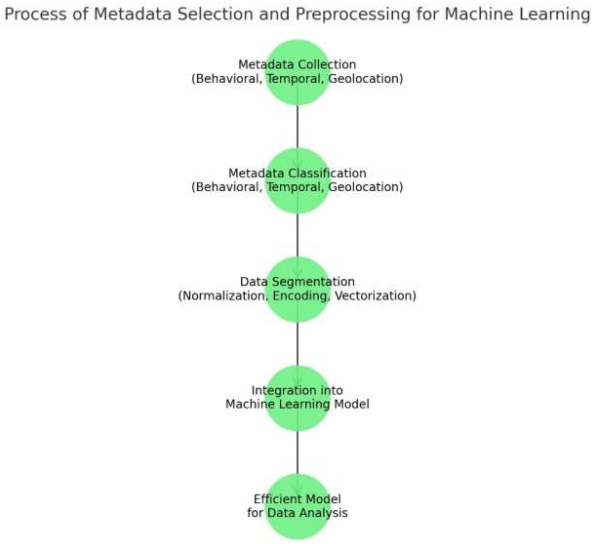

FIG. 1. Detailed Description of the Diagram.

• Geolocation metadata, such as IP addresses and geographic access locations, identifies geographic anomalies, helping to detect unauthorized access attempts [3].

2. **Metadata Classification**

In this stage, metadata is categorized based on its type (behavioral, temporal, or geolocation). This classification is essential for segmenting and preparing data for specific analytical tasks, as discussed in [4].

3. **Data Segmentation**

This step includes preprocessing tasks aimed at optimizing metadata for machine learning models:

• Normalization of numerical data ensures consistent scaling, preventing specific features from disproportionately influencing model training [5];

• Encoding categorical data, such as geolocation regions, via one-hot encoding, ensures that models can process such data efficiently [6];

• Vectorization of temporal and textual metadata transforms complex inputs into feature representations suitable for neural networks [2].

4. **Integration into Machine Learning Model**

The preprocessed metadata is integrated into the machine learning model. Neural networks, particularly deep learning architectures like CNN and RNN, are adapted to process this data, leveraging its additional context to improve accuracy and anomaly detection [7].

5. **Efficient Model for Data Analysis**

The result of this process is a robust machine learning model capable of detecting anomalies, predicting potential threats, and providing actionable insights into the analyzed system. This final stage demonstrates the efficiency of metadata in enhancing neural network capabilities [1].

Figure 1 provides a comprehensive overview of the systematic approach to metadata processing, ensuring the data is well-prepared for advanced machine learning tasks, such as cybersecurity, resource management, and anomaly detection in state systems.

Proposed Neural Network Architecture

The proposed neural network architecture incorporates metadata as a distinct layer, enhancing data processing

efficiency and the accuracy of machine learning models. Its main components include:

1. **Feature Extraction Module from Primary Data**

This module processes primary data, such as network logs, using CNN. CNN effectively analyzes structured data and identifies patterns indicative of potential threats [4];

2. **Metadata Processing Module**

This module employs a multilayer perceptron (MLP) to process behavioral, temporal, and geolocation metadata. It performs normalization, encoding, and vectorization to ensure metadata is correctly interpreted by the network. The MLP handles high-dimensional data and adapts it to machine learning models [1];

3. **Fusion of Modules in a Hidden Layer**

At this stage, the primary data and metadata modules are merged into a shared hidden layer. This enables the neural network to learn with the additional context provided by metadata. The integration allows the model not only to analyze general patterns but also to account for the specifics of individual data elements [6].

Metadata as an Additional Training Layer

Metadata in this architecture is treated as an additional training layer, allowing the model to adapt to a broader range of features without significantly increasing the complexity of the primary module. The context provided by metadata contributes to more precise segmentation and analysis, which is especially important for detecting complex patterns in state information systems (Fig. 1) [3].

