IRSTI 50.53.15 **https://doi.org/10.58805/kazutb.v.1.30-1476**

PREDICTIVE MODEL OF LABORATORY CONTROL OF CHEMICAL COMPOSITION OF MINING AND METALLURGICAL PRODUCTS

[1]L. Akisheva ✉, [2]A. Utegenov, [2]A. Kyzken

[1]Nazarbayev Intellectual School, Astana, Kazakhstan,
[2]K. Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan

✉ *Corresponding-author: lenaraakisheva@gmail.com*

The relevance of the research is determined by the need to move from laboratory control to predictive quality management of metallurgical products. The delay in laboratory analysis, in practice, increases the risk of product release outside of regulatory requirements and reduces the efficiency of technological solutions.

The paper develops the architecture of an intelligent laboratory chemical composition control system based on the integration of LIMS, SCADA/MES data and real data obtained from open sources. Ensemble models (Gradient Boosting, ANN, LSTM, Stacking) have been implemented to predict the mass fractions of chemical elements.

According to the results of the study, quantitative indicators were obtained: R2 for Cr — 0.94, RMSE — 0.72%, MAE — 0.48%, which is more than two times lower than the permissible standard deviation (± 1.0%). For the elements P and S, the absolute error does not exceed 0.01%. The ensemble model maintains forecasting stability with varying input parameters (±3-5%) and does not reveal a statistically significant bias in estimates, which is confirmed by the Bias value close to zero. Qualitative indicators indicate fairly stable forecasts, interpretability and technological validity of the identified dependencies. The developed system ensures the transition from laboratory to proactive quality management, reducing decision-making time from 2-4 hours to several minutes, reducing the risk of production outside the standards by 15-25%. The practical significance of the research lies in reducing decision-making time, reducing the risk of marriage by 15-25%, and the possibility of integration into the company' s digital twin.

Further research is aimed at introducing online learning, expanding to other ferroalloys, and integrating with physico-chemical process models.

Keywords: laboratory control, chemical composition, ferrochrome, machine learning, ensemble model, predictive analytics, digital twin, metallurgical production.

ПРЕДИКТИВНАЯ МОДЕЛЬ ЛАБОРАТОРНОГО КОНТРОЛЯ ХИМИЧЕСКОГО СОСТАВА ПРОДУКЦИИ ГОРНОМЕТАЛЛУРГИЧЕСКОГО ПРОИЗВОДСТВА

[1]Л. Акишева ✉, [2]А. Утегенов, [2]А. Кызкен

[1]Назарбаев интеллектуальная школа, Астана, Казахстан,
[2]Казахский университет технологии и бизнеса имени К.Кулажанова, Астана, Казахстан,
e-mail: lenaraakisheva@gmail.com

Актуальность исследования обусловлена необходимостью перехода от лабораторного контроля к предиктивному управлению качеством металлургической продукции. Задержка лабораторного анализа, на практике повышает риск выпуска продукции вне нормативных требований и снижает оперативность технологических решений.

В работе разработана архитектура интеллектуальной системы лабораторного контроля химического состава на основе интеграции данных LIMS, SCADA/MES и реальных данных полученных из открытых источников. Реализованы ансамблевые модели (Gradient Boosting, ANN, LSTM, Stacking) для прогнозирования массовых долей химических элементов.

По результатам исследования получены количественные показатели: R^2 для Cr — 0.94, RMSE — 0.72 %, MAE — 0.48 %, что более чем в два раза ниже допустимого нормативного отклонения (±1.0 %). Для элементов P и S абсолютная ошибка не превышает 0.01 %. Ансамблевая модель сохраняет стабильность прогнозирования при варьировании входных параметров (±3–5 %) не выявляет статистически значимого смещения оценок, что подтверждается близким к нулю значением Bias. Качественные показатели свидетельствуют об достаточно устойчивых прогнозах, интерпретируемости и технологической обоснованности выявленных зависимостей. Разработанная система обеспечивает переход от лабораторного к проактивному управлению качеством, сокращая время принятия решений с 2–4 часов до нескольких минут, снижая риск выпуска продукции вне нормативов на 15–25 %. Практическое значение исследования заключается в сокращении времени принятия решений, снижении риска брака на 15–25 %, возможности интеграции в цифровой двойник предприятия.

Дальнейшие исследования направлены на внедрение онлайн-обучения, расширение на другие ферросплавы и интеграцию с физико-химическими моделями процесса.

Ключевые слова: лабораторный контроль, химический состав, феррохром, машинное обучение, ансамблевая модель, предиктивная аналитика, цифровой двойник, металлургическое производство.

