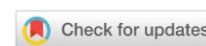


INFORMATION TECHNOLOGY, COMPUTER SCIENCE AND MANAGEMENT ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ



UDC 004.896:004.032.26

Original Empirical Research

<https://doi.org/10.23947/2687-1653-2026-26-3-2619>

An Intelligent Model for Predicting Spike Type Technological Process Disturbances in Aluminum Electrolysis Based on Ensemble Learning Technologies

Anton S. Mikhalev , Tatiana G. Penkova ✉,

Ludmila F. Nozhenkova

Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russian Federation, <https://ror.org/047f8bv22>✉ penkova_t@icm.krasn.ru

EDN: QSBEGG

Abstract

Introduction. Maintaining high technical and economic performance in aluminum production is a pressing operational challenge. The use of artificial intelligence makes it possible to mitigate the impact of process irregularities such as the deformation of the anode bottom resulting in spike formation. The causes and consequences of spike formation have been thoroughly studied. Dozens of parameters signaling this disturbance are identified. However, existing solutions typically detect deviations only after they have occurred or predict related states. Industrial realities demand adequate, advance warning of the problem. The objective of this study is to develop and validate an intelligent model for the early prediction of spike type disturbances under aluminum production.

Materials and Methods. The study has analyzed over 70 process parameters based on average daily monitoring data from eight electrolyzers at a pilot site (2020–2025). The positive class was defined based on the pre-failure interval spanning 72 to 24 hours prior to the recorded disturbance. Ensemble methods were used for prediction:

- gradient boosting (XGBoost, CatBoost, LightGBM);
- FEDOT automated pipeline;
- neural network ensembles (RTDL-RM, RTDL-NE, TabM).

Quality metrics: Accuracy, Precision, Recall, F1.

Results. The LightGBM model was identified as optimal, with the following metrics: Accuracy — 0.88; Recall — 0.79; F1 — 0.79. For events without disturbances, it correctly classified the “normal” state in 95% of cases (900 observations), while incorrectly identifying “disturbance” in 5% of cases (46 observations). For events involving disturbances, 62.5% of states were correctly identified (10 observations), whereas disturbances were missed in 37.5% of cases (6 observations). Validation using data from an actual technological process showed the model ability to detect spikes 24–72 hours prior to their actual registration. During the validation, six spike formation events were recorded, and the model detected signs of the developing disturbance in five of them (83.3%).

Discussion. The research results are consistent with the physical and technological characteristics of spike formation. The proposed solution complements existing methods for the local monitoring of anodes and electrolyzers through analyzing parameters derived from daily monitoring data. Validation confirms known challenges for predictive models in industrial electrolysis: difficulty in event detection, class imbalance, and changing process characteristics. The monitoring data from eight electrolyzers at the pilot site are insufficient. Further testing is required to confirm the model generalizability — over longer periods and at other sites.

Conclusion. The ability of the model to predict a disturbance 24–72 hours prior to the detection of a spike formation was confirmed, as was its sufficient effectiveness given the constraints imposed by sample size and the variability of process conditions. Future research will involve data augmentation, model validation at other production sites, and a comprehensive analysis of average daily and instantaneous monitoring parameters to localize the disturbance.

Keywords: process disturbances in aluminum production, prediction of spike defects, LightGBM model, anode bottom deformation, monitoring of anodes and electrolyzers

Acknowledgements. The authors would like to thank the lead technicians of production automation at RUSAL Engineering and Technology Center LLC for their expert assessment and the opportunity to test the developed models.

Funding Information. The research was done within the framework of the project “Mathematical and information modeling of complex physical and industrial processes and systems” (FWES-2024-0013 project code) from the state assignment of the Institute of Computational Modelling, the Siberian Branch of the Russian Academy of Sciences (124012900550-1 state registration number of the R&D project), <https://gisnauka.ru/nioktr/detail/B5CV3ILS5GQHT0KT8Y7VHDA6>

For Citation. Mikhalev AS, Penkova TG, Nozhenkova LF. An Intelligent Model for Predicting Spike Type Technological Process Disturbances in Aluminum Electrolysis Based on Ensemble Learning Technologies. *Advanced Engineering Research (Rostov-on-Don)*. 2026;26(3):2619. <https://doi.org/10.23947/2687-1653-2026-26-3-2619>

Оригинальное эмпирическое исследование

Интеллектуальная модель прогнозирования технологических нарушений электролиза алюминия типа «конус» на основе методов ансамблевого обучения

А.С. Михалев , Т.Г. Пенькова ✉, Л.Ф. Ноженкова 

Институт вычислительного моделирования Сибирского отделения Российской академии наук, г. Красноярск, Российская Федерация, <https://ror.org/047f8bv22>

✉ penkova_t@icm.krasn.ru

Аннотация

Введение. Обеспечение высоких технико-экономических показателей выпуска алюминия — актуальная производственная задача. Использование искусственного интеллекта позволяет ограничить влияние таких технологических нарушений, как деформация подошвы анода с формированием конусов. Достаточно хорошо изучены причины и последствия образования конусов. Названы десятки параметров, которые сигнализируют о нарушении. Однако известные решения фиксируют уже возникшие отклонения или предсказывают смежные состояния. Производственные реалии требуют адекватного заблаговременного уведомления о проблеме. Цель представленной работы — создание и апробация интеллектуальной модели для раннего прогнозирования нарушений типа «конус» при производстве алюминия.

Материалы и методы. Исследовались более 70 технологических параметров, данные среднесуточного мониторинга восьми электролизеров опытного участка (2020–2025 гг.). Положительный класс сформирован по предаварийному интервалу 72–24 ч до регистрации нарушения. Для прогнозирования применялись ансамблевые методы:

- градиентный бустинг (XGBoost, CatBoost, LightGBM);
- автоматизированный конвейер FEDOT;
- нейросетевые ансамбли (RTDL-RM, RTDL-NE, TabM).

Метрики качества: Accuracy, Precision, Recall, F1.

Результаты исследования. Оптимальной признана модель LightGBM: Accuracy — 0,88; Recall — 0,79; F1 — 0,79. В категории событий без нарушений она корректно классифицировала состояние «норма» в 95 % случаев (900 наблюдений). В 5 % случаев (46 наблюдений) ошибочно распознано «нарушение». В категории событий с нарушениями верно распознаны 62,5 % состояний (10 наблюдений), в 37,5 % случаев (6 наблюдений) нарушения пропущены. Апробация на данных реального технологического процесса доказала способность модели выявлять конусы за 24–72 ч до фактической регистрации. При апробации зарегистрировано шесть событий обнаружения конуса, модель выявила признаки развития нарушения для пяти (83,3 %).

