ORIGINAL ARTICLE
Electrica 2026; 26: 1-8

Big Data Stream Processing Models for Building a Predictive System for Controlling Biological Risks in CIP Processes

Javanshir Zeynalov[1], Vitaly Kargin[2], Beyali Ahmedov[3], Viktor Artemyev[4], Anar Hajiyev[3], Gultaj Zeynalzade[1]

[1]Nakhchivan State University, Nakhchivan, Azerbaijan
[2]Department of Automated Control Systems for Biotechnological Processes, BIOTECH University, Moscow, Russia
[3]Department of Machine Design and Mechatronics, Azerbaijan Technical University, Baku, Azerbaijan
[4]Department of Computer Science, Plekhanov Russian University of Economics, Moscow, Russia

Cite this article as: J. Zeynalov, V. Kargin, B. Ahmedov, V. Artemyev, A. Hajiyev and G. Zeynalzade, "Big data stream processing models for building a predictive system for controlling biological risks in CIP processes," *Electrica,* 2026, 26, 0300, doi: 10.5152/electrica.2026.25300.

WHAT IS ALREADY KNOWN ON THIS TOPIC?

- *Clean-in-Place (CIP) processes are critical for ensuring biological safety in pharmaceutical and food industries; however, conventional monitoring approaches rely on delayed and discrete microbiological testing.*
- *Most existing CIP control systems are reactive in nature and do not effectively utilize continuous sensor data streams for real-time prediction of biological risks.*
- *Big Data streaming technologies have been widely applied in industrial analytics; yet, their systematic use for predictive control of biological risks in CIP processes remains limited.*

Corresponding author:
Gultaj Zeynalzade

E-mail:
gultac.z@ndu.edu.az

Received: September 24, 2025
Revision Requested: November 5, 2025
Last Revision Received: January 8, 2026
Accepted: February 19, 2026
Publication Date: September 17, 2026

DOI: 10.5152/electrica.2026.25300

Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ABSTRACT

With the rapid growth of technological and biological data, the development of intelligent predictive control systems has become a pressing task to enhance the reliability and safety of industrial processes. This article explores streaming data processing models in the context of building a predictive control system for biological risks in Clean-in-Place (CIP) processes. Architectural approaches are examined for integrating streaming analytics with machine learning and predictive modeling tools that enable real-time identification of potential microbiological contamination risks, assessment of their dynamics, and forecasting of critical situations. Special attention is given to the selection and adaptation of streaming technologies such as Apache Kafka, Apache Flink, and Spark Streaming, considering the specific characteristics of biotechnological processes, high data variability, and the necessity of compliance with regulatory requirements in pharmaceutical and food production. It is demonstrated that hybrid models combining statistical methods, machine learning algorithms, and digital twins of CIP processes not only improve prediction accuracy but also allow proactive risk management within automated systems. The article also addresses the integration of predictive control with existing Quality Management Systems and Manufacturing Execution Systems platforms, thereby creating a unified digital monitoring ecosystem. The results of experimental modeling are presented, confirming the effectiveness of streaming models in minimizing the probability of biological risks and optimizing operational costs.

Index Terms— Digital twin, industrial analytics, machine learning, monitoring, predictive control

I. INTRODUCTION

The current development of the biotechnological, pharmaceutical, and food industries is accompanied by a radical increase in the complexity of production processes and the tightening of regulatory requirements for product quality and biological safety. In this context, the improvement of sanitation systems for process equipment becomes of paramount importance, with the Clean-in-Place (CIP) methodology occupying a central position. This approach ensures multilevel disinfection without dismantling equipment; however, the efficiency and reproducibility of CIP procedures directly determine the likelihood of microbiological risks. Any deviations in process parameters—from insufficient exposure time of cleaning solutions to fluctuations in temperature or disinfectant concentration—may result in residual contamination, leading to reduced sterility, increased production costs, and potential threats to end-user health [1-2].