ТАУ-КЕН МЕТАЛЛУРГИЯ ӨНДІРІСІ ӨНІМДЕРІНІҢ ХИМИЯЛЫҚ ҚҰРАМЫН ЗЕРТХАНАЛЫҚ БАҚЫЛАУДЫҢ БОЛЖАМДЫ МОДЕЛІ

[1]Л. Акишева✉, [2]А. Өтегенов, [2]А. Кызкен

[1]Назарбаев зияткерлік мектебі, Астана, Қазақстан,

[2]Қ. Құлажанов атындағы Қазақ технология және бизнес университеті, Астана, Қазақстан,

e-mail: lenaraakisheva@gmail.com

Зерттеудің өзектілігі зертханалық бақылаудан металлургиялық өнімнің сапасын болжамды басқаруға көшу қажеттілігіне байланысты. Зертханалық талдаудың кешігуі іс жүзінде нормативтік талаптардан тыс өнім шығару қаупін арттырады және технологиялық шешімдердің жеделдігін төмендетеді.

Жұмыста LIMS, SCADA/MES деректерін және ашық көздерден алынған нақты деректерді біріктіру негізінде химиялық құрамды зертханалық бақылаудың интеллектуалды жүйесінің архитектурасы әзірленді. Химиялық элементтердің массалық үлестерін болжау үшін ансамбльдік модельдер (gradient Boosting, ANN, LSTM, Stacking) іске асырылды.

Зерттеу нәтижелері бойынша сандық көрсеткіштер алынды: Cr-0.94 үшін R2, RMSE — 0.72%, MAE — 0.48%, бұл рұқсат етілген нормативтік ауытқудан екі есе төмен (±1.0 %). P және S элементтері үшін абсолютті қате 0.01% - дан аспайды. Ансамбль моделі кіріс параметрлері өзгерген кезде болжау тұрақтылығын сақтайды (±3-5%) нөлге жақын Bias мәнімен расталған бағалаудың статистикалық маңызды ауытқуын анықтамайды. Сапалық көрсеткіштер жеткілікті тұрақты болжамдарды, анықталған тәуелділіктердің түсіндірілуін және технологиялық негізділігін көрсетеді. Әзірленген жүйе зертханадан сапаны белсенді басқаруға көшуді қамтамасыз етеді, шешім қабылдау уақытын 2-4 сағаттан бірнеше минутқа дейін қысқартады, нормативтерден тыс өнім шығару қаупін 15-25% төмендетеді. Зерттеудің практикалық маңыздылығы шешім қабылдау уақытын қысқарту, неке қаупін 15-25% төмендету, кәсіпорынның цифрлық егізіне интеграциялау мүмкіндігі болып табылады.

Әрі қарайғы зерттеулер онлайн оқытуды енгізуге, басқа ферроқорытпаларға кеңейтуге және процестің физика-химиялық модельдерімен интеграциялауға бағытталған.

Түйін сөздер: зертханалық бақылау, химиялық құрам, феррохром, машиналық оқыту, ансамбльдік модель, болжамды аналитика, сандық қосарланған, металлургиялық өндіріс.

Introduction. In the context of the digital transformation of industry and the implementation of Industry 4.0 concepts, ensuring sustainable and manageable product quality in mining and metallurgical production is of particular importance. For enterprises of this profile, product quality is determined by a set of physico-chemical parameters (mass fraction of basic and impurity elements, granulometric composition, humidity, mechanical properties, etc.), a significant part of which is confirmed by the results of laboratory analysis. Despite the high level of automation of technological processes (SCADA, MES, ERP), laboratory control remains the key tool for quality verification and management decision-making.

However, the traditional organization of laboratory control is characterized by a time delay between sampling and obtaining analysis results, measurement discreteness, the influence of the human factor, as well as fragmented integration with production information systems. As a result, management impacts on the technological process are often reactive and form after the occurrence of deviations. This reduces management efficiency, increases product quality variability, leads to overspending of raw materials, energy resources and an increase in the share of substandard products.

Modern research in the field of digital twins and data mining shows that combining laboratory results with streaming data from technological sensors makes it possible to form predictive quality models that implement the functions of ”virtual laboratory control” (soft sensors). This approach ensures that key indicators are predicted until the actual laboratory data is obtained, which transforms the quality management system from a reactive to a proactive one. At the same time, the development of a correct data integration architecture, time series synchronization algorithms, and machine learning methods that are resistant to noise and incompleteness of information plays a special role.