Обсуждение. Результаты согласуются с физико-технологическими особенностями возникновения конуса. Предложенное решение дополняет методы локального контроля анодов и электролизеров благодаря анализу параметров среднесуточного мониторинга. Апробация подтвердила известные проблемы прогностических моделей в условиях промышленного электролиза: сложность регистрации событий, дисбаланс классов, изменение характеристик. Данных мониторинга восьми электролизеров опытного участка недостаточно, и для подтверждения универсальности модели нужна дополнительная проверка — с другой продолжительностью, на других участках.

Заключение. Подтвердились способность модели прогнозировать нарушение за 24–72 часа до регистрации конуса и ее достаточная эффективность при ограничениях, обусловленных объемом выборки и изменчивостью технологических условий. Дальнейшие исследования предполагают аугментацию данных, проверку модели на других участках производства, комплексный анализ параметров среднесуточного и мгновенного мониторинга для локализации нарушения.

Ключевые слова: технологические нарушения при производстве алюминия, прогнозирование нарушений типа «конус», модель LightGBM, деформация подошвы анода, контроль анодов и электролизеров

Благодарности. Авторы выражают благодарность ведущим специалистам автоматизации производства ООО «РУСАЛ Инженерно-технологический центр» за экспертную оценку и возможность апробации моделей.

Финансирование. Работа выполнена в рамках проекта «Математическое и информационное моделирование сложных физических и производственных процессов и систем» (код (шифр) проекта FWES-2024-0013) из государственного задания ИВМ СО РАН — обособленного подразделения ФИЦ КИЦ СО РАН (номер государственного учета НИР 124012900550-1), <https://gisnauka.ru/nioctr/detail/B5CV3ILS5GQHT0KT8Y7VHDA6>

Для цитирования. Михалев А.С., Пенькова Т.Г., Ноженкова Л.Ф. Интеллектуальная модель прогнозирования технологических нарушений электролиза алюминия типа «конус» на основе методов ансамблевого обучения. *Advanced Engineering Research (Rostov-on-Don)*. 2026;26(3):2619. <https://doi.org/10.23947/2687-1653-2026-26-3-2619>

Introduction. The development of the aluminum industry requires increasing the production volume of primary aluminum [1]. Achieving high technical and economic performance indicators necessitates advanced methods for controlling complex technological processes. Their objective is to provide the timely assessment of the state of the entire production complex [2] and its individual components [3]. Key limiting factors include process disturbances occurring during electrolysis [4], such as the deformation of the anode bottom, which leads to the formation of characteristic protrusions known as spikes [5] (Fig. 1).

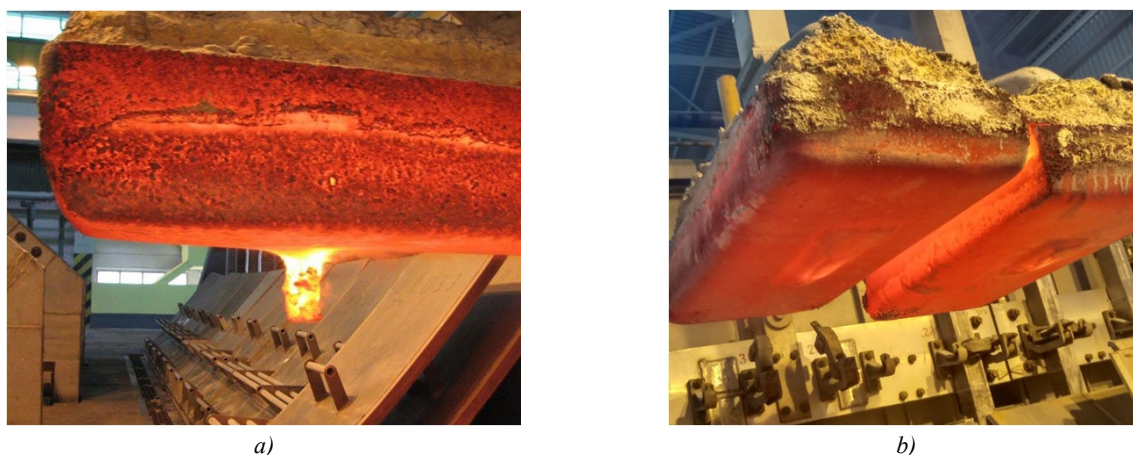


Fig. 1. Anode surface deformation:
a — typical spike; b — “lag” spike

Possible causes of spike formation:

- nonuniformity of the carbon anode structure and composition;
- disturbance of the process specifications during anode installation or replacement [6];
- changes in electrolysis process conditions related to disturbances in the heat balance and alumina concentration [7].

This disturbance alters the anode-cathode distance and current distribution within the electrolyzer. Consequently, spike formation can lead to a short circuit between the anode and cathode, increased current fluctuations, intense foaming, and increased noise levels [8].

The formation of spikes must be recognized as a systemic issue that significantly impacts the technological and energy-related performance of the electrolysis process. Spike growth leads to a sharp redistribution of current, localized overheating, short circuits, and a reduction in current efficiency — a drop that can reach several percent when such irregularities occur on a large scale. Even a slight reduction in the frequency of these events at an industrial scale can substantially improve energy efficiency, process stability, and aluminum output.

Studies on process parameter values under various operating conditions in aluminum production show that spike formation is accompanied by extreme values across a dozen parameters [9]. Due to the harsh conditions and the complexity of the technological process, timely diagnosis of this disturbance is extremely difficult, and the deviation is often detected only at a critical stage. Furthermore, automation is significantly complicated by numerous interdependent parameters, manual measurement methods, and the process inherent inertia. Under these conditions, analyzing accumulated data to identify indirect signs of developing malfunctions becomes particularly important. Intelligent methods and big data processing technologies enable the extraction of non-trivial, previously unknown, and interpretable insights. This approach makes it possible to explain performance degradation and to detect and predict deviations in the technological process in a timely manner [10].

Machine learning methods have become key tools for analysis [11] and optimization of production processes [12]. They are widely applied to diagnose process disturbances in the aluminum industry [13]:

- reconstruction of values for parameters that are difficult to measure;
- identification of relationships between the operating indicators of industrial electrolyzers and prediction of the probability of changes in their process state.

Particular attention is paid to the virtual estimation of key electrolysis process parameters that are difficult or impossible to monitor directly in real time:

- alumina concentration;
- melt temperature;
- anode-cathode distance [14].

While such models address the task of assessing the current state of the process, they do not allow for the early detection of the onset of process disturbances.