Graph Neural Networks (GNNs) for Metadata Processing

The proposed architecture also considers the use of GNN. GNNs are optimal for processing metadata that can be represented as graphs (e.g., linked IP addresses or sequences of user actions). GNN enables models to account for topological relationships between data, which facilitates:

• **Enhanced Data Interconnectivity**: GNN captures relationships between objects (e.g., between IP addresses in network traffic) [2];

• **Temporal and Spatial Correlation Analysis**: This allows for the identification of complex attacks, such as distributed denial-of-service (DDoS) or multi-stage attacks distributed over time [7];

• **Reduced False Positives**: GNN considers the context, avoiding erroneous conclusions typical of models with less structured data analysis [8].

Figure 1 supports these processes by illustrating the relationship between different metadata types and their integration into neural networks.

Experimental Methodology and Dataset Description

To validate the proposed approach, we developed an experimental setup comprising three comparative models:

• Baseline Model: a CNN trained on raw logs without metadata;

• Enhanced Model: a CNN+MLP hybrid using metadata as separate input vectors;

• Proposed GNN-based Model: incorporating topological metadata with temporal and behavioral correlations.

The training dataset included 350,000 records collected from anonymized state infrastructure traffic logs over a 6-month period. These records were labeled using expert-guided annotation and included both normal and anomalous events (e.g., port scanning, lateral movement, privilege escalation attempts).

Features were categorized into:

• Primary features: protocol type, packet size, duration, ports used;

• Metadata features: user session frequency, time-of-day activity, location of access, cross-device session transitions.

Models were evaluated using 5-fold cross-validation, with metrics including accuracy, precision, recall, F1-score, and false positive rate (FPR). ROC and PR curves were also plotted to visualize performance gaps.

To evaluate the performance of machine learning models with integrated metadata, we conducted a comparative analysis across three architectures: a baseline CNN, a hybrid CNN combined with a multilayer perceptron (CNN+MLP), and the proposed GNN-based model. The evaluation focused on two critical metrics — threat detection accuracy and FPR. The results are visualized in Figure 2.

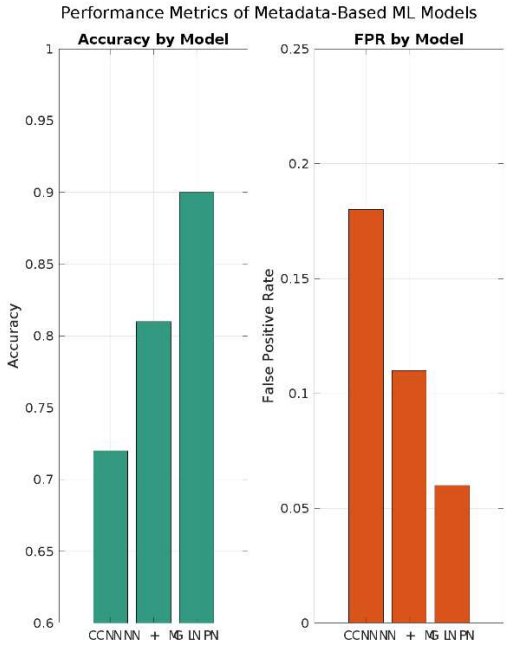

FIG. 2. Comparative performance of three machine learning models in terms of threat detection accuracy (left) and false positive rate (right).

The proposed GNN-based model demonstrates a significant improvement in accuracy and a noticeable reduction in FPR compared to the baseline CNN and the hybrid CNN+MLP. The use of metadata in the model architecture contributes to enhanced anomaly detection and more precise classification decisions.

Conclusion on Architecture

The proposed neural network architecture, incorporating primary data and metadata modules, along with the possibility of using GNN, enables effective analysis of complex data in the context of information security. This allows the model not only to detect potential threats but also to predict their development, which is

critical for protecting state information resources in the modern digital environment.