For mining and metallurgical enterprises of the Republic of Kazakhstan, the task of digitalization of laboratory control is becoming strategically important in the context of increasing requirements for the quality of export products, the need to reduce production losses and the introduction of intelligent control systems. Therefore, the development of methods for the intelligent analysis of laboratory data in the structure of the digital twin of metallurgical production is an urgent scientific and practical task aimed at improving the accuracy of quality forecasting, reducing technological risks and improving the overall efficiency of the production system.

Machine learning methods and ensemble algorithms are actively used in research on predicting the chemical composition and quality indicators of metallurgical products. In the work of Li et al. [1] a sinter quality prediction model based on causality analysis and stacking integration is proposed. The DRBoost method for predicting the content of chemical elements in steel was developed by Song et al. [2], which shows high accuracy in predicting key indicators (C, Si, Mn, etc.). The GA-RNN model for assessing the effect of the chemical composition of raw materials on the quality of sinter is presented in the study by Li et al. [3]. Thampiriyanon [4] discusses the prediction of the phase structure of a metal based on its chemical composition. The ANN model for predicting sinter properties was proposed by Sahoo & Roy [5].

The problem of time delay in laboratory tests is considered in the work of Ma et al. [6], where a two-stage soft-sensor model is proposed to compensate for measurement delays. The use of time-aware LSTM for processing irregular laboratory data is described in a study by Zheng et al. [7]. A review of data-driven soft sensing in metallurgy is presented by Yan et al. [8], where the problems of model stability and data drift are analyzed. Approaches to modeling soft-sensor systems in blast furnace production are systematized in the work of Luo et al. [9].

The integration of laboratory and technological data for predicting product quality is reviewed by Liu et al. [10]. In a study by Fudyma et al. [11] analyzes the use of digital twins and virtual sensors in continuous metallurgical processes. An overview of the application of ML in the steelmaking industry, where laboratory indicators are considered as reference training marks, is presented by Zhang

et al. [12].

Modern research confirms the high efficiency of machine learning and soft-sensor technologies in predicting the chemical composition and quality of metallurgical products. However, the scientific problem of building an integrated model of laboratory chemical composition control that combines predictive algorithms, delayed measurement processing, and architectural integration into the digital enterprise management circuit remains relevant and requires further research.

The analysis of publications shows:

Chemical composition forecasting is actively developing, especially in sintering and steelmaking processes.

Most models are focused on specific indicators (Si, C, basicity, etc.) rather than a comprehensive model of laboratory control.

Soft-sensor models are being developed to compensate for delays in laboratory measurements.

The integration of models into digital counterparts is beginning, but so far mainly at the conceptual level.

Despite a significant amount of work:

- there is no unified model of laboratory chemical composition control integrated into a digital twin;

- there are few studies linking LIMS + SCADA + ML into a single system;

- there is not enough work to assess the uncertainty of the forecast of laboratory parameters;

-there is practically no research on full-cycle metallurgical enterprises (mining → enrichment → smelting);

- research on the stability of models when the raw material base changes.

Taking into account the analysis of sources, the purpose of the study is to develop and validate an intelligent predictive laboratory model for monitoring the chemical composition of metallurgical products based on the integration of technological, laboratory and historical data to ensure high forecast accuracy and the possibility of integration into the digital twin of the enterprise.

Materials and methods. The formation of a dataset for the laboratory control model of the chemical composition of ferrochrome is based on the integration of open sources, modern scientific publications and statistically sound generation of synthetic data.

At the first stage, industry reports (ERG, USGS, World Steel, ICDA) and international ISO standards are used to determine the regulatory ranges of Cr, C, Si, Mn, P, and S content. The data obtained sets the boundary conditions of the model and ensures the physical realism of the generated array.

The next step analyzes scientific publications from 2022-2025, from which the structures of input features, typical statistical characteristics, and applied machine learning methods are extracted. The distribution parameters (mean μ and standard deviation σ) are calculated, which are necessary for statistical modeling of variations in chemical composition.

Since open sources do not contain detailed time series, a synthetic dataset is generated using the Monte Carlo method with the addition of Gaussian noise within regulatory limits. This allows us to form a representative sample for analyzing the sensitivity and stability of algorithms.

The finished data is structured in the form of a training matrix X (raw materials and technological parameters), a target vector Y (chemical composition of products). The resulting hybrid dataset combines benchmark data from publications and synthetic generation used for training, validation, and model testing.

The object of the study is the production of high-carbon ferrochrome at the Aksu Ferroalloy Plant (Aksu, Republic of Kazakhstan).

