Development of an Intelligent Oil Field Management System based on Digital Twin and Machine Learning

NURDAULET TASMURZAYEV*, BIBARS AMANGELDY, YEDIL NURAKHOV, SHONA SHINASSYLOV, SAMSON DAWIT BEKELE
Joldasbekov Institute of Mechanics and Engineering,
Building 2, Shevchenko Street, Medeusky District, Almaty,
KAZAKHSTAN

*Corresponding Author

Abstract: - This article introduces an innovative approach to oil field management using digital twin technology and machine learning. A detailed experimental setup was designed using oil displacement techniques, equipped with sensors, actuators, flow meters, and solenoid valves. The experiments focused on displacing oil using water, polymer, and oil, from which valuable data was gathered. This data was pivotal in crafting a digital twin model of the oil field. Utilizing the digital twin, ML algorithms were trained to predict oil production rates, detect potential equipment malfunctions, and prevent operational issues. Our findings highlight a notable 10-15% improvement in oil production efficiency, underscoring the transformative potential of merging DT and ML in the petroleum industry.

Key-Words: - Digital Twin, Machine learning, Artificial Neural Networks, Oil displacement, Intelligent Control System, Industrial Internet of Things, Sensor Network, SCADA, Energy Efficiency.

Received: March 9, 2023. Revised: October 5, 2023. Accepted: November 23, 2023. Published: December 31, 2023.

1 Introduction

The global petroleum industry finds itself at a crossroads, facing a variety of difficulties including depleting reserves, surging operational costs, and escalating environmental concerns. Innovative solutions must be developed to improve operational effectiveness and advance sustainable practices in light of these challenges. This paper introduces a groundbreaking Intelligent Oil Field Management System, harnessing the synergistic power of Machine Learning (ML) and Digital Twin (DT) technologies.

The fusion of ML and DT technology, central to our methodology, implicitly relies on several assumptions. Firstly, the data from the oil field, crucial for ML model training, is assumed to be both reliable and accurate. This assumption is vital as ML, a subset of Artificial Intelligence, leverages extensive data to evolve. By creating a dynamic virtual representation of real-world objects and systems, DT technology, in concert with ML, opens up new possibilities for operational optimization, analysis, and monitoring in real-time, [1]. This integration, much like the one described in [2], significantly improves operational efficiency and predictive accuracy, enhancing oil recovery and simplifying drilling operations, [3], [4], [5], [6]. Similarly, our use of DT is justified as seen in [7], emphasizing the creation of dynamic virtual models for real-time operational optimization. Our approach assumes that the integration of these technologies will represent various oil field components comprehensively, from pipelines to reservoirs, facilitating informed decision-making and preventive maintenance, and that our findings are representative of broader, real-world scenarios. This underscores the scalability and applicability of our methodology in oil field management.

The use of ML, namely Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN), was crucial to our effort to improve the efficiency of oil field management. The ANN's architecture, which was based on the nervous system of humans, was made up of interconnected neurons that were dispersed throughout the input, hidden, and output layers. This setup made it possible to precisely define intricate connections between inputs and outputs, which is essential in situations where system parameters are unclear. Furthermore, a specialized variant of RNN, the Long Short-Term Memory (LSTM) network, was employed to adeptly manage time-series data, thereby allowing for precise predictions of operational parameters based on both historical and real-time data streams.

Our ML models were deployed on a single-board Raspberry PI module, chosen for its cost-effectiveness, compact size, and sufficient computational power for preliminary data processing and analysis. This configuration turned out to be the pivot for forecasting data and making decisions. An extensive data cleaning phase was paramount to ensuring the accuracy and reliability of our machine learning models. This process involved filtering out noise and inconsistencies from the training data, creating a more refined dataset for model training.

The choice of Root Mean Square Error (RMSE) and Stochastic Gradient Descent (SGD) in our project was driven by specific needs. RMSE is valuable as it emphasizes larger errors in predictions, which is crucial for our objective of minimizing inaccuracies in forecasting oil field operational parameters, [8], [9]. On the other hand, SGD was chosen for its ability to handle large datasets efficiently and converge faster, making the learning process quicker and more effective, [10], [11]. This is particularly beneficial in our setup with extensive data, where timely insights are vital for optimizing oil production operations.