A distinct area of research focuses on predicting process disturbances and abnormal operating conditions. Statistical analysis of production data, expert knowledge, and intelligent models have formed the basis for a review of methods for detecting disturbances and process deviations in aluminum electrolyzers [15]. The primary objective of research in this field is the identification of deviations. It should be noted that predicting a specific disturbance requires analyzing its preceding dynamics. The complexity of diagnostics stems from the multidimensional nature of the electrolysis process, the interdependences between process parameters, and the variability of operating conditions.

To detect anodes developing a spike, an analysis of anode currents using machine learning methods is proposed in [16]. This solution makes it possible to identify a significant proportion of disturbances several days before the signs of spike formation become apparent to operating personnel. Anode currents provide high sensitivity to localized changes in a specific anode. However, this approach requires a system capable of measuring individual currents and analyzing local electrical characteristics.

In [17], an industrial model is proposed for predicting spike formation using Hidden Markov Models. The authors highlight challenges specific to this task:

- class imbalance;
- difficulty of reliably recording events;
- cross-correlation of process parameters;
- changes in data statistical characteristics resulting from variations in electrolyzer operating modes.

The results obtained demonstrate the ability to predict spikes using machine learning. At the same time, it is necessary to increase the stability of models to changes under process conditions, reduce the number of false warnings, and take into account the dynamics of production parameters.

Known solutions mainly fix existing deviations or predict adjacent states. So far, insufficient attention has been paid to building intelligent models for early prediction of disturbances. There has been some success in predicting disturbances with clearly defined changes in individual process parameters [18]. Examples include the anode effect [19], alumina concentration imbalances, and various mechanical and electrical malfunctions [20]. However, the challenge of early prediction regarding spike formation remains unresolved. This requires a comprehensive analysis of multidimensional monitoring data to identify — based on available process parameters — the latent patterns underlying the onset and progression of such a disturbance. In addition to high prediction accuracy tailored to the specific production facility, the predictive model must also be interpretable. For industrial application, classifying the state of the electrolyzer is insufficient. It is essential to determine which process parameters contribute most significantly to the prediction, which indicators foreshadow disturbances, and the extent to which the resulting relationships align with the physical and technological nature of the process. Models are required to analyze feature importance and identify anomalous patterns.

The objective of the present work is to develop an intelligent model for the early prediction of spike type process disturbances in aluminum production. The proposed model is based on a comprehensive analysis of process parameters using ensemble learning methods. Three tasks are addressed to achieve this aim.

1. Generation and preprocessing of a process data array for the application of machine learning methods.
2. Development of an intelligent predictive model in the form of an ensemble of binary classification models capable of determining the state of an object — “normal” or “disturbance” — based on new values of process parameters.
3. Approbation of the model using real monitoring data at a pilot aluminum production site.

Materials and Methods. From 2020 to 2025, monitoring was conducted at the pilot site of an aluminum production complex. The operation of eight electrolyzers with baked anodes using RA-550 technology was tracked. Data from the control system and the automated process control system yielded more than 70 parameters for analysis, characterizing the operation of individual units (electrolyzers, potrooms) under various operating conditions. Some of these indicators included: metal level, electrolyte level, electrolyte temperature, metal tapping duration, electrolyte chemical composition parameters, parameters of the alumina and aluminum fluoride feeding system, power supply modes, and parameters for regulating the interpolar distance between the metal and the carbon anode bottom, voltage, current, and noise parameters, characteristics of recorded process disturbances, and indicators of electrolyzer service life and process operations.

To formalize the development of the predictive model, a sequence of data processing and analysis stages was defined (Fig. 2).

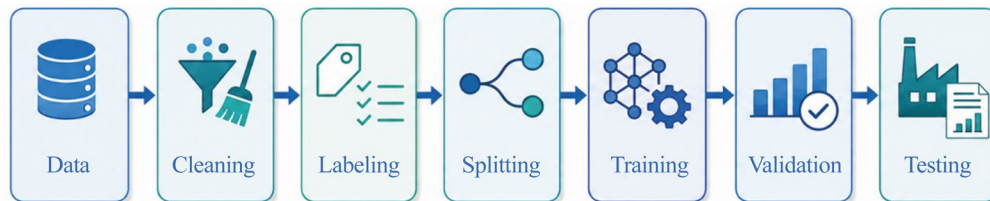


Fig. 2. Schematic of the intelligent model for predicting spike type disturbances

At the initial stage, a production monitoring dataset is compiled. The data are then cleaned and preprocessed: missing values are addressed, record validity is verified, and the feature space is prepared. Subsequently, observations are labeled as either “normal” or “disturbance”. Machine learning outcomes depend heavily on the quality and completeness of the input data and the accuracy of the labeling: the more complete and more precise the source information, the higher the reliability of the predictions. Finally, the data are split into training, validation, and test sample, taking into account the temporal structure of the process. The technological process is characterized by a pronounced time dependence, inertia, and the recurrence of states across consecutive periods. Accordingly, it is appropriate to employ a temporal split, where the model is trained on earlier observations and validated on subsequent ones. This approach makes it possible to evaluate the model ability not only to reproduce known patterns but also to predict the progression of a process disturbance under conditions approximating real-world operations. The next stage involves training and tuning the predictive model and selecting optimal parameters. This is followed by validation and a comparative assessment of the prediction results based on selected quality metrics. The final stage entails testing the model, specifically, verifying its ability to predict the spike in advance using real process data.

The spike can form quite rapidly: the time elapsed from the appearance of the defect to the short-circuiting of the anode and cathode is sometimes around three days [21]. An increase in voltage and current noise is observed as early as the second or third day [22]. Consequently, the three-day interval preceding the detection of the disturbance is regarded in this study as a significant pre-failure period for analyzing monitoring parameters and developing an early predictive model.

When constructing the training sample, the day of actual disturbance registration was taken into account, that is, the day on which the spike was detected during visual inspection of the anode removed from the electrolyzer. Data from the day of disturbance registration were excluded from the training sample. For each event registered at time T , observations falling within the pre-failure interval from $T-72$ h to $T-24$ h inclusive were assigned to the positive class.

During the preprocessing stage of the monitoring data, missing values were imputed using the multiple imputation by chained equations (MICE) algorithm [23]. Each parameter containing missing values was treated in turn as the target variable, while the remaining parameters served as features. Bayesian Ridge regression was employed as the base model for estimating the missing values. Prior to the iterative process, initial missing values were filled with the median of the corresponding feature. Subsequently, the parameters with missing values were processed sequentially: the model was trained on observations with known values and then used to predict the missing values in the corresponding records.