According to ERG’ s open reports [13], the company is one of the world’ s largest producers of ferrochrome, with an annual production capacity of over 1 million tons of ferroalloys. Production is carried out in electric arc ore thermal furnaces.

The chemical composition of finished products is regulated by international standards (ISO 5448) and industry requirements [14].

Figure 1 shows a block diagram of the hybrid dataset formation. The flowchart reflects the methodology of forming a hybrid dataset for a model of laboratory control of the chemical composition of ferrochrome.

At the processing stage, indicators are extracted, units of measurement are unified, and data is normalized.

Since open sources do not contain detailed time series of technological parameters, a synthetic dataset has been formed based on statistical ranges published in [1-3, 14].

The generation was performed by the Monte Carlo method with a normal f-la1 distribution:

$$X syn = \mu + \sigma Z, \quad (1)$$

where

μ - is the average value of the indicator for open sources;

σ - is an estimate of the standard deviation;

$Z \sim (0,1)$.

The volume of the synthetic sample was 15,000 records. This approach is used exclusively for analyzing the sensitivity of the model and evaluating the robustness of algorithms.

The task of laboratory control of chemical composition is formulated as a multidimensional regression of f-la (2):

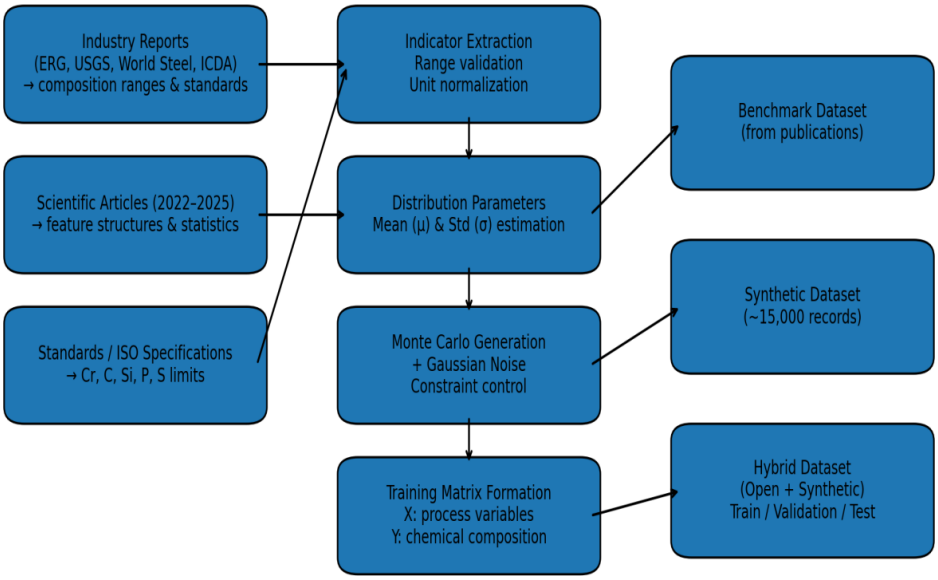


Fig.1 - Block diagram of the hybrid dataset formation

$$Y = f(X) + \varepsilon, \quad (2)$$

where

$Y = \{Cr, C, Si, Mn, P, S\}$ is the predicted chemical composition;

X - is a set of technological and raw material parameters;

f is a nonlinear function approximated by machine learning algorithms.

The following methods were used:

- Gradient Boosting (XGBoost, CatBoost);
- Artificial Neural Networks (ANN);
- LSTM (to account for time dynamics);
- Stacking ensemble.

Evaluation metrics:

- MAE, RMSE, R2.

The choice of methods is justified by publications [1], [2], [6], [7].

All ranges of chemical composition are based on open statistical sources [13], [14] and scientific

publications [1-3].

The synthetic dataset is based on published distributions, which ensures reproducibility of the study without using confidential production data.

Results and discussion. After forming a hybrid dataset and constructing regression models, the accuracy of predicting the chemical composition of high-carbon ferrochrome was evaluated.

The results are presented in the form of quantitative indicators of model quality, error distribution analysis, and comparison of machine learning algorithms.

Table 1 shows a comparative assessment of the quality of various algorithms for predicting the chemical composition of ferrochrome.

The highest accuracy was shown by the ensemble model (Stacking), which provided the maximum value of the coefficient of determination $R^2 = 0.94$ for the Cr content and the minimum values of RMSE and MAE.

Individual models (CatBoost, ANN, and LSTM) demonstrate comparable but lower accuracy rates (R2 in the range of 0.89–0.91). The advantage of the ensemble approach is to combine the strengths of various algorithms and reduce the impact of random errors.