There are numerous examples in the literature that demonstrate the use of ANN and RNN in the field of oil recovery and production optimization. An investigation highlighted ANN's ability to forecast CO2 storage capacity and oil recovery, shedding light on the technology's potential to manage the inherent uncertainty in oil recovery procedures, [12]. Another investigation showcased the deployment of RNN for modeling oil field production, highlighting the importance of adeptly handling substantial data for precise oil data prediction, [13]. These studies demonstrate the growing importance of ML in oil field management and, when combined with DT, open the door to a new era of operational excellence in the petroleum sector.

Our project aimed to create a smarter way to manage oil fields using DT technology and ML, targeting enhanced operational efficiency. The primary objective of this study was to revolutionize oil field management by effectively integrating these advanced technologies, thereby addressing key industry challenges like resource depletion and environmental impact. We set up a special stand based on oil displacement technology for our experiments, which is instrumental in efficiently extracting more oil from the ground. This stand was equipped with various sensors, valves, and flow meters to analyze the efficacy of different fluid injections in oil displacement. The data collected was crucial in creating a virtual model (Digital Twin) of the oil field, subsequently used to train our ML algorithms. These algorithms are intricately designed to predict oil output, identify potential equipment failures, and preemptively address operational issues, [14], [15].

Additionally, the incorporation of SCADA systems ensures ease and safety in operation, offering real-time monitoring and control. The unique contribution of our research lies in the real-time modeling capabilities of our system, achieved through the innovative fusion of ML and DT technology. This approach, combining real-time data analysis with adaptive control strategies, significantly enhances operational efficiency, as evidenced by a noticeable 10-15% increase in oil production. It marks a significant innovation in oil field management, underlining our system's impact and distinctiveness in the field. Ultimately, our research strives to set a new benchmark in sustainable and efficient oil field management, contributing to the broader goal of creating more environmentally conscious and resource-efficient practices in the industry.

2 Architecture of the System

Building a bridge between theory and practice is crucial in developing our DT and ML-based Oil Field Management System. Our architecture is carefully designed to mimic real-world oil field conditions, while also providing a controlled setting for detailed testing and data gathering. The architecture of our system is depicted in Figure 1.

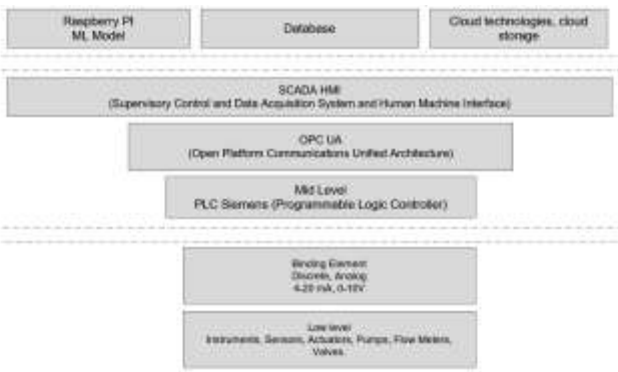

Fig. 1: Architecture of the System

The core constituents of our system are described in the following subsections.

2.1 Oil Reservoir Simulator

At the core of our experimental setup is the Oil Reservoir Simulator, which has been carefully engineered to replicate the underground conditions of an oil reservoir. It comes equipped with an array of sensors for monitoring essential parameters like pressure, temperature, and oil flow, along with electromagnetic valves and flow meters. These sensors utilize the IEEE 754, 32-bit floating-point format to convert measured field values, ensuring high precision in data acquisition.

Our experimental setup comprises terminal blocks for connecting sensors and actuators, fluid reservoirs, electromagnetic valves, connecting pipes, a logic block (which is an industrial controller), and a router for internet connectivity as seen in Figure 5 and Figure 6. This intricate setup not only enables the simulation of various oil field scenarios but also lays the foundation for data collection and analysis—crucial for the creation of a DT and training of our ML models.

2.2 SCADA Integration

The SCADA system is like the nervous system of our setup. It collects real-time data from the Oil Reservoir Simulator and other devices. SCADA is very important for keeping track of what's happening in Digital Oil Fields as mentioned in [16]. The SCADA seamlessly connects with devices such as Siemens SIMATIC S7-1200 and Siemens SIMATIC IoT 2040.