To construct an intelligent predictive model, two groups of methods were investigated: decision tree-based ensembles and neural network ensembles. Ensemble methods rely on combining multiple base models, which enhances prediction accuracy and robustness. In this study, the following implementations of gradient boosting on decision trees were selected for predicting disturbances [24]: XGBoost, CatBoost, and LightGBM. These algorithms belong to the class of industrial-grade solutions. They are used to handle heterogeneous feature spaces and demonstrate robust performance in classification tasks. The XGBoost (extreme gradient boosting) algorithm is an optimized implementation of gradient

boosting that combines loss function minimization with model complexity regularization. This approach aims to enhance the algorithm generalization capability and reduce the risk of overfitting. A distinctive feature of the CatBoost (categorical boosting) algorithm is a specialized scheme for handling categorical variables and constructing sequential estimates, which mitigates the likelihood of biases associated with feature encoding. The LightGBM (light gradient-boosting machine) algorithm is designed for the efficient processing of large datasets, delivering high training speeds while maintaining acceptable accuracy. In addition to individual gradient boosting algorithms, the study employs the FEDOT library. This library utilizes evolutionary methods and is designed for the automated construction of composite machine learning pipelines. It is used to search for model structures and to select methods for data preprocessing and result aggregation. This approach allows for the evaluation of not only predefined models but also automatically generated combinations of methods, as well as the comparison of their performance under identical training and testing conditions.

In addition to decision tree-based ensemble algorithms, deep learning approaches utilizing neural network ensembles, such as RTDL-RM, RTDL-NE, and TabM, are being investigated [25]. The RTDL-RM (RTDL-revisiting-models) approach involves applying standard neural network architectures to tabular data. High performance can be achieved through ensembling, activation function tuning, normalization, and regularization. The RTDL-NE (RTDL-num-embeddings) modification employs specialized representations for numerical features, where each numerical parameter is transformed into a vector. This embedding enables models to more effectively capture complex nonlinear relationships between the features and the target variable. The TabM approach combines several expert models that jointly form the final prediction. To reduce computational complexity, common parameters and individual low-rank adapters are used, built according to the LoRA (low-rank adaptation) principle. This allows increasing the number of models in the ensemble without a significant growth of the number of trained parameters.

The most efficient algorithms and optimal settings for intelligent models are selected based on classification quality assessment metrics: Accuracy, Precision, Recall, and F1-score [26].

To build and evaluate intelligent models, the initial dataset from daily average monitoring is partitioned taking into account the temporal structure of the production process (Fig. 3).

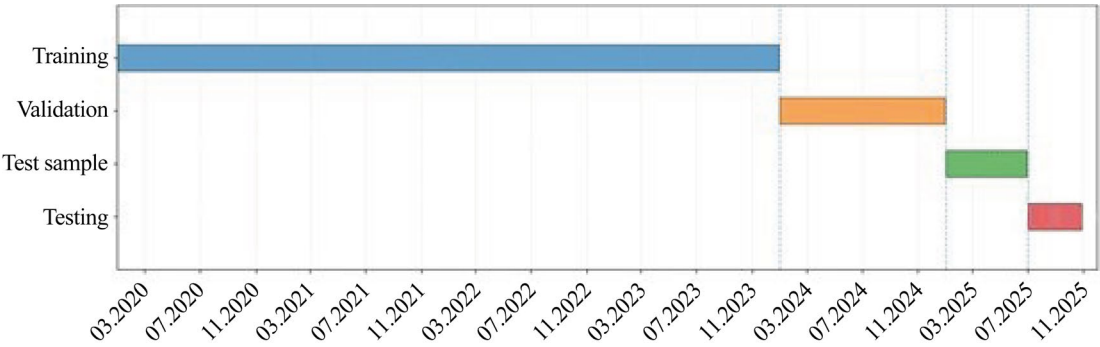


Fig. 3. Schematic of temporal data splitting into training, validation, test sample, and testing

Data from January 1, 2020, to December 31, 2023 (17,768 records), are used as the training sample. The validation sample is composed of data from January 1, 2024, to December 31, 2024 (2,230 records). The test sample includes data from January 1, 2025, to June 30, 2025 (1,342 records). Additionally, a testing period — from July 1, 2025, to October 28, 2025 (962 records) — has been designated to verify the predictive model performance.

This temporal splitting allows reducing the risk of information leakage between samples and evaluate the model ability to predict disturbance on data that was not involved in training and setting. The training sample is intended for building models, the validation sample is for selecting hyperparameters and choosing a classification threshold, the test sample is for independent comparison of the quality of algorithms, the testing sample is for testing the model under conditions close to industrial operation.

Table 1 presents the distribution of records into the “normal” and “disturbance” classes for each sample.

Table 1

Distribution of Observations by Class in the Samples

Class	Training	Validation	Test sample	Testing
Normal	16429	1902	1234	946
Disturbance	1339	328	108	16

An imbalance is evident: the number of instances of normal electrolyzer operation significantly exceeds the number of pre-failure states preceding spike detection. This distribution affects model training and evaluation. Recognizing the dominant class yields high accuracy, yet rare pre-failure states are poorly detected.

To compensate for class imbalance, class weights are applied. These increase the contribution of errors on positive class instances to the loss function and reduce model bias toward the normal operating mode. Model performance is evaluated using Precision, Recall, and F1-score metrics, which are significantly more informative than Accuracy when the target event is rare.

Computations were performed on a Windows 11 workstation equipped with GeForce RTX 4080 GPU (12 GB). The software implementation was developed in Python 3.13 using PyCharm IDE. Data processing and analysis utilized the NumPy, pandas, and scikit-learn libraries. Gradient boosting models were implemented using the XGBoost, CatBoost, and LightGBM libraries. Automated model construction was carried out using the FEDOT library. Neural network models were implemented using PyTorch, RTDL, and TabM. A fixed seed for the pseudorandom number generator was used in all computations involving stochastic operations.

Research Results. Computational experiments made it possible to select parameters that provide maximum values for quality assessment metrics. These data became the basis for the development of an intelligent predictive model. Hyperparameters with ranges and optimal values for each model type are given in Tables 2–4.