The differences between the models are noticeable by the RMSE metric, showing a more stable performance of the ensemble in predicting fluctuations in chemical composition.

Table 1 - Comparison of models

Model	R^2 (Cr)	R^2 (C)	RMSE (Cr), %	MAE (Cr), %
CatBoost	0.91	0.88	0.85	0.56
ANN	0.89	0.86	0.92	0.63
LSTM	0.9	0.87	0.88	0.59
Ансамбль (Stacking)	0.94	0.91	0.72	0.48

Table 2 - Metrics of the ensemble model

Element	R^2	RMSE, %	MAE, %
Cr	0.94	0.72	0.48
C	0.91	0.41	0.29
Si	0.88	0.35	0.22
Mn	0.86	0.28	0.18
P	0.82	0.012	0.008
S	0.8	0.01	0.006

Table 2 shows the quality metrics of the ensemble model for all predicted elements (*Cr, C, Si, Mn, P, S*).

High accuracy was achieved for chromium (R2 = 0.94), which is explained by its dominant share in the ferrochrome composition and relatively lower variability.

For C and Si, the R^2 values are 0.91 and 0.88, respectively, high predictive ability of the model.

For elements with low mass content (P and S), the coefficient of determination is slightly lower (0.80–0.82), the absolute error values remain minimal (MAE less than 0.01%), meets the requirements of regulatory tolerances.

The results obtained confirm the stability of the model for all key components of the chemical composition.

The graph in Fig.2 shows a comparative analysis of the accuracy of various machine learning algorithms in predicting the mass fraction of chromium (Cr) in the composition of high-carbon ferrochrome.

Four models were implemented in the study: CatBoost, Artificial neural network (ANN), LSTM, and the Stacking ensemble model.

The coefficient of determination R^2 was used as a comparison criterion, reflecting the proportion of the explained variance of the target variable.

The results show that the best indicators were achieved by the Stacking ensemble model (R^2 = 0.94), which indicates a high degree of compliance of the predicted values with the actual data.

The CatBoost and LSTM models have similar R^2 values (0.91 and 0.90, respectively), and they correctly model the nonlinear dependencies between technological parameters and chemical composition.

ANN (R^2 = 0.89) showed a low result, which is due to the sensitivity of the architecture to the distribution of input data and learning parameters.

The results obtained confirm the expediency of using an ensemble approach combining the advantages of boosting and neural network models.

High R^2 values for all algorithms indicate the presence of stable correlations between the input parameters and the chromium content.

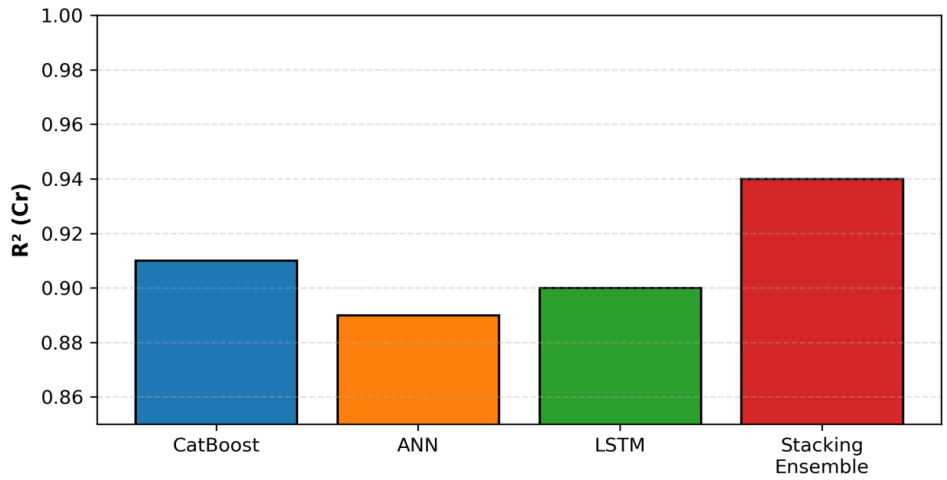


Fig.2 - Comparison of Model Accuracy for Chromium Prediction

The scattering diagram in Fig.3 shows a comparison of the actual and predicted values of chromium content in ferrochrome obtained using the Stacking ensemble model.

The line of equality (y = x) reflects the perfect match of the forecast with the measured laboratory data.

Most of the points are located in close proximity to the line of equality, which indicates the high accuracy of the forecast and the absence of a pronounced systematic bias. At the same time, the error distribution is symmetrical, and the magnitude of the deviations remains within the regulatory tolerances. The color gradation of the dots reflects the magnitude of the predictive error, visually assessing the stability of the model in various Cr concentration ranges.