It facilitates real-time data acquisition from the Oil Reservoir Simulator and other connected apparatus, creating a dynamic representation of the experimental stand. This setup is pivotal for keeping a tab on the ongoing processes and conditions within our simulated oil field environment, thereby mirroring the potential real-world scenarios of oil fields.

2.3 Cloud Server and Database Integration

All the data from the SCADA system and sensors are sent to a cloud server. This server keeps the data safe and makes it easy to access whenever needed. Using a cloud server is a modern way to handle big amounts of data with high reliability as discussed in some studies, [17]. It manages all the computing, storage, and network resources, making sure everything is well-organized and optimized, [18]. Its SQL system ensures a structured data storage approach.

2.4 SCADA Interface (Human Machine Interface)

The SCADA interface is a window to our experimental setup, offering a real-time visualization that empowers operators and researchers to steer experiments and make judicious decisions based on visual feedback. This interface is instrumental for monitoring, controlling, and tweaking various processes in a simulated setting, thereby forming a crucial link between data acquisition and actionable insights.

Figure 2 shows the SCADA HMI for our experimental stand.

Fig. 2: SCADA Human Machine Interface

2.5 Historical Data Analysis

Preservation of all experimental data for historical analysis is an essential aspect of our system. It enables a thorough assessment of the impact of diverse operational strategies over time. By sifting through historical data, we can discern trends and patterns that are instrumental for informed future decision-making and for enhancing efficiency over prolonged operational timelines. Live data recording in MySQL database can be seen in Figure 3.

	ttype	device	tvalue	
16:28:54	5	999	23	3
16:28:49	3	999	22	3
16:28:46	2	999	22	3
15:48:15	10	999	23	3
15:48:12	9	999	23	3
15:48:11	8	999	23	3
15:48:06	6	999	22	3
15:48:05	5	999	23	3
15:48:02	4	999	23	3
15:47:59	3	999	23	3
15:27:44	12	999	19	3
15:27:43	9	999	24	3
15:27:43	11	999	336	3
15:27:42	10	999	24	3
15:27:34	7	999	23	3
15:27:34	6	999	23	3
15:27:33	5	999	23	3
15:27:26	2	999	23	3
15:17:15	12	999	18	3
15:17:14	11	999	337	3
15:17:13	10	999	24	3
15:17:07	8	999	25	3
15:17:07	7	999	23	3
15:17:06	6	999	23	3
15:17:05	5	999	24	3
15:17:04	4	999	24	3

Fig. 3: Live data registration in MySQL database

2.6 Machine Learning Methods

Our ML models learn from the collected data from our experimental stand to predict how oil will move and to fine-tune operational settings. These models use different methods to understand oil field behavior. Employing a blend of regression and classification algorithms such as random forest and XGBoost, these models delve into the integrated data to forecast oil displacement behavior, thereby aiding in experiment optimization. This segment of our architecture is the heart for harnessing data-driven insights to augment operational efficacy.

2.7 LSTM Modelling

To support data forecasting and decision-making, a single-board Raspberry PI computer is used. On this platform, we deploy our LSTM neural network and train it using data recorded in the database. The neural network accepts an input data of shape (12, 8), where 12 is the number of time steps (12 time slots) and 8 is the number of features per time step (representing the data from 8 sensors at each time slot). Following the input layer, the network features an LSTM layer with 128 units, which is particularly suited for time-series data due to its ability to remember information over long periods and to capture temporal dependencies. It is followed by a dense layer of 256 neurons, and an output layer of one neuron. Figure 4 shows the architecture of the neural network used.

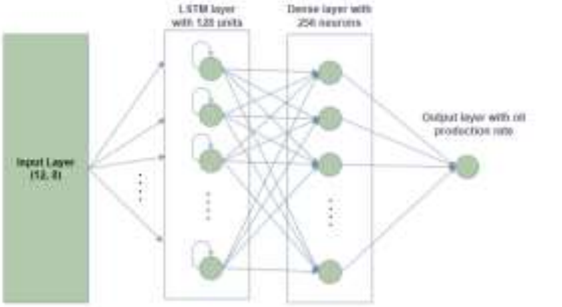

Fig. 4: Architecture of the neural network

2.8 Optimization and Intelligent Control

The architecture also embraces optimization and intelligent control mechanisms to tweak operational parameters for enhanced performance and efficiency. Intelligent control is achieved through the SCADA system which has been utilized in various research for the scientific management of oil fields, [19]. The decision support system nested within the Digital Twin furnishes real-time recommendations for optimizing oil displacement methods. It's adept at alerting operators about potential issues such as equipment malfunctions, thus proactively averting costly downtime.