Table 2

Hyperparameters of the XGBoost Model

No.	Hyperparameter	Range	Value
1	Estimators (number of trees)	[50, 5000]	4500
2	Max depth (maximum tree depth)	[3, 25]	18
3	Learning rate (training step)	[1e-5, 1]	0.008
4	Subsample (sample proportion)	[0.5, 1]	0.701
5	Col sample by level (proportion of features per level)	[0.5, 1]	0.623
6	Col sample by tree (proportion of features per tree)	[0.5, 1]	0.714
7	Alpha (alpha — L1 regularization coefficient)	[1e-8, 10]	3.302
8	Lambda (lambda — L2 regularization coefficient)	[1e-8, 10]	8.654
9	Gamma (gamma — splitting threshold)	[1e-8, 10]	0.067

Table 3

Hyperparameters of the CatBoost Model

No.	Hyperparameter	Range	Value
1	Learning rate (training step)	[1e-5, 1]	0.031
2	Depth (tree depth)	[3, 16]	10
3	RSM (proportion of randomly selected features)	[0.5, 1]	0.752
4	L2 leaf reg (L2 regularization coefficient)	[1, 10]	2.511
5	Leaf estimation iterations (number of leaf evaluation iterations)	[1, 10]	1

Table 4

Hyperparameters of the LightGBM Model

No.	Hyperparameter	Range	Value
1	Estimators (number of trees)	[50, 5000]	4000
2	Max depth (maximum tree depth)	[3, 10]	10
3	Learning rate (training step)	[1e-5, 1]	0.011
4	Num leaves (number of leaves)	[1, 100]	75
5	Feature fraction (proportion of features)	[0.5, 1]	0.561
6	Min data in leaf (minimum amount of data in the leaf)	[10, 500]	25
7	Min child weight (minimum splitting weight)	[1e-3, 50]	0.038
8	Reg lambda1 (L1 regularization coefficient)	[1e-8, 10]	0.452
9	Reg lambda2 (L2 regularization coefficient)	[1e-8, 10]	7.022

Using the FEDOT pipeline, the structure of the algorithm is determined, including machine learning models XGBoost, Random Forest and Logistic Regression. The following hyperparameters are defined for the Random Forest model:

- criterion for splitting nodes (criterion) — “entropy”;
- proportion of features randomly selected when constructing each tree (max_features) — 0.302;
- minimum number of objects to split an internal node (min_samples_split) — 6;
- minimum number of samples per leaf node (min_samples_leaf) — 2;
- use of bootstrap sampling during tree construction (bootstrap) — true.

For the XGBoost model, the following hyperparameters are defined:

- base boosting model type (booster) — gbtrees;
- tree construction method (tree_method) — auto;
- use of categorical features (enable_categorical) — true;
- use of an evaluation sample during training (use_eval_set) — true;
- early stopping (early_stopping_rounds) — 30.

The default configuration is used for the Logistic Regression model. Additionally, normalization with standard scaling (StandardScaler) and data resampling with default parameters are used.

Table 5 presents the results of the quality assessment of the predictive models for the architectures under consideration, based on the test sample. This sample comprises monitoring data from the pilot site operations. Square brackets indicate 95% confidence intervals calculated by the bootstrap method on the test sample. These intervals allow for a comparison of the models based not only on point estimates of the metrics but also on the stability of the results obtained.

Table 5

Predictive Models Quality Assessment Results

Model	Quality assessment metrics							
	Accuracy		Precision		Recall		F1	
XGBoost	0.87	[0.85; 0.89]	0.78	[0.72; 0.83]	0.69	[0.62; 0.75]	0.72	[0.66; 0.77]
CatBoost	0.87	[0.85; 0.90]	0.77	[0.71; 0.82]	0.76	[0.70; 0.82]	0.77	[0.72; 0.82]
LightGBM	0.88	[0.86; 0.90]	0.78	[0.73; 0.84]	0.79	[0.73; 0.85]	0.79	[0.74; 0.84]
FEDOT	0.87	[0.84; 0.89]	0.80	[0.74; 0.86]	0.74	[0.68; 0.80]	0.77	[0.71; 0.82]
RTDL-RM	0.86	[0.84; 0.88]	0.76	[0.70; 0.82]	0.76	[0.69; 0.82]	0.76	[0.70; 0.81]
RTDL-NE	0.86	[0.84; 0.89]	0.76	[0.71; 0.82]	0.78	[0.72; 0.84]	0.77	[0.71; 0.82]
TabM	0.84	[0.82; 0.87]	0.74	[0.68; 0.80]	0.78	[0.71; 0.84]	0.76	[0.70; 0.81]

The analysis of the data in Table 5 allows for an assessment of the model efficiency regarding the data under study.

LightGBM shows the best overall performance among all the algorithms examined, leading in terms of Accuracy (0.88), Recall (0.79), and F1 (0.79). The high Recall values indicate an ability to efficiently detect disturbances (low rate of false negatives), which is crucial for early warning.

XGBoost and CatBoost demonstrate a comparable level of Accuracy (0.87). However, CatBoost significantly outperforms XGBoost in terms of Recall (0.76 vs. 0.69) and F1-score (0.77 vs. 0.72). This indicates that CatBoost is more reliable at detecting, whereas XGBoost misses them more frequently.

The algorithm, automatically generated using the FEDOT library, achieves Accuracy of 0.87, a result comparable to that of LightGBM. Furthermore, its Precision (0.80) is the highest among all the models evaluated, meaning it yields the fewest false positives. This fosters confidence in the generated predictions and reduces the workload on personnel who would otherwise have to respond to false warnings. In terms of the F1-score (0.77), the algorithm performs on par with CatBoost.

Ensemble-based neural network models also yield competitive results. Within this group, RTDL-NE achieves the best F1-score (0.77) and high Recall (0.78), highlighting the importance of efficient numerical feature encoding when working with tabular data. While TabM also achieves high Recall (0.78), it shows the lowest Accuracy (0.84) and Precision (0.74). This is likely due to its architecture, specifically, its parameter-efficient ensembling approach, which requires additional setting for the specific sample.

Thus, the LightGBM architecture model is optimal for predicting technological disturbances of the spike type with an initial set of monitoring data. It demonstrates the best balance between detection completeness (Recall) and overall accuracy (F1).

To validate the proposed intelligent model using real-world data from a pilot site, it was integrated into the information and analytical system for the operational control of the aluminum production process [27]. The system can issue warnings regarding potential disturbances when a specified probability threshold is exceeded. A signal is transmitted to the on-duty operator or process engineer, who — based on the current operating parameters of the electrolyzers — determines whether an inspection and corrective actions are required.

The model performance is evaluated on a validation sample through comparing predictions against actual instances of disturbances. The sample comprises 962 daily records: 946 representing the normal process state and 16 representing a pre-failure state associated with six recorded spike detection events. The evaluation results are presented in a confusion matrix (Fig. 4) showing the possible combinations of predicted and actual values.