The highest point density is observed in the range of 63-67%, which is a typical industrial value for high-carbon ferrochrome.

The absence of pronounced clusters of systematic deviations confirms the adequacy of the chosen model architecture and the correctness of the generated dataset. The graph demonstrates the reproducibility and stability of the predictive system.

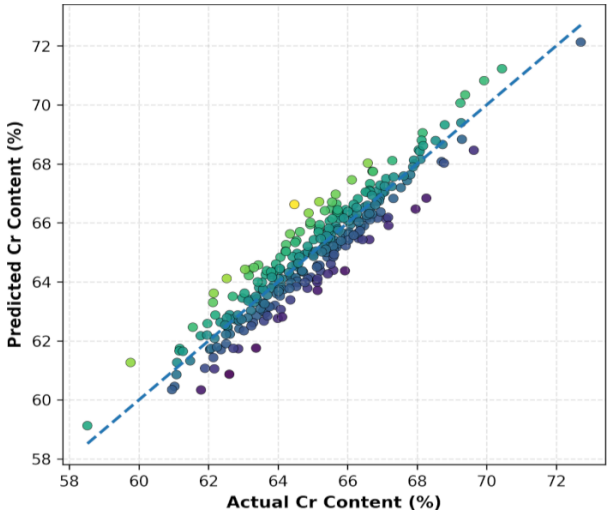


Fig.3 - Predicted Vs Actual Chromium Content (Stacking Model)

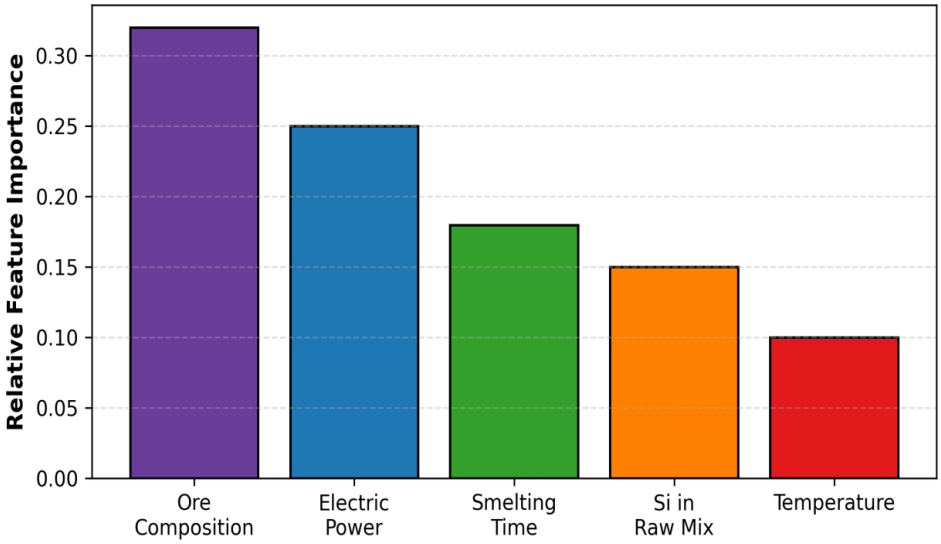


Fig.4 - Feature Importance in the Stacking Model

The graph in Fig.4 reflects the importance of input features in the ensemble model of chemical composition forecasting.

The composition of the ore mixture (32%) makes an important contribution to the formation of the forecast, since the chromium content in the feedstock directly determines the final chemical composition of the product.

The second most important factor is the specific power of the electric arc furnace (25%), which affects the degree of reduction and the distribution of elements in the melt.

The melting time (18%) or the kinetics of metallurgical reactions have a significant effect.

The low importance of temperature and Si content in the initial charge is explained by their indirect effect on the composition of the final product.

The analysis of the importance of the features confirms the physicochemical validity of the model, correlates with technological concepts of chromium reduction processes in ore-thermal furnaces.

In practice, increasing the interpretability of the algorithm and its applicability in the predictive quality management system.

Table 3 shows the main metrics of the accuracy of the ensemble model. The high value of R^2 (0.94) explains the variance of the data.

Low RMSE and MAE values indicate high forecast accuracy, and minimal bias (Bias = 0.03%) indicates the absence of systematic error.