At the same time, industrial digitalization and the widespread adoption of Internet of Things technologies are driving an exponential increase in the volume of process data. Intelligent sensor networks, SCADA systems, Manufacturing Execution Systems (MES) platforms, and digital parameter loggers generate continuous data streams [3] that capture both the macroparameters of CIP temperature profiles, hydrodynamic regimes, reagent dosing and hidden markers associated with biological risks dynamics of microbial contamination, washing efficiency, and biofilm

WHAT THIS STUDY ADDS TO THE TOPIC?

- *This study introduces an integrated Big Data streaming architecture based on Apache Kafka, Apache Flink, and Spark Streaming for real-time prediction of biological risks in CIP processes.*
- *A formal predictive control framework combining streaming analytics, machine learning techniques, and digital twins is developed to enable proactive and preventive risk management.*
- *The proposed solution supports low-latency decision-making and seamless integration with Manufacturing Execution Systems and Quality Management Systems platforms, ensuring compliance with international regulatory standards.*

inactivation [4]. Given such high variability and nonstationarity of the data, traditional statistical methods and classical quality control approaches areinsufficient for providing comprehensive, real-time risk assessment. For this reason, the use of Big Data and streaming data processing technologies becomes increasingly relevant, as they enable the integration of heterogeneous data sources and their intelligent analysis with minimal latency. The application of streaming analytics models combined with machine learning algorithms opens the way to shifting from reactive control toward predictive management of biological risks. In this context, the construction of digital twins of CIP processes is of particular importance, as they validate predictive models with empirical data and support adaptive control of sanitation parameters.

From a scientific perspective, the development of predictive control systems based on Big Data streaming requires solving a set of complex challenges: designing high-throughput data processing architectures, selecting optimal tools such as Apache Kafka, Apache Flink, and Spark Streaming, among others, integrating machine learning methods and deep neural networks into real production environments [5], and ensuring compliance of the developed solutions with international quality standards and regulatory frameworks such as GMP, ISO, and HACCP. The practical significance of this approach lies not only in improving the efficiency of CIP and reducing the probability of microbiological risks but also in optimizing operational costs through the reduction of equipment downtime, rational use of chemical agents and energy resources. Furthermore, the implementation of predictive control enables the formation of a unified digital quality management ecosystem that ensures process transparency and full compliance with Industry 4.0 requirements [6-8].

The present study is aimed at developing the methodological and technological foundations for constructing an intelligent predictive control system for biological risks in CIP processes based on Big Data streaming. The article systematizes current approaches to streaming analytics in industrial biotechnology, presents algorithmic solutions for risk prediction, and substantiates the prospects for integrating these technologies into real-world production processes.

II. FORMULATION OF THE PROBLEM

In the context of the intensive development of Industry 4.0 and the transition toward the concept of digital manufacturing, a contradiction persists between the high level of automation of CIP processes and the lack of adequate tools for predictive control of biological risks. Traditional monitoring systems rely on discrete microbiological test samples, which inevitably lead to delays in corrective actions. In contrast, streaming data generated by sensor circuits capture the dynamics of temperature regimes, electrical conductivity of cleaning solutions, hydrodynamic parameters, and chemical concentrations. However, existing analytical methods do not implement mechanisms for integrating these signals into a comprehensive risk prediction model [9-10].

According to (1), each component of the observation stream reflects a distinct sensor characteristic. On this basis, the biological risk at time t can be represented by a probabilistic functional that establishes the relationship between process parameters and the likelihood of contamination.

$$X(t)=\left(x_1(t),x_2(t),\ldots,x_n(t)\right), \tag{1}$$

where each component function corresponds to the recorded sensory-electrical characteristics of the CIP cycle. According to (1), the biological risk at time t can be represented by a probabilistic functional:

$$R(t)=P\left(f\left(X(t)\right)>\delta\right), \tag{1a}$$

where f is a mapping that approximates the hidden dependence between process parameters and the level of contamination, and δ is the critical threshold corresponding to sanitary and hygienic standards.

The task is to construct such a mapping model f_θ which, based on the streaming structure of data and accounting for their electrical, thermal, and chemical-technological nature, would enable the prediction of risk exceedance beyond the admissible limit within a minimal time interval Δt [11]. This necessitates the synthesis of a Big Data streaming architecture capable of simultaneously minimizing the delay functional and maximizing predictive accuracy, which constitutes the central problem of the present study [2].