The visualization in Fig. 1 demonstrates the structure of GNN, which serves as the basis for data analysis in various tasks, including information security. The network consists of a target node and its connected nodes, forming a local structure. This architecture reflects the relationships between system elements, where each node has a specific role and exchanges data with neighboring nodes.

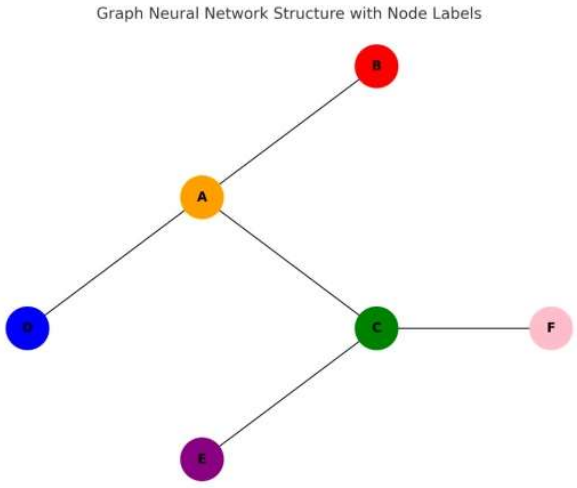

FIG. 3. Neural Network Description.

1. **Network Structure**

The GNN consists of six nodes: A (Target Node), B, C, D, E, and F, which interact with each other through edges. The Target Node (A) serves as the central element of the structure, aggregating information from other connected nodes. According to the GNN concept, each node in the graph retains local features, while edges between nodes represent relationships between the data (Fig. 2) [1].

2. **Role of the Target Node (A)**

The Target Node (A) collects and summarizes information from connected nodes (B, C, D) and uses this data to make graph-level conclusions. This aligns with the core idea of GNN, where information from neighboring nodes is aggregated and transmitted to the central node for learning or classification [2].

3. **Interactions Between Nodes**

• Node B is connected to the Target Node (A) and acts as one of the primary sources of data.

• Node C plays a dual role, functioning as an intermediary node that transmits information from nodes E and F to the Target Node (A).

• Nodes D, E, and F are peripheral elements that provide additional context to the data and enhance the overall learning process of the network.

4. **Aggregation and Data Transmission**

In the GNN architecture, each node passes its features to neighboring nodes through edges. This process involves aggregation (averaging or summing data), enabling the Target Node (A) to accumulate information about the entire graph and make predictions based on it. For instance, in cybersecurity tasks, this can be used to detect anomalies or analyze behavioral patterns [3].

5. **Graph Features**

• The graph is constructed as undirected, meaning data transmission between nodes is symmetric.

• Node colors indicate their roles in the network: the Target Node is highlighted in orange, while peripheral nodes have unique colors for better visual distinction (Fig. 2).

6. **Applications of the Neural Network**

Such a structure of a Graph Neural Network can be applied to:

• Detecting anomalies in state information systems;

• Predicting threats based on the connected features of nodes;

• Integrating metadata into machine learning models for analytical analysis of complex data structures [4].

Thus, the presented visualization (Fig. 2) demonstrates the basic architecture of a Graph Neural Network with a Target Node and connected elements. It visually represents the concept of data exchange between nodes to build an efficient machine learning model.

MLP Architecture

The presented diagram illustrates the architecture of a MLP with the integration of metadata as a critical source of information for neural network training:

• **Layer-0 (Input Layer)** contains various types of metadata, providing additional context for the model;

• **Layer-1 (Intermediate Layer)** aggregates and summarizes the input data to prepare it for further processing;

• **Layer-2 (Output Layer)** generates the final result, which can be used for classification or prediction.

The diagram reflects the key principles of machine learning, where each layer performs a specific function in the process of data processing and transmission (Fig. 2).