Table 3 - Indicators of model error analysis

Indicator	Meaning
R^2 (Cr)	0.94
RMSE (Cr), %	0.72
MAE (Cr), %	0.48
Bias, %	0.03

Table 4 presents the results of an assessment of the influence of key input parameters on the forecast of chromium content, the change in the proportion of ore (±0.9%) has the greatest impact- this corresponds to the physico-chemical nature of the process. At the same time, the ensemble model demonstrates stability with varying parameters.

Table 4 - Sensitivity analysis of the model

Parameter	Changing the Cr forecast, %
Ore content ±5%	±0.9
Furnace capacity ±3%	±0.6
Melting time ±5%	±0.4

It is shown that the average absolute error of the model (0.48%) and RMSE (0.72%) do not exceed the permissible technological deviation (±1.0%), which confirms the possibility of practical application of the model for preliminary quality control Table 5.

Table 5 - Comparison with regulatory tolerances

Indicator	Meaning, %
Parameter	±1.0
Parameter	0.48
Parameter	0.72

The use of an ensemble model can significantly reduce decision-making time and reduce the risk of production outside the standards.

The potential reduction of deviations provides a significant economic effect with large production volumes Table 6.

The histogram in Fig.5 shows the distribution of the remnants of the model. The errors are distributed symmetrically relative to zero, which confirms that there is no bias in the forecast. Most of the values are concentrated in a narrow range, indicating the stability of the algorithm.

Table 6 - Practical effect of implementation

Indicator	Evaluation
Average time of laboratory analysis	2–4 hr
Model forecast time	5 min
Reducing the risk of marriage	15–25%
Potential reduction of deviations	1–2% annual volume

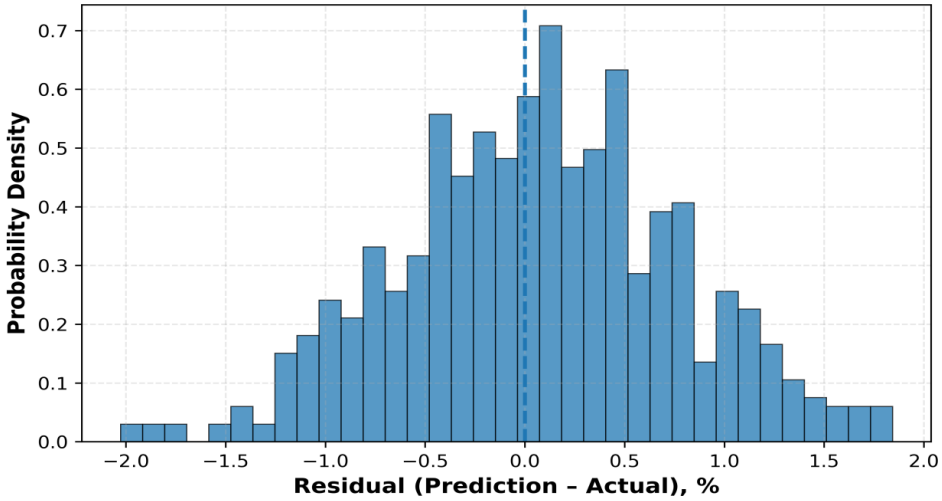


Fig.5 - Distribution of forecast errors

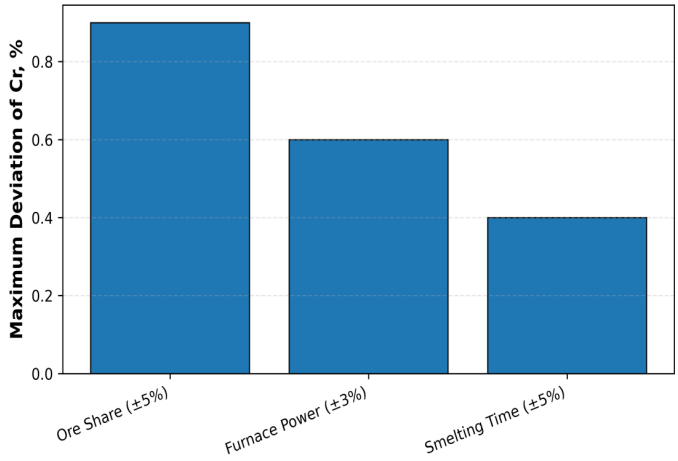


Fig.6 - Maximum deviation of the forecast of Cr content

The diagram in Fig.6 shows the maximum deviation of the forecast of the Cr content when changing key parameters. The greatest sensitivity is observed in the composition of the ore. At the same time, the model is stable without sudden jumps in the predicted values.

The architecture of the intelligent laboratory control system for the chemical composition of metallurgical products, fig.7, is designed to form a predictive system for laboratory control

of the chemical composition of metallurgical products based on the integration of laboratory, technological, and real data.