Given that each measurement of CIP-process parameters represents an element of a complex stochastic field, the mathematical model of predictive control must account for the instability of dynamics, the multidimensionality of input signals, and the existence of hidden correlations between electrical, hydrodynamic, and biological processes [12]. This raises the need to formulate the problem in the space of streaming functions, where each parameter is subject to noise and drift, and the system itself exhibits properties of nonlinearity and nonstationarity.

Let the observation stream $\boldsymbol{X}(\boldsymbol{t})$ from (1) belong to the functional space *F*. Then, the risk prediction can be interpreted as the problem of minimizing the functional

Equation (2) thus formalizes predictive risk assessment as an optimization problem. The balance between accuracy and robustness is captured by the regularization parameter, ensuring that the model can operate effectively under real-time industrial conditions.

$$J[f]=E\left[L\left(f\left(X(t)\right),Y(t)\right)\right], \tag{2}$$

In conditions where each measurement of CIP process parameters constitutes an element of a complex stochastic field, the mathematical model of predictive control must take into account the instability of dynamics, the multidimensionality of input signals, and the presence of hidden correlations between electrical, hydrodynamic, and biological processes. This necessitates the formulation of the problem in the space of streaming functions, where each parameter is subject to noise and drift, and the system itself exhibits nonlinear and nonstationary properties [13].

By using (1) in (2), let the observation stream $\boldsymbol{X}(\boldsymbol{t})$belong to the functional space $L^2([0,T])$. The mapping function defined in (3) establishes a predictive relationship between real-time sensor data and the probability of biological risk.

$$J[f]=\int_0^T\left(R(t)-\check{R}(t)\right)^2dt+\lambda\int_0^T\nabla f\left(X(t)\right)^2dt, \tag{3}$$

where $\check{R}(t)$ is the estimate produced by the predictive control model, and λ is a regularization parameter reflecting the balance between prediction accuracy and the robustness of the algorithm to fluctuations in electrical and chemical-technological signals [14-16].

The problem of predictive control is thus reduced to a variational optimization under constraints on computational latency. In industrial settings, the key criterion is not so much the asymptotic accuracy of the model, but rather its ability to operate in real time under continuous inflows of high-frequency data streams. Consequently, the search for a solution must focus on the synthesis of Big Data streaming architectures integrated with machine learning, in which each new data element is immediately transformed into a corrective adjustment of the prediction.

III. METHODOLOGY

The formulated problem of predictive control of biological risks in CIP processes based on Big Data streaming requires a transition from traditional discrete monitoring systems toward the integration of continuous sensor data streams into a unified intelligent architecture. The key direction for addressing this challenge lies in the use of the streaming computation paradigm, wherein each data element is immediately incorporated into analysis and prediction processes. This approach not only reduces decision-making latency but also establishes the foundations for implementing the principle of predictive control.

Formalizing the problem, the observation stream $X(t)$ is considered as the input to a dynamic forecasting system, where the output is the risk estimate $\hat{R}(t)$. The solution reduces to constructing a mapping:

$$\hat{R}(t)=f_0\left(X(t)\right),$$

where f_0 is a parameterized model trained on historical and real-time data using machine learning methods. Adaptive learning is ensured by (4), which specifies the rule for online parameter updates under streaming conditions.

$$\theta_{t+1}=\theta_1-\eta\nabla_0J[f_0], \tag{4}$$

where η is the learning rate, and the functional *J* reflects the balance between prediction accuracy and robustness to noise.

The solution architecture assumes a multi-level structure. At the lower level, data stream preprocessing is performed: filtering of electrical signals, removal of noise components, and normalization of hydrodynamic and thermodynamic data. The middle level implements streaming analytics models Spark Streaming and Flink, where online approximation of nonlinear dependencies between CIP parameters and the probability of risk exceeding the admissible threshold is carried out. The upper level is integrated with the digital twin of the CIP system, which enables not only risk prediction but also the synthesis of control actions that minimize the probability of its occurrence [17]. The connection between controllable process variables and risk probability is formalized in (5), which frames the optimization of temperature, flow, and reagent concentration.

$$u^*(t)=\text{argmin}_{u(t)}R\left(X(t),u(t)\right), \tag{5}$$

where $u(t)$ are the control variables of the CIP process—temperature, solution flow rate, and reagent concentration—and *R* is the probability of biological risk occurrence. The problem is thus reduced to closing the loop, from streaming registration of electrical and technological parameters to predictive risk assessment and further to automatic corrective action on the process.