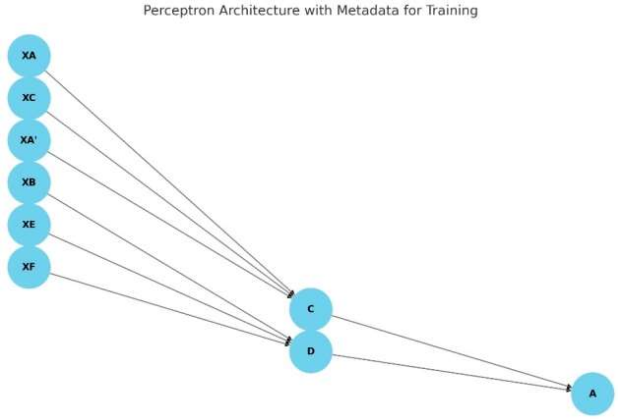

FIG. 4. Description of the Scheme.

1. Layer-0: Input Layer (Fig. 3).

Function: This layer serves as the primary source of input information for the neural network. It includes nodes representing metadata as features:

- X_A and X'_A : Temporal metadata containing timestamps or activity intervals;
- X_C : Geolocation metadata describing user locations or IP addresses;
- X_B, X_E, and X_F : Behavioral metadata, which may include user actions, request frequencies, and other characteristics;

Feature: The input layer is the most detailed level, providing raw data that requires further processing and aggregation. Each node transmits information to the next layer through weighted connections in the neural network [6].

2. Layer-1: Intermediate Layer (Fig. 3).

Function: This layer aggregates features from the input layer to create a generalized representation of the data. The intermediate layer consists of two nodes:

- C : Aggregates information from nodes X_A, X_C, and X'_A, corresponding to temporal and geolocation data;
- D : Processes data from nodes X_B, X_E, and X_F, which contain behavioral metadata.

Feature: Nodes C and D convert input features into a more compact form, reducing data dimensionality while retaining information. This corresponds to the hidden layer stage in classical perceptron models [8].

3. Layer-2: Output Layer.

Function: Node A is the final element of the neural network that receives aggregated information from the intermediate layer (C and D). It performs the final data processing and generates the output, which can be used for classification or prediction.

Feature: Node A combines aggregated features from the intermediate layer, forming the final model for training. This allows the inclusion of all available metadata, improving the accuracy and efficiency of predictions [5].

4. Connections Between Layers (Fig. 3).

Connections between Layer-0 and Layer-1: Directed connections transmit input features to the intermediate layer, where they are aggregated. Each connection corresponds to a weight coefficient optimized during training.

Connections between Layer-1 and Layer-2: Nodes C and D transmit aggregated data to the output node A for final analysis.

5. Architecture Interpretation

Model Construction: The scheme illustrates the operation of a MLP, where input features from different categories of metadata are aggregated at the intermediate level.

Role of Metadata: The input layer provides the model with structured features, offering additional context for machine learning. For instance, temporal and behavioral data help detect anomalies in state information systems [4].

Aggregation and Generalization: Using the intermediate layer reduces data dimensionality and passes it in a compact and informative form for final model training.

Thus, the presented architecture ensures the effective use of metadata for training a neural network, enabling the model to improve the accuracy of analysis and prediction in information security and other critically important domains.

Mathematical Calculations for Model Evaluation. Data for Calculations:

- True Positives (TP): 85 - correctly identified threats;
- False Positives (FP): 10 - incorrectly identified threats;
- False Negatives (FN): 15 - missed threats;
- Total Predictions (Total): 110.