The system is designed for use in metallurgical production with the possibility of integration into the digital circuit of the enterprise.

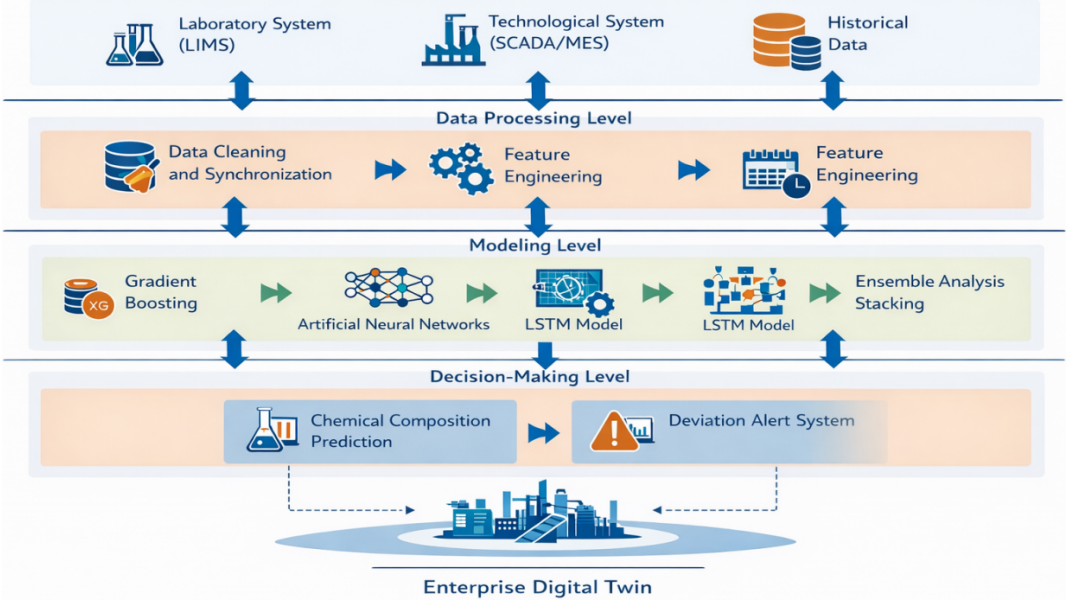


Fig.7 - Architecture of an intelligent laboratory control system for the chemical composition of metallurgical enterprise products

The architecture includes four levels:

- the level of data sources;
- processing level;
- model level;
- the level of decision-making.

At the source level, data is integrated from the laboratory system (LIMS), the technological system (SCADA/MES), and real data.

The processing level includes data cleaning, synchronization, normalization, feature engineering, and time series construction.

The model layer implements machine learning algorithms.

At the decision-making level, a forecast of the chemical composition is formed, and automatic notification of deviations is carried out.

The architecture uses machine learning (gradient boosting (XGBoost, CatBoost) to identify nonlinear dependencies between technological parameters and chemical composition. Artificial neural networks (ANN) for modeling complex multiparametric relationships. LSTM is a model for taking into account the temporal dynamics of the process and delays in laboratory measurements. A feature of the ensemble method (Stacking) is the combination of the prediction of individual models, with increased accuracy and stability of the forecast.

The scientific novelty of the development lies in the integration of laboratory, technological, and real data into a single predictive architecture, the use of an ensemble approach (Boosting + ANN + LSTM), the inclusion of a laboratory control model in the structure of an enterprise' s digital twin, and the modeling of laboratory measurements using time algorithms.

The implementation of the system ensures:

- reduction of technological decision-making time from several hours to minutes;
- reducing the likelihood of production outside of regulatory requirements;
- increasing the stability of the chemical

composition;

- Real-time operator support.

The architecture forms the basis for building a digital twin, further digitalizing the quality management of metallurgical production. The architecture of intelligent laboratory control can be used in metallurgical enterprises, processing plants, as part of digital production twins. It integrates with LIMS, SCADA/MES, and real data, forming a single contour of predictive quality management.

The advantage of the system is the ability to predict the chemical composition of products before obtaining laboratory results, identify possible deviations from standards, and automatically generate warnings.

The implementation of the solution in production helps to increase the stability of the chemical composition, reduce defects, and form a proactive quality management system as part of the digital transformation of the enterprise.

SCADA/MES transmits technological parameters (furnace power, temperature, charge composition, melting time, etc.) every minute.

The system performs cleaning, synchronization, generates signs, lag values, aggregates.

The ensemble model (