IV. FORMALIZATION OF THE PREDICTIVE CONTROL MODEL

To rigorously substantiate the conclusions presented in the comparative analysis, it is appropriate to complement this section with a mathematical model reflecting the operation of the streaming analytics architecture in CIP processes. Equation (6) characterizes the full sensor stream as a stochastic process, thereby capturing the inherent variability of CIP cycles.

$$X(t)=\left\{x_1(t),x_2(t),\ldots,x_n(t)\right\},\mathrm{t}\in[0,T], \tag{6}$$

where the components of $x_1(t)$include electrical, hydrodynamic, thermodynamic, and chemical-technological signals [3]. Robustness

of the predictive model is incorporated in (7), where a regularization term accounts for fluctuations in the data streams.

$$J(\theta)=E\left[\left(R(t)-f_0\left(X(t)\right)\right)^2\right]+\lambda\theta^2, \tag{7}$$

where $R(t)$ is the true risk determined by retrospective data, and the regularization term reflects the $\lambda\theta^2$ requirements for model robustness to signal fluctuations.

Equation (8) specifies the closed-loop structure of predictive control, linking sensor feedback with corrective process actions.

$$u^*(t)=\text{argmin}_{u(t)}E\left[R(X(t),u(t)\right], \tag{8}$$

where $u(t)$ is the control action, and the risk R is modeled as a function of sensor streams and control variables. By incorporating the digital twin, (9) reformulates the control task as an optimization over trajectories in the state space.

$$\min_{u(t)}\int_0^T (R\left(X(t),u(t)\right)+\alpha X(t)-X^{virt}(t)^2)dt, \tag{9}$$

where $X^{virt}(t)$ is the predicted trajectory of the digital twin, and the coefficient α defines the weight of convergence of the real process to its virtual representation.

Thus, the mathematical model captures the key elements of the architecture: the continuous data stream (Kafka), online learning and prediction (Flink), analysis and retraining on large historical datasets (Spark Streaming), as well as the control and regulatory loop (MES and QMS). The dynamic evolution of the CIP system is described by (10), which combines state variables, control inputs, and random disturbances

$$\frac{dX(t)}{dt}=AX(t)+Bu(t)+\xi(t), \tag{10}$$

where $X(t)\in R^n$ is the vector of observed parameters electrical signals, temperature, concentration, and hydrodynamic characteristics; $u(t)\in R^m$ is the control action adjustment of temperature, flow rate, and reagent dosing; A, B are the system dynamics matrices; is the random perturbation process modeling the stochastic nature of contamination. The probabilistic formulation of biological risk is presented in (11), translating system states into measurable indicators.

$$R(t)=P\left\{CX(t)\geq\delta\right\}, \tag{11}$$

where C is the observation matrix that maps the state parameters to the risk indicator, and δ is the sanitary-critical threshold that constrains acceptable contamination levels. The corresponding optimization problem is set out in (12):

$$\min_{u(t)}E\left[\int_0^T I\left(X(t),u(t)\right)dt\right], \tag{12}$$

where the cost function *I(X(T)*, given in (13), balances process performance against penalties for excessive use of energy or chemicals.

$$I\left(X(t),u(t)\right)=R(t)+\alpha u(t)^2, \tag{13}$$

and the parameter α >0 introduces a penalty for excessive use of control resources (energy, chemical).