	0	1		0	1
0 (normal)	TN true negative decision	FP false positive decision (type I error)	0 (normal)	900	46
1 (spike)	FN false negative decision (type II error)	TP true positive decision	1 (spike)	6	10

Fig. 4. Confusion matrix for the test data of the spike type disturbance predictive model

Class 0 indicates the absence of disturbance, while Class 1 indicates the presence of disturbance. If the predicted and actual classes match, the classification result is considered true; if they do not match, it is considered false. Thus, the confusion matrix displays the following results:

- true negative (TN) — no disturbance, prediction matches reality;
- true positive (TP) — disturbance, prediction matches reality;
- false positive (FP) — model predicted disturbance, but none exists in reality (Type I error);
- false negative (FN) — model predicted no disturbance, but one exists in reality (Type II error).

For the predictive model, the result is considered positive at a probability threshold of 0.4. This threshold is established based on an analysis of classification quality and represents a compromise between the completeness of disturbance detection and the acceptable number of false warnings (Table 6).

Table 6

Model Quality Assessment at Various Probability Thresholds

Quality metric	Probability threshold								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1	0.55	0.65	0.72	0.78	0.63	0.58	0.52	0.48	0.46

The confusion matrix obtained from the model testing shows that, in the category of events without disturbances (upper row of the matrix), the proposed model correctly classified the “normal” state in 95% of cases (900 observations), while in 5% of cases (46 observations) the “disturbance” state was misclassified. In the category of events with disturbances (bottom row of the matrix), the proportion of correctly recognized conditions was 62.5% (10 observations), and in 37.5% of cases (6 observations), disturbances went unnoticed.

Table 7 shows predictions for individual events during testing. The prediction results are considered during the pre-failure period — a three-day interval preceding the day the spike is detected (registered). Due to the proximity of events on the same electrolyzer, for the event 04.08.2025, only one day of the pre-failure state is taken into account.

Spike Prediction Results during Testing Period

No.	Electrolyzer No.	Spike detection date	Pre-failure date	Prediction result	
				probability	class
1	7	27.07.2025	24.07.2025	0.27	0
			25.07.2025	0.62	1
			26.07.2025	0.74	1
2	6	02.08.2025	01.08.2025	0.56	1
			31.07.2025	0.68	1
			30.07.2025	0.82	1
3	6	04.08.2025	03.08.2025	0.74	1
4	8	11.08.2025	08.08.2025	0.31	0
			09.08.2025	0.57	1
			10.08.2025	0.76	1
5	5	31.08.2025	28.08.2025	0.25	0
			29.08.2025	0.65	1
			30.08.2025	0.72	1
6	5	12.09.2025	09.09.2025	0.18	0
			10.09.2025	0.26	0
			11.09.2025	0.30	0

During the testing of the LightGBM model, identified as optimal for prediction, six instances of spike detection were recorded. The model detected signs of the developing disturbance in five of these cases, representing a success rate of 83.3%. The lead time for the detected disturbances ranged from one to three days: in one instance, the warning was issued three days prior to the recorded disturbance; in three instances, two days prior; and in one instance, one day prior. The lead time horizon was defined as the maximum number of days between the initial positive prediction and the date the spike was recorded.

Discussion. Thus, the results of comparing predicted values with recorded disturbances in the electrolyzers of the pilot site confirm the model ability to detect spike type process prior to their actual registration, with a practical early prediction horizon of 24–72 hours.

The results are consistent with the physical and process characteristics associated with the onset of a spike type disturbance [5]. Several days may elapse between the initial formation of the defect and a pronounced change in the electrolyzer operating mode. Changes in key parameters are observed by the second or third day [22]. Therefore, the prediction horizon of 24 to 72 hours adopted in this study should be considered a technologically justified interval, within which changes in monitored parameters can serve as early indicators of the disturbance development.

Furthermore, the results of the model validation, consistent with [16], confirmed the feasibility of predicting disturbance several days prior to a critical change in the process state. However, there is a distinction: whereas the prediction in [16] was based on individual anode currents, the present study achieved a warning with a comparable lead time (“several days before the spike becomes evident”) without utilizing a specialized current measurement system. The solution proposed by the authors can be viewed as a supplement to methods for the local monitoring of anodes and electrolyzer conditions based on the analysis of parameters from daily operational monitoring.

The validation results reflect known challenges associated with applying predictive models in industrial electrolysis, including the difficulty of event recording [17], class imbalance, and shifts in statistical characteristics [15]. The decline in prediction performance when moving from the test sample to industrial-scale validation may also be attributed to the variability of the actual process and the rarity of the specific type of disturbances under consideration. Thus, the share of observations of a pre-failure state in the testing sample is less than 2%. Therefore, even with a high percentage of correctly recognized normal states, a relatively small number of false positive decisions significantly reduces the predictive value of the model.

False positive warnings are partly attributable to the human factor. For instance, when signs of spike formation appear, operators move (jerk) the anodes and check for and clear non-breakthroughs at the alumina feed points. All this significantly changes the dynamics of the parameters, prevents the development of the disturbance and its actual registration. In such cases, the model identifies signs of a pre-failure state, but the event is not included in the violation log, and the warning is classified as a false positive. This interpretation requires separate verification and comparison of the time of warning generation with records of process operations and subsequent dynamics of controlled parameters.

False negative results are attributed to:

- absence of sufficiently pronounced precursors of the disturbance in the daily average parameters;
- individual characteristics of how the disturbance develops in specific electrolyzers;
- accuracy of the operational classification of disturbances, given that the moment an event is recorded does not always coincide with the actual onset of its development.

Therefore, despite the use of temporal data splitting, external temporal validation of the model is required.

The testing results are significantly affected by changes in operational conditions. Over time, control settings, anode conditions, raw material composition, anode feeding modes, and maintenance protocols change. This means that the feature distributions differ from those present during model training. Furthermore, the study is based on monitoring data from a small number of electrolyzers at a pilot site. Confirming the model generalization capability requires additional verification across other production sites and time periods. Therefore, despite the use of temporal data splitting, external temporal testing of the model is required.

From a practical standpoint, the results enable the use of an intelligent model for the early prediction of spike type disturbances in real-world industrial settings. This creates a time buffer for additional verification and the implementation of preventive process adjustments. It should be noted that overall classification accuracy alone is insufficient for the deployment of intelligent systems. The balance between model sensitivity and the workload placed on process personnel in the case of false-positive and false-negative decisions plays a critical role. The importance of logging warnings and probability values, along with personnel actions and confirmed disturbance events, should be emphasized. This will make it possible to monitor prediction quality and the false warning rate, refine the model activation threshold, and build a representative sample of confirmed cases for retraining.