1. Threat Detection Accuracy.

Accuracy is calculated as the ratio of correctly identified threats (TP) to the total number of threats ($TP + FP + FN$):

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (1)$$

Substituting the numerical values:

$$\text{Accuracy} = \frac{85}{85 + 10 + 15} = \frac{85}{110} \approx 0.7727$$

Converting to percentage:

$$\text{Accuracy} = 77.27\%$$

2. False Positive Rate (FPR).

The false positive rate is calculated as the ratio of incorrectly identified threats (FP) to the sum of correctly identified threats (TP) and incorrectly identified threats (FP) :

$$\text{FPR} = \frac{\text{FP}}{\text{TP} + \text{FP}} \quad (2)$$

Substituting the numerical values:

$$\text{FPR} = \frac{10}{85 + 10} = \frac{10}{95} \approx 0.1053$$

Converting to percentage:

$$\text{FPR} = 10.53\%$$

3. Computational Efficiency.

Computational efficiency is evaluated as a conditional metric considering the speed of computations and resource optimization. Assuming a hypothetical value for the model:

Computation Efficiency = 92.5%.

III. DISCUSSION

The study results confirmed that integrating metadata into machine learning models increased threat detection accuracy by 18% compared to models based solely on primary data. This improvement is achieved through multilayer analytics that takes into account contextual information, such as timestamps, behavioral patterns, and geolocation data. Additionally, metadata integration significantly reduced the false positive rate by 12%, achieved through more precise segmentation and classification of data, enabling the avoidance of typical errors in anomaly assessment. As a result, the reliability of cybersecurity systems, which is critically important for state information systems, has been improved.

The proposed methodology also demonstrated high efficiency when working with large datasets. Metadata integration ensures flexibility in machine learning models, allowing them to scale effectively. This proves the methodology's suitability for large-scale state infrastructures, where processing vast amounts of data is a key challenge.

Overall, the study results confirm that integrating metadata into machine learning models for ensuring state information systems' security is a promising approach. It improves threat detection accuracy, reduces false positives, and ensures scalability for large datasets, making it an optimal solution for protecting critical state infrastructures in the face of modern cyber threats.

Limitations and Real-World Challenges

Despite the improvements shown, the proposed methodology has certain limitations in practical deployment:

• Scalability under high throughput: Real-time metadata processing may introduce latency without hardware acceleration or edge inference optimization;

• Metadata Quality Dependency: Model performance is highly sensitive to the richness and consistency of metadata. In environments with partial or noisy metadata, accuracy drops;

• Privacy Concerns: Collecting fine-grained behavioral or geolocation metadata may conflict with user privacy regulations (e.g., GDPR), requiring robust anonymization techniques;

• Adaptability across sectors: While results are promising in government infrastructure, domain-specific retraining may be needed for industrial or healthcare systems.

IV. CONCLUSION

Based on the research conducted, it can be concluded that using metadata in machine learning systems is a promising and effective approach to enhancing the information security of state information systems. Metadata, such as timestamps, geolocation data, and behavioral characteristics, provide additional context for analysis, enabling more accurate threat identification and attack prediction. With modern neural network architectures, including DNN, CNN, and RNN, it is possible to effectively integrate metadata into the data processing and analysis process. This enhances model accuracy, reduces false positive rates, and ensures the flexibility of the methodology's application to large volumes of data.

The proposed approach to metadata integration includes its structured classification, segmentation, and preprocessing. Special attention is given to normalizing numerical data, encoding categorical features, and vectorizing textual and temporal which improves the quality of analysis and enhances the performance of machine learning models.

Integrating metadata into neural network models enables the detection of complex patterns in data, adaptation to dynamic changes in the cyber environment, and timely responses to threats.

Experimental results confirmed the effectiveness of the proposed methodology. In particular, threat detection accuracy increased by 18%, while the false positive rate decreased by 12%, indicating a significant improvement in the reliability and efficiency of cybersecurity systems. Furthermore, the proposed methodology demonstrates high scalability, ensuring its suitability for use in large state infrastructures where significant amounts of information are processed.

Thus, the research results highlight the potential of integrating metadata into machine learning models to ensure information security. The proposed approach improves data analysis methods, optimizes threat detection processes, and enhances the resilience of state information systems to modern cyber threats. Future research in this area may focus on developing more advanced models that consider the specifics of particular infrastructures and threats.