To solve the problem, the apparatus of stochastic dynamic programming can be applied. The value function, *V(X,t)* which characterizes the minimal expected value of risk from time *t* to the prediction horizon *T*, is given by (14), where the Hamilton–Jacobi–Bellman

TABLE I. COMPARATIVE ANALYSIS OF BIG DATA STREAMING TOOLS FOR PREDICTIVE CONTROL IN CIP PROCESSES

Component	Primary Role in Architecture	Advantages	Limitations	Specific Application in CIP
Apache Kafka	Transport layer, message broker, distributed event log	High throughput, fault tolerance, guaranteed data delivery	No data analytics, acts only as a mediator	Continuous collection of electrical, hydrodynamic, and chemical sensor signals from CIP cycles
Apache Flink	Core of event-driven stream analytics	True real-time processing, minimal latency, precise event semantics	Higher setup complexity, resource-intensive	Instant detection of anomalies (temperature, conductivity, concentration) and real-time risk prediction
Spark Streaming	Analytical layer and model training	Scalability, MLlib integration, handling of large historical datasets	Micro-batch paradigm → higher latency compared to Flink	Model training and retraining, long-term contamination pattern analysis, integration with digital twin
MES platforms	Manufacturing execution and control	Operational automation, equipment integration	Dependent on enterprise architecture	Transforming predictive risk assessments into corrective actions (adjusting temperature, concentration, CIP duration)
QMS	Quality management system	Compliance with GMP/ISO, process documentation	No analytics, focused on quality control and certification	Registering predictive outcomes, ensuring conformity to regulatory and sanitary standards

CIP, clean-in-place.

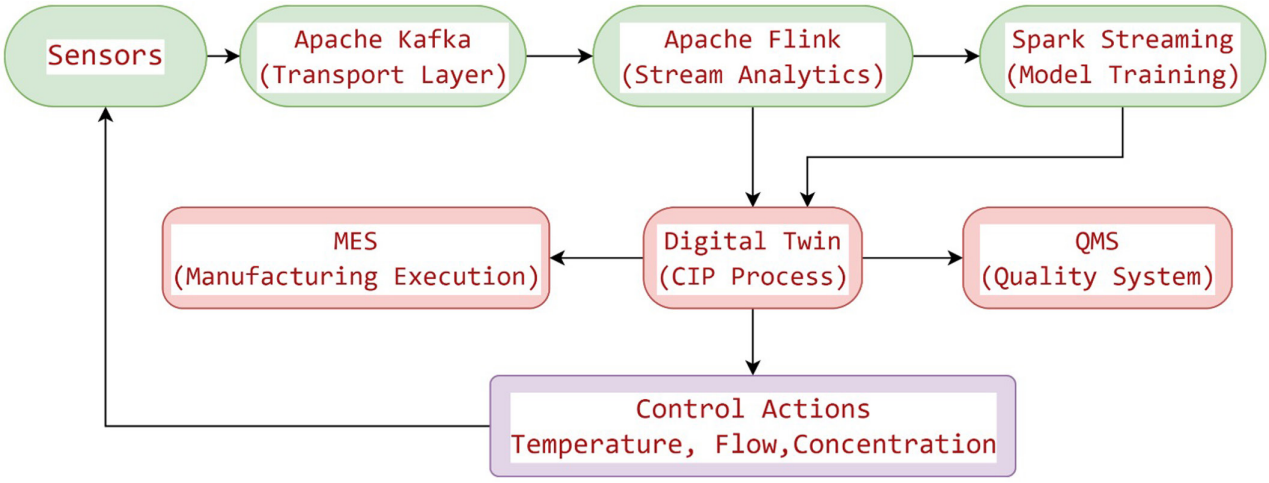

Fig. 1. Schematic model of predictive control architecture for CIP processes.

formulation governs the optimal control policy under uncertainty and adaptive decision-making.