Conclusion. An intelligent model is proposed for the early prediction of spike type process disturbances in aluminum production. The study involved compiling a set of process data from eight RA-550 electrolyzers covering the period from 2020 to 2025. Preprocessing of the monitoring data was performed, including data cleaning, imputation of missing values, labeling, and splitting the data into subsets while accounting for the temporal structure of the production process.

Ensemble methods were used to predict the spike: gradient boosting (XGBoost, CatBoost, LightGBM), the automated FEDOT pipeline, and neural network ensembles (RTDL-RM, RTDL-NE, TabM). Based on the results of computational experiments, model hyperparameters yielding the highest quality metric values were determined. LightGBM proved to be the most efficient, achieving the best metrics for Accuracy (0.88), Recall (0.79), and F1 (0.79).

The proposed intelligent model was integrated into an information-analytical system for the real-time monitoring of aluminum production and validated using data from an actual technological process. A comparison of predictions with actual process disturbances showed that, at a sensitivity threshold of 0.4, the model correctly classified the “normal” state in 95% of cases. The proportion of correctly identified “disturbance” states was 63%. The false positive rate was 5%, and the false negative rate was 37.5%. False results arise, in particular, from personnel actions aimed at preventing a disturbance from developing before it is officially registered. Other causes include incorrect production labeling of disturbances, changes in process conditions, and the limited size of the sample.

The study has shown that the model can detect process disturbances 24–72 hours before they are actually recorded.

Testing results and expert assessments by process engineers confirm the model sufficient efficiency and applicability under actual operating conditions.

Future research will involve expanding the training sample, validating the model for other sites of aluminum production, and conducting a comprehensive analysis of average daily and instantaneous monitoring parameters to pinpoint the origin of the disturbance.

References

1. Sizyakov VM, Polyakov PV, Bazhin VYu. Current Trends and Strategic Objectives in the Production of Aluminum and Its Alloys in Russia. (In Russ.) *Non-ferrous Metals*. 2022;7:16–23. <https://www.rudmet.ru/journal/2131/article/35492> (accessed: 04.08.2026).
2. Heli Liu, Dhawan Saksham, Merrill Shen, Kangan Chen, Vincent Wu, Liliang Wang. Industry 4.0 in Metal Forming Industry Towards Automotive Applications: A Review. *International Journal of Automotive Manufacturing and Materials*. 2022;1(1):16–27. URL: <https://www.sciltp.com/journals/ijamm/articles/2504000083> (accessed: 04.08.2026).
3. Bochkaryov PYu, Korolev RD, Bokova LG. Comprehensive Assessment of the Manufacturability of Products. *Advanced Engineering Research (Rostov-on-Don)*. 2023;23(2):155–168. <https://doi.org/10.23947/2687-1653-2023-23-2-155-168>
4. Sadler B. Critical Issues in Anode Production and Quality to Avoid Anode Performance Problems. *Journal of Siberian Federal University. Engineering & Technologies*. 2015;8(5):546–568. <https://doi.org/0.17516/1999-494X-2015-8-5-546-568>
5. Mikhalev YG, Polyakov PV, Yasinskiy AS, Polyakov AA. Spikes Generation on Anode of Aluminium Reduction Cell. *Non-ferrous metals*. 2018;9:43–48. <https://doi.org/10.17580/tsm.2018.09.06>
6. Polyakov PV, Vlasov AA, Mikhalev YuG, Yanov VV. On Cone Formation on Burnt Anode Face in Aluminum Electrolyzers. *Metallurgist*. 2017;60(9/10):1087–1093. <https://doi.org/10.1007/s11015-017-0411-2>
7. Penkova T, Senashova M, Korobko A. Multidimensional Analysis of Aluminum Production Monitoring Data in Basic Operation Modes. *CEUR Workshop Proceedings*. 2020;2727:128–136. URL: <https://ceur-ws.org/Vol-2727/paper17.pdf> (accessed: 04.08.2026).
8. Belousova NV, Sharypov NA, Shakhrai SG, Bezrukikh AI. Coal Foam in an Aluminum Electrolyzer: Problems and Some Solutions. *Non-ferrous Metals*. 2017;8:43–49. <https://doi.org/10.17580/tsm.2017.08.06>
9. Metus A, Penkova T. Analysis of Aluminium Electrolysis Data in the Context of Extreme Values of Technological Parameters. *CEUR Workshop Proceedings*. 2020;2727:92–98. URL: <https://ceur-ws.org/Vol-2727/paper12.pdf> (accessed: 04.08.2026).
10. Zhi-Hua Zhou. *Machine Learning*. Singapore: Springer; 2021. 460 p. URL: <https://link.springer.com/content/pdf/bfm:978-981-15-1967-3/1?pdf=chapter+toc> (accessed: 15.07.2026).
11. Tugashova LG, Zatonitskiy AV. Machine Learning-Based Condition Assessment Method for Shell-and-Tube Heat Exchangers to Improve Energy Efficiency. *Advanced Engineering Research (Rostov-on-Don)*. 2026;26(2):2237. <https://doi.org/10.23947/2687-1653-2026-26-2-2237>
12. Cheng Ji, Wei Sun. A Review on Data-Driven Process Monitoring Methods: Characterization and Mining of Industrial Data. *Processes*. 2022;10(2):335. <https://doi.org/10.3390/pr10020335>
13. Muntin AV, Zhikharev PYu, Ziniagin AG, Brayko DA. Artificial Intelligence and Machine Learning in Metallurgy. Part 1. Methods and Algorithms. *Metallurgist*. 2023;67(5/6):886–894. <https://doi.org/10.1007/s11015-023-01576-3>
14. Zhikharev PYu, Muntin AV, Brayko DA, Kryuchkova MO. Artificial Intelligence and Machine Learning In Metallurgy. Part 2. Application Examples. *Metallurgist*. 2024;67:1545–1560. <https://doi.org/10.1007/s11015-024-01648-y>
15. Abd Majid NA, Taylor MP, Chen JJ, Young BR. Aluminium Process Fault Detection and Diagnosis. *Advances in Materials Science and Engineering*. 2015;2015:1–11. <https://doi.org/10.1155/2015/682786>
16. Martel A. Spike Detection Using Advanced Analytics and Data Analysis. In book: Martin O. (ed) *Light Metals 2018*. Cham: Springer; 2018. P. 485–490. https://doi.org/10.1007/978-3-319-72284-9_64
17. Faraj M, Sayed K, Al Hosani A, Bakuteev A, Shyamala M, Pervez K, et al. Anode Spike Model – A Case Study of Challenges and Future Directions. In: *TRAVAUX 53. Proc. 42nd International ICSOBA Conference*. Lyon: ICSOBA; 2024. P. 1553–1573. URL: <https://icsoba.org/proceedings/42nd-conference-and-exhibition-icsoba-2024/?doc=131> (accessed: 04.08.2026).
18. Gang Yan, Ximing Liang. Predictive Models of Aluminum Reduction Cell Based on LS-SVM. In: *International Conference on Digital Manufacturing & Automation*. New York City: IEEE; 2010. P. 99–102. <https://doi.org/10.1109/ICDMA.2010.12>
19. Mikhalev A, Lugovaya N, Penkova T, Puzanov I, Zavadyak A. Application of Ensemble Algorithms to Detect Anode Effects in Aluminum Production. *CEUR Workshop Proceedings*. 2021;3047:79–85. URL: <https://ceur-ws.org/Vol-3047/paper11.pdf> (accessed: 04.08.2026).