2.9 Intelligent Remote Module for Real-Time Operation

The foundation of our intelligent remote-control module is laid by the sensors, which are specifically designed to capture both discrete and analog signals, typically in the range of 4-20 milliamperes and 0-10 volts.

Once these signals are collected, they are forwarded to the Programmable Logic Controller (PLC). The PLC is adept at interpreting these signals, converting both discrete and analog readings into a digital format which can then be processed further.

Upon processing, the PLC communicates with the OPC Server and Gateway. This server plays a dual role: firstly, it ensures the secure and efficient transmission of data to the cloud for storage. This cloud storage not only acts as a backup but also as a centralized data repository that can be accessed from various endpoints. Secondly, the OPC Server transmits this data to the SCADA HMI system. Here, the data is visualized, providing operators with a real-time overview of the system and allowing them to make informed decisions.

Simultaneously, the processed data is also stored in a dedicated database. From this database, the Raspberry Pi, equipped with a Machine Learning (ML) model, fetches the data for training. By training on this data, the ML model can derive patterns and insights which it then uses to send intelligent, predictive outputs back to the SCADA system. This mechanism ensures that the SCADA system isn't just a passive display but an interactive control panel that benefits from the predictive capability of the ML model.

3 Results

We undertook a range of tests using our oil displacement technology set up to gauge the efficacy of our digital twin and machine learning-focused oil field management approach. These tests simulated situations typical of real-world oil fields, especially secondary and enhanced oil extraction techniques. Our experiments yielded several notable results: The marriage of digital twins and machine learning significantly boosted oil extraction in our simulation, surpassing traditional techniques. By perpetually refining operational metrics, a greater volume of oil was mobilized from the reserve.

Furthermore, the real-time monitoring capabilities and predictive maintenance aspects of the digital twin system enabled us to proactively identify potential equipment failures. This proactive approach was rigorously validated through deliberately introducing components with known defects, ensuring a robust test of the system's predictive acumen. The experimental stand built can be seen below.

The crucial connection points and the meticulous detail in our experimental stand can be seen in Figure 6. Moreover, the logical and control block with PLC of the digital twin stand can be seen in Figure 7.

Our SCADA system recorded live metrics like reservoir pressure, temperature, and oil migration. This data was fed into our digital twin system, generating a live model of the testing environment. The system's control and operational management are conducted via the SCADA interface, as illustrated in Figure 8.

Fig. 5: Experimental digital oil field test bench

Fig. 6: Terminal block of digital twin stand

Fig. 7: Logical and control block with PLC of digital twin stand

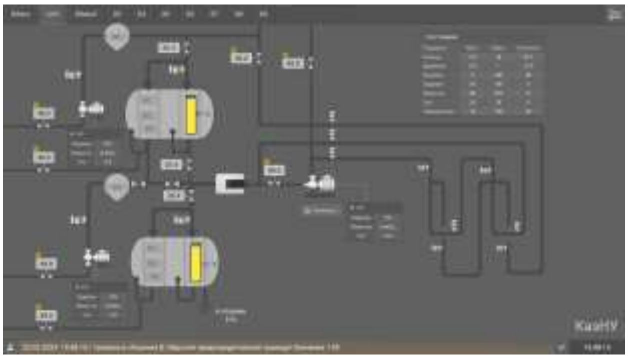

Fig. 8: Visualization in SCADA

Employing the developed LSTM algorithm, we decoded the combined data to predict oil displacement tendencies.