$$-\frac{\partial V(X,t)}{\partial t}=\min_{u\in U}\left\{l(X,x)+\nabla_x V(X,t)\bullet(AX+Bu)+\frac{1}{2}\mathrm{Tr}\left(\sum\nabla_x^2 V(X,t)\right)\right\}, \tag{14}$$

where $\sum$ is the covariance matrix of the noise process $\xi(t)$

The formation of predictive control actions is reduced to solving the HJB equation in a streaming mode, which ensures adaptive updating of the control policy under the continuous inflow of sensor data. To systematize the key results of the analysis and to demonstrate the functional interrelation between the architectural elements of the proposed solution, it is advisable to move from the descriptive part to a structured comparison. This representation makes it possible to clearly identify the role of each streaming tool and corporate platform in the overall predictive control loop, to highlight their strengths and limitations, and to show the specifics of their application in CIP processes. For this reason, Table I below summarizes the characteristics of Apache Kafka, Apache Flink, Spark Streaming, as well as their integration with MES and QMS systems [5, 8]. Fig. 1 presents the schematic model of the predictive control architecture for CIP processes. As can be seen, the integration of streaming data platforms (Kafka, Flink, and Spark Streaming) with the digital twin, MES, and QMS ensures a closed-loop system, where predictive analytics are directly linked to corrective actions and quality assurance.

The schematic model clearly demonstrates that predictive control of CIP processes relies on the synergy of multiple technological layers. Sensor data are reliably transported through Apache Kafka, analyzed in real time by Apache Flink, and enriched with historical learning via Spark Streaming. The integration with the digital twin enables adaptive simulations, while MES platforms transform analytical insights into corrective process actions. Finally, QMS ensures that all outcomes comply with international regulatory standards. To provide a clearer understanding of the role of each technological component in the predictive control architecture, a comparative analysis is summarized in Table I. This table highlights the primary function, advantages, limitations, and specific applications of Kafka, Flink, Spark Streaming, as well as their integration with MES and QMS in CIP processes.

This table reflects the distribution of functions and the roles of each component within the overall solution. Kafka provides the fundamental reliability of data streams, Flink performs instant event analysis, Spark Streaming is responsible for "deep" analytics and model training, MES generates control actions, and QMS consolidates the results within the quality system. The comparative analysis confirms that only the comprehensive use of these components ensures the achievement of predictive control objectives. Kafka forms the foundation of reliability and continuity of data flow, Flink serves as the core of event-driven analytics with minimal latency, Spark Streaming is responsible for advanced learning and the identification of long-term patterns, while integration with MES and QMS closes the full quality management cycle by transforming analysis results into corrective actions and compliance protocols. Thus, this combination creates a synergistic effect in which technical resilience, analytical depth, and regulatory significance merge into a unified digital ecosystem of predictive control of biological risks in CIP processes.

V. RESULTS AND CONCLUSIONS

The conducted study demonstrates that the development of predictive control systems for biological risks in CIP processes based on Big Data streaming is not only promising but also methodologically necessary for the advancement of modern biotechnological and pharmaceutical industries. In the context of the exponential growth of technological information, traditional discrete monitoring methods prove incapable of ensuring the required level of responsiveness, robustness, and accuracy in sanitary risk assessment. The developed concept of integrating streaming analytics tools with corporate management systems forms a holistic architecture in which each component performs a strictly defined function. Apache Kafka ensures the continuity and reliability of sensor signal transport, Apache Flink implements true streaming processing with minimal latency, and Spark Streaming provides tools for large-scale risk model training and the identification of long-term patterns. Together, they form a resilient analytical loop capable of functioning in real time and ensuring timely detection of anomalies in CIP cycles.

The integration of predictive models with MES platforms and QMS systems closes the management loop: predictive risk assessments are transformed into control actions on the technological process and are simultaneously recorded in the quality system, ensuring full compliance with international GMP and ISO standards. This implementation transforms the monitoring system from a reactive tool into an active element of predictive and preventive biosafety management. The mathematical model developed within the study confirms the possibility of formalizing the predictive control problem in terms of risk functional optimization, taking into account the stochastic nature of sensor streams and computational latency constraints. The inclusion of a digital twin of the CIP process in the architecture expands the analytical capabilities of the system and enables optimization of control actions in the space of possible trajectories, ensuring resilience to fluctuations and improving the reliability of predictions.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Artificial Intelligence Usage Statement: The authors declared that no artificial intelligence tool was used in the preparation of the manuscript.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – J.Z., G.Z.; Design – J.Z., V.K.; Supervision – J.Z., G.Z.; Resources – V.K., B.A.; Materials – B.A., A.H.; Data Collection and/or Processing – V.A., A.H.; Analysis and/or Interpretation – J.Z., V.A.; Literature Search – G.Z., V.A.; Writing – G.Z., J.Z.; Critical Review – V.K., A.H.