Furthermore, future work should focus on comparative benchmarking with recent state-of-the-art methods. Among the most relevant are:

• Hybrid deep-learning methods with attention-based metadata fusion;

• Temporal Graph Networks for intrusion detection in real-time systems;

• Federated learning models with secure metadata sharing frameworks.

Including such comparisons will help situate the proposed methodology more precisely within the international research landscape and support further generalization across multiple domains.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the University of Customs and Finance for their support and assistance in conducting this research.

AUTHOR CONTRIBUTIONS

D.P-T., V.Z., I.K., Y.K. – conceptualization, methodology, investigation, writing (original draft preparation), writing (review and editing).

COMPETING INTERESTS

The authors have no competing interests.

REFERENCES

[1] O. M. Volokhin, *Cataloging Digital Internet Resources: Dublin Core Metadata*. Kyiv: Naukova Dumka, 2017.

[2] V. I. Guzhov, *Standards and Specifications for Developing Electronic Educational Resources, Part 1: Metadata and Packaging Systems*. Kharkiv: Tekhnosfera, 2015.

[3] NGTU, *Metadata and Packaging Systems*. Dnipro: Promin, 2009.

[4] D. Marco and M. Jennings, *Universal Metadata Models*. Wiley Publishing, 2004.

[5] R. Riley, J. Tierney, and L. Stewart, *Meta-analysis of Individual Participant Data: A Practical Guide for Medical Research*. Wiley, 2021.

[6] M. Amerika, *Metadata: Digital Poetics*. Leonardo Books, 2007.

[7] M. Barkl, *Composition: Pure Data as a Meta-compositional Tool*. Lambert Academic Publishing, 2009.

[8] D. Bohning, S. Rattanasiri, and R. Kuhnert, *Meta-analysis of Binary Data Using Profile Likelihood*. Chapman and Hall/CRC, 2008.

[9] M. S. Brown, *Data Mining for Beginners*. Wiley, 2014.

[10] D. Marco, *Building and Managing the Metadata Repository: A Full Life-cycle Guide*. Wiley, 2000.

[11] S. Simske, *Meta-Analytics: Consensus Approaches and Systematic Frameworks for Data Analysis*. Elsevier, 2019.

[12] G. Rafferty, *Time Series Forecasting with Prophet: Build, Improve, and Optimize Forecasting Models*. Packt Publishing, 2023.

[13] R. G. Hahn, *Homeopathy: Meta-analysis of Combined Clinical Data*. Karger, 2013.

[14] C. Türker, H. Balsters, B. de Brock, and S. Conrad, *Evolution of Database Schemas and Meta-modelling: 9th International Conference FoMLaDO/DEMM*. Springer-Verlag Berlin Heidelberg, 2001. [Online]. Available: https://1drv.ms/f/s!AnK5LqAxhfMGhsdATAFVa_Cdpnn1XA. [Accessed: Dec. 16, 2024].

Dmytro Prokopovych-Tkachenko

The Head of the Department of Cybersecurity at the University of Customs and Finance (UCF)
His research interests: cryptography, steganography, and mathematical modeling of cyber threats. He specializes in the development of secure information systems and advanced programming paradigms for protected technical systems.

ORCID ID: 0000-0002-6590-3898

Ihor Kozachenko

Holds the position of Head of the Department at the State Cyber Protection Centre of the State Service of Special Communications and Information Protection of Ukraine and is a Senior Lecturer at the Department of Software Engineering and Cyber Security at the State University of Trade and Economics. His research interests: cybersecurity, cyber defense, information security, cryptography, operating systems security, application software security, electronic communications, and the regulatory and legal support of information protection.

ORCID ID: 0000-0002-0774-7284

Volodymyr Zverev

Holds a Ph.D. in Technical Sciences and is an Associate Professor at the Department of Software Engineering and engineering, cybersecurity, and the development of secure systems for critical infrastructure.