Upon comparing the actual temperature measurements with the predictions from our neural network, we found a strong correspondence between the two, indicating a high level of predictive accuracy by the network. The difference between real and predicted data were minimal, showing the model's efficiency. This minimal difference was quantitatively assessed using RMSE, which showed low values, reflecting the model's ability to closely mirror actual conditions. The consistent decrease in prediction inaccuracies across successive epochs further substantiates the model's learning efficacy and its capacity to refine its predictions over time. The results of our analysis showed an r2 score range of 90–95%. Moreover, our findings highlight a notable 10-15% improvement in oil production efficiency, providing insight into the transformative potential of merging DT and ML in the petroleum industry.

This study emphasizes the potential of a machine learning-based tool to enhance the management and operational efficiency of oil field test setups. Our roadmap envisions the integration of features such as oil transfer speed, tank pressure, and oil consistency. The advisory platform of the digital twin also played a crucial role in identifying and addressing equipment defects, including those induced intentionally as part of our system's resilience and fault-tolerance testing. This approach allowed us to simulate potential issues and assess the system's response to such anomalies, thereby enhancing the robustness and reliability of our oil displacement methods.

4 Conclusion

The integration of digital twins and machine learning in oil field management marks a monumental shift for the oil and gas sector. This union harnesses real-time analytics, predictive foresight, and data-fueled decision-making, leading to enhanced safety, efficiency, and optimal resource deployment.

In our pilot setup mimicking oil displacement technology, we showcased the tangible advantages of this methodology. Merging Digital Twin tech, SCADA systems, and machine learning, we dynamically fine-tuned oil field operations, showing a significant enhancement in oil production efficiency by 10-15%.

Key outcomes from our endeavor:

Our intelligent framework routinely amplified oil extraction rates by dynamically refining operation metrics and offering on-the-spot recommendations.

The ability to predict maintenance needs, coupled with preemptive alerts, curtailed operational downtime and bolstered safety—warding off equipment breakdowns and potential hazards.

The newfound precision in resource distribution ensured optimal utilization, thus ramping up production outputs.

The implementation of our system in real-world oil fields can revolutionize industry practices. By integrating our intelligent framework into existing infrastructure, oil companies can expect significant improvements in operational efficiency and safety. The system's predictive maintenance capabilities and dynamic operation adjustments can be particularly beneficial in large-scale operations, where they can lead to substantial cost savings and reduced environmental impact. This practical application highlights the system's potential for widespread adoption and its ability to address current industry challenges.

Future work will explore enhancing the real-time capabilities of the system, particularly focusing on refining the accuracy of predictions and the efficiency of the digital twin model. We also plan to address current limitations such as scalability to larger fields and integration with various types of oil extraction technologies.

References:

[1] F. Shen, S. S. Ren, X. Y. Zhang, H. W. Luo, and C. M. Feng, "A Digital Twin-Based Approach for Optimization and Prediction of Oil and Gas Production," *Math. Probl. Eng.*, vol. 2021, p. e3062841, Sep. 2021, doi: 10.1155/2021/3062841.

[2] C. Nunes, E. J. S. Pires, and A. Reis, "Machine Learning and Deep Learning applied to End-of-Line Systems: A review," *WSEAS Transactions on Systems*, vol. 21, pp. 147–156, 2022, https://doi.org/10.37394/23202.2022.21.16.

[3] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, "Application of machine learning and artificial intelligence in oil and gas industry," *Pet. Res.*, vol. 6, no. 4, pp. 379–391, Dec. 2021, doi: 10.1016/j.ptlrs.2021.05.009.

[4] R. van Dinter, B. Tekinerdogan, and C. Catal, "Predictive maintenance using digital twins: A systematic literature review," *Inf. Softw.*

Technol., vol. 151, p. 107008, Nov. 2022, doi: 10.1016/j.infsof.2022.107008.

[5] Z. Ren, J. Wan, and P. Deng, "Machine-Learning-Driven Digital Twin for Lifecycle Management of Complex Equipment," *IEEE Trans. Emerg. Top. Comput.*, vol. 10, no. 1, pp. 9–22, Jan. 2022, doi: 10.1109/TETC.2022.3143346.

[6] N. O. Nikitin, I. Revin, A. Hvatov, P. Vychuzhanin, and A. V. Kalyuzhnaya, "Hybrid and Automated Machine Learning Approaches for Oil Fields Development: the Case Study of Volve Field, North Sea," *Comput. Geosci.*, vol. 161, p. 105061, Apr. 2022, doi: 10.1016/j.cageo.2022.105061.