Declaration of Interests: The authors have no conflict of interest to declare.

Funding: The authors declared that this study has received no financial support.

REFERENCES

1. A. Jamarani, S. Haddadi, R. Sarvizadeh, M. Haghi Kashani, M. Akbari, and S. Moradi, "Big data and predictive analytics: A systematic review of recent trends," *Artif. Intell. Rev.*, vol. 57, no. 7, 2024. [CrossRef].
2. P. Goel, A. Datta, and M. S. Mannan, "Application of big data analytics in process safety and risk management," IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 2017, 2017, pp. 1143–1152. [CrossRef].
3. O. Khattach et al., "End-to-end architecture for real-time IoT analytics and predictive insights," *Sensors, PMC*, 2025.
4. T. Kolajo, O. Daramola, and A. Adebiyi, "Big data stream analysis: A systematic literature review," *J. Big Data*, vol. 6, no. 1, Art. 47, 2019. [CrossRef].
5. H. Zheng, A. R. Paiva, and C. S. Gurciullo, *Advancing from Predictive Maintenance to Intelligent Maintenance with AI and IoT*, 2020.
6. O. Alotaibi, E. Pardede, and S. Tomy, "Cleaning big data streams: A systematic literature review," *Technologies*, vol. 11, no. 4, Art. 101, 2023. [CrossRef].
7. H. Kumar, "Big data streaming platforms: A review," *Int. J. Comput. Sci. Manag.*, vol. 3, no. 2, pp. 33–41, 2022.
8. A. Di Pace, G. Fenza, M. Gallo, V. Loia, A. Meglio, and F. Orciuoli, "'Implementing the cognition level for Industry 4.0 by integrating augmented reality and manufacturing execution systems,' *arXiv*," 2020, *arXiv*. https://arxiv.org/abs/2011.10482.
9. Q. Zhu, "Research on the optimization of performance evaluation system of listed companies under the new development pattern," *Acad. J. Bus. Manag.*, vol. 5, No. 7, pp. 103–108, 2023. [CrossRef].
10. M. Arshad, C. W. Onn, A. Ahmad, and G. Mogwe, "Big data analytics and AI as success factors for online video streaming platforms," *Front. Big Data*, vol. 8, Art. 1513027, 2025. [CrossRef].
11. T. Vergilio, A.-L. Kor, and D. Mullier, "A unified vendor-agnostic solution for big data stream processing in a multi-cloud environment," *Appl. Sci.*, vol. 13, no. 23, Art. 12635, 2023. [CrossRef].
12. L. Zhao, "Event prediction in the big data era: A systematic survey," *ACM Comput. Surv.*, vol. 54, no. 5, Art. 1, 2021. [CrossRef].
13. Z. Feng, Y. Chen, Y. Xu et al., "'Intelligent perioperative system: towards real-time big data analytics in surgery risk assessment,' *arXiv*," 2017, *arXiv*. Available: https://arxiv.org/abs/1709.10192.
14. I. Verenich, M. Dumas, M. La Rosa, and F. M. Maggi, "'Survey and cross-benchmark comparison of remaining time prediction methods in business process monitoring,' *arXiv*," 2018, *arXiv*. https://arxiv.org/abs/1805.02896.
15. G. Nagarajan, and D. B. L.d, "Predictive analytics on big data – An overview," *Informatica*, vol. 43, no. 4, pp. 289–300, 2019. [CrossRef].
16. A. I. Torre-Bastida et al., "Bio-inspired computation for big data fusion, storage and analysis: A survey," *Neurocomputing*, vol. 423, pp. 357–372, 2021. [CrossRef].
17. M. Canizo, E. Onieva, A. Conde, S. Charramendieta, and S. Trujillo, "'Real-time predictive maintenance for wind turbines using Big Data frameworks,' *arXiv*," 2017, *arXiv*. Available: https://arxiv.org/abs/1709.07250.

Javanshir Zeynalov was born on February 26, 1965, in Bilav village of Ordubad district. In 1982, he was admitted to the Faculty of Automation of Production Processes at the Azerbaijan Institute of Oil and Chemistry. After completing his military service from 1983 to 1985, he graduated from the institute in 1989 and has since been working at Nakhchivan State University. From 1992 to 1995, he pursued postgraduate studies at the Institute of Space Research. In 1996, he defended his PhD dissertation in Engineering, and in 2013, he successfully defended his Doctor of Sciences dissertation in Mathematics (Computer Science). Since 2000, he has held the position of Associate Professor at the Department of Communication and Information Technologies, and he currently serves as Dean of the Faculty of Architecture and Engineering at Nakhchivan State University.

Vitaliy Kargin is a Professor in the Department of Automated Control Systems for Biotechnological Processes at ROSBIOTECH, Institute of Industrial Engineering, Information Technologies and Mechatronics. He holds a Candidate of Technical Sciences degree (awarded in 2008) and a Doctor of Technical Sciences degree (awarded in 2021). His academic title of Associate Professor was awarded in 2011. Professor Kargin has approximately 20 years of total work experience, with 18 of those in teaching. His research focuses on adaptive automatic control systems and power electromagnetic drives, especially in food production. He has also completed multiple advanced training programs in institutions like Saratov State Agrarian University and Saint Petersburg Polytechnic University.

Bayali Ahmadov was born on March 15, 1965. On December 16, 2021, Bayali Ahmadov successfully defended his dissertation for the degree of Doctor of Technical Sciences, specialty 3313.02.—"Machines, equipment and processes" at the ED 2.32 Specialized Scientific Council operating under the Azerbaijan Technical University and by the decision of the Higher Attestation Commission under the President of the Republic of Azerbaijan dated October 7, 2022, he was awarded the scientific degree of Doctor of Science in the field of technical sciences. He is the author of more than 100 scientific works.

Viktor Artemyev, Senior Lecturer at the Department of Informatics of the Plekhanov Russian University of Economics, building on the engineering tradition of the Moscow Automobile and Road Construction State Technical University and a unique combination of two master's programs, creates an educational and scientific environment where classical computer science theory is organically intertwined with the practice of digital transformation, artificial intelligence, and big data. His professional activity reflects the key priorities of Industry 4.0, including the development of cyber-physical systems, intelligent platforms, Internet of Things technologies, digital twins, and robotic complexes for the automation of production and managerial processes. As the author of certificates of authorship for computer software products and methodological manuals on SCADA systems, a certified specialist in industrial automation and digital control systems, Editor-in-Chief of the scientific journal Technologies in Management, and an expert in professional championships on cross-cutting digital competencies, he combines in his work a deep academic tradition with advanced tools of the future, revealing to students the interconnectedness of science and practice, business and government, society and technology, and forming a holistic vision of the role of intelligent systems in building a sustainable digital world.

Gultaj Zeynalzada was born on July 20, 1996, in Shahbuz district. She was admitted to the Academy of Public Administration under the President of the Republic of Azerbaijan, majoring in Computer Science. She graduated with honors in 2018. Between 2018 and 2020, she pursued a master's degree in Computer Science with a specialization in "Mathematical and Information Support of Economic Activity" at Azerbaijan State University of Economics, also graduating with honors. From 2020 to 2023, she worked as a senior engineer at the Department of Information and Communication Technologies and E-Education of Nakhchivan State University. Since 2023, she has worked as a senior lecturer at the Department of Electronics and Information Technologies of the same university.

Anar Hajiyev was born on June 29, 1991. He completed a Bachelor's degree in Engineering at Azerbaijan Technical University (2009–2013) and later obtained a Master's degree from the same institution (2014–2016). He currently serves as Deputy Dean for International Affairs and Research at Azerbaijan Technical University. His research interests include oil and gas mechanical systems, with a focus on sucker-rod pumping units, as well as the integration of sustainable and green energy technologies into higher education. He has led several international and EU-funded initiatives.