20. Nazatul Aini Abd Majid, Mark P Taylor, John JJ Chen, Marco A Stam, Albert Mulder, Brent R Young. Aluminium Process Fault Detection by Multiway Principal Component Analysis. *Control Engineering Practice*. 2011;19(4):367–379. <https://doi.org/10.1016/j.conengprac.2010.12.005>
21. Puzanov II, Zavadyak AV, Klykov VA, Makeev AV, Plotnikov VN. Continuous Monitoring of Information on Anode Current Distribution as Means of Improving the Process of Controlling and Forecasting Process Disturbances. *Journal of Siberian Federal University. Engineering & Technologies*. 2016;9(6):788–801. <https://doi.org/10.17516/1999-494X-2016-9-6-788-801>
22. Polyakov PV, Sharypova NA, Osipova VA, Pinykh AA. Mathematical Modeling of Current Distribution in the Presence of Abnormalities on the Reduction Cell Anode Bottom. *Non-ferrous Metals*. 2019;913(1):25–30. <https://doi.org/10.17580/tsm.2019.01.04>
23. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45(3):1–67. <https://doi.org/10.18637/jss.v045.i03>
24. Bentéjac C, Csörgő A, Martínez-Muñoz G. A Comparative Analysis of Gradient Boosting Algorithms. *Artificial Intelligence Review*. 2021;54(3):1937–1967. <https://doi.org/10.1007/s10462-020-09896-5>
25. Fort S, Huiyi Hu, Lakshminarayanan B. Deep Ensembles: A Loss Landscape Perspective. *ArXiv preprint*. 2019:2. <https://doi.org/10.48550/arXiv.1912.02757>
26. Hossin M, Sulaiman MN. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. 2015;5(2):1–11. <https://doi.org/10.5121/ijdkp.2015.5201>
27. Zavadyak AV, Nozhenkova LF, Puzanov II, Penkova TG, Korobko AA, Korobko AV, et al. *Intelligent Support Tools for Managing the Detection of Process Disturbances and Assessing the Process State of an Aluminum Production Complex*. Certificate of Software State Registration No. 2021662399, 2021. 1 p. (In Russ.)

About the Authors:

Anton S. Mikhalev, First-category Programmer, Institute of Computational Modeling SB RAS (50/44, Akademgorodok, Krasnoyarsk, 660036, Russian Federation), [SPIN-code](#), [ORCID](#), [ScopusID](#), [ResearcherID](#), [ResearchGate](#), asmikhalev@yandex.ru

Tatiana G. Penkova, Cand.Sci. (Eng.), Associate Professor, Head of the Applied Informatics Department, Senior Research Fellow, Institute of Computational Modeling SB RAS (50/44, Akademgorodok, Krasnoyarsk, 660036, Russian Federation), [SPIN-code](#), [ORCID](#), [ScopusID](#), [ResearcherID](#), [ResearchGate](#), penkova_tg@icm.krasn.ru

Ludmila F. Nozhenkova, Dr.Sci. (Eng.), Professor, Chief Research Fellow, Institute of Computational Modeling SB RAS (50/44, Akademgorodok, Krasnoyarsk, 660036, Russian Federation), [SPIN-code](#), [ORCID](#), [ScopusID](#), [ResearcherID](#), expert@icm.krasn.ru

Claimed Contributorship:

AS Mikhalev: formal analysis, software, validation.

TG Penkova: investigation, methodology, writing – original draft preparation.

LF Nozhenkova: conceptualization, supervision.

Conflict of Interest Statement: the authors declare no conflict of interest.

All authors have read and approved the final manuscript.

AI Declaration: the authors declare that generative artificial intelligence (AI) technologies were not used in the preparation of the manuscript and that AI is not a co-author of the publication: authorship is assigned only to natural persons who have made a substantial intellectual contribution to the research.

Об авторах:

Антон Сергеевич Михалев, программист 1-й категории, Институт вычислительного моделирования Сибирского отделения Российской академии наук (660036, Российская Федерация, г. Красноярск, ул. Академгородок, 50/44), [SPIN-код](#), [ORCID](#), [ScopusID](#), [ResearcherID](#), [ResearchGate](#), asmikhalev@yandex.ru

Татьяна Геннадьевна Пенькова, кандидат технических наук, доцент, заведующий отделом «Прикладная информатика», старший научный сотрудник, Институт вычислительного моделирования Сибирского отделения Российской академии наук (660036, Российская Федерация, г. Красноярск, ул. Академгородок, 50/44), [SPIN-код](#), [ORCID](#), [ScopusID](#), [ResearcherID](#), [ResearchGate](#), penkova_tg@icm.krasn.ru

Людмила Федоровна Ноженкова, доктор технических наук, профессор, главный научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук (660036, Российская Федерация, г. Красноярск, ул. Академгородок, 50/44), [SPIN-код](#), [ORCID](#), [ScopusID](#), [ResearcherID](#), expert@icm.krasn.ru

Заявленный вклад авторов:

А.С. Михалев: формальный анализ, разработка программного обеспечения, валидация результатов.

Т.Г. Пенькова: проведение исследования, разработка методологии, написание черновика рукописи.

Л.Ф. Ноженкова: разработка концепции, научное руководство.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Все авторы прочитали и одобрили окончательный вариант рукописи.

Декларация ИИ: авторы заявляют, что при подготовке текста рукописи не использовались технологии генеративного искусственного интеллекта (ИИ), ИИ не является соавтором публикации: авторство присвоено только физическим лицам, внёсшим существенный интеллектуальный вклад в исследование.

Received / Поступила в редакцию 25.05.2026

Reviewed / Поступила после рецензирования 30.06.2026

Accepted / Принята к публикации 16.07.2026