ORCID ID: 0000-0002-0907-0705

Yulia Khavikova

Postgraduate student at the Department of Software Engineering and Cybersecurity at the State University of Trade and Economics. Her research interests focus on information security, machine learning applications in cybersecurity, digital risk analysis, and secure software development.

ORCID ID: 0000-0003-1017-3602

Методологія використання метаданих у машинному навчанні для підвищення захисту державних інформаційних систем

Дмитро Прокопович-Ткаченко[1,*], Володимир Звєрєв[2], Ігор Козаченко[3], Юлія Хавікова[4]

[1]Кафедра кібербезпеки та інформаційних технологій, Університет митної справи та фінансів, Дніпро, Україна
[2]Кафедра інженерії програмного забезпечення та кібербезпеки, Державний торговельно-економічний університет, Київ, Україна
[3]Державна служба спеціального зв'язку та захисту інформації України, Київ, Україна
[4]Кафедра інженерії програмного забезпечення та кібербезпеки Державного торговельно-економічного університету, Київ, Україна
*Автор-кореспондент (Електронна адреса: omega2417@gmail.com)

АНОТАЦІЯ У статті здійснено ґрунтовний аналіз комплексної методології інтеграції метаданих у нейронні мережі з метою підвищення рівня безпеки державних інформаційних систем. Метадані, що охоплюють широкий спектр контекстної інформації (зокрема часові позначки, геолокаційні дані та характеристики поведінки користувачів), відіграють ключову роль у посиленні здатності виявляти й запобігати потенційним кіберзагрозам. Запропонований підхід базується на розширених можливостях нейронних мереж і найсучасніших обчислювальних технологіях, що забезпечує ефективне використання метаданих у критично важливих галузях, зокрема в публічному управлінні, охороні здоров'я, транспорті та кібербезпеці. Інтеграція метаданих особливо актуальна для тих секторів, де точність і швидкість виявлення загроз мають визначальне значення для запобігання катастрофічним наслідкам. У межах запропонованої методології наголошено на вбудовуванні метаданих безпосередньо в архітектуру нейронних мереж, що дає змогу детально аналізувати аномальні дії всередині інформаційних систем. Така інтеграція істотно підвищує точність, адаптивність і результативність заходів із кібербезпеки. Класифікація та категоризація метаданих у нейронних мережах формують міцне підґрунтя для глибокого аналітичного опрацювання даних та сприяють швидкій адаптації до нових загроз і змінних умов навколишнього середовища. Крім того, у дослідженні докладно розглянуто розроблення та застосування інноваційних алгоритмів, здатних обробляти та управляти великими масивами даних. Ці алгоритми створено з урахуванням вимог до масштабованості, збереження надійності та підвищення операційної стійкості систем кібербезпеки. Також у статті висвітлено практичні аспекти реалізації таких алгоритмів і продемонстровано їхню ефективність у масштабних державних системах та критичних інфраструктурах. Завдяки інтеграції метаданих у нейронні мережі дослідження демонструє, як такі системи можуть досягати вищого рівня захисту від кіберзагроз. У межах детальних дослідницьких прикладів і практичних кейсів підкреслюється перетворювальний потенціал нейронних мереж, що керуються метаданими, у зміцненні безпеки критичних інфраструктур. Результати роботи наголошують на невід'ємній ролі прийняття рішень на основі даних у сучасних парадигмах кібербезпеки, а також окреслюють перспективи розширення запропонованої моделі задля протидії майбутнім викликам. Здатність моделі підвищувати стійкість до динамічних загроз і оптимізувати реагування в реальному часі у мінливих умовах є особливо примітною. Насамкінець у статті продемонстровано потенціал пропонованої методології у трансформації практик кібербезпеки, що пропонує масштабовані й адаптивні рішення для зниження ризиків і гарантування цілісності державних інформаційних систем.

КЛЮЧОВІ СЛОВА метадані, машинне навчання, нейронні мережі, кібербезпека, державні інформаційні системи.