[7] A. Lektauers, J. Pecerska, V. Bolsakovs, A. Romanovs, J. Grabis, and A. Teilans, "A Multi-Model Approach for Simulation-Based Digital Twin in Resilient Services," *WSEAS Transactions on Systems and Control*, vol. 16, pp. 133–145, 2021, https://doi.org/10.37394/23203.2021.16.10.

[8] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.

[9] T. Chai, "Root Mean Square," in Encyclopedia of Mathematical Geosciences, B. S. Daya Sagar, Q. Cheng, J. McKinley, and F. Agterberg, Eds., in Encyclopedia of Earth Sciences Series, *Cham: Springer International Publishing*, 2020, pp. 1–3. doi: 10.1007/978-3-030-26050-7_280-1.

[10] M. Tang, L. Qiao, Z. Huang, X. Liu, Y. Peng, and X. Liu, "Accelerating SGD using flexible variance reduction on large-scale datasets," *Neural Comput. Appl.*, vol. 32, no. 12, pp. 8089–8100, Jun. 2020, doi: 10.1007/s00521-019-04315-5.

[11] L. Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent," in *Proceedings of COMPSTAT'2010, Y. Lechevallier and G. Saporta, Eds., Heidelberg: Physica-Verlag HD*, 2010, pp. 177–186. doi: 10.1007/978-3-7908-2604-3_16.

[12] H. Vo Thanh, Y. Sugai, and K. Sasaki, "Application of artificial neural network for predicting the performance of CO2 enhanced oil recovery and storage in residual oil zones," *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41598-020-73931-2.

[13] R. Al-Shabandar, A. Jaddoa, P. Liatsis, and A. J. Hussain, "A deep gated recurrent neural network for petroleum production forecasting," *Mach. Learn. Appl.*, vol. 3, p. 100013, Mar. 2021, doi: 10.1016/j.mlwa.2020.100013.

[14] N. Tasmurzayev, B. Amangeldy, Y. Nurakhov, D. Akhmed-Zaki, and Zh. Baigarayeva, "Intelligent Thermal Accumulator Operation Control System based on Renewable Energy Sources:" in *Proceedings of the 19th International Conference on Informatics in Control, Automation, and Robotics, Lisbon, Portugal: SCITEPRESS - Science and Technology Publications*, 2022, pp. 737–742. doi: 10.5220/0011319300003271.

[15] K. E. Nolan, W. Guibene, and M. Y. Kelly, "An evaluation of low power wide area network technologies for the Internet of Things," *2016 Int. Wirel. Commun. Mob. Comput. Conf. IWCMC*, pp. 439–444, Sep. 2016, doi: 10.1109/IWCMC.2016.7577098.

[16] Z. Sun and X. Tian, "SCADA in oilfields," *Meas. Control*, vol. 43, pp. 176–178, Jul. 2010, doi: 10.1177/002029401004300603.

[17] X. Yang, O. Bello, L. Yang, D. Bale, and R. Failla, "Intelligent Oilfield - Cloud Based Big Data Service in Upstream Oil and Gas," Mar. 2019. doi: 10.2523/IPTC-19418-MS.

[18] D. Du, X. Zhang, Q. Guo, B. Zhang, and G. Zhang, "Smart Oilfield Technology," in *Proceedings of the International Field Exploration and Development Conference 2018, J. Lin, Ed., in Springer Series in Geomechanics and Geoengineering. Singapore: Springer*, 2020, pp. 685–694. doi: 10.1007/978-981-13-7127-1_63.

[19] H. Wang, Y. Xu, B. Shi, C. Zhu, and Z. Wang, "Optimization and intelligent control for operation parameters of multiphase mixture transportation pipeline in oilfield: A case study," *J. Pipeline Sci. Eng.*, vol. 1, no. 4, pp. 367–378, Dec. 2021, doi: 10.1016/j.jpse.2021.07.002.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research was funded by a grant from the Ministry of Education and Science of the Republic of Kazakhstan BR18574136 "Development of deep learning and intellectual analysis methods for solving complex problems of mechanics and robotics" (2022-2024).

Conflict of Interest

The authors have no conflicts of interest to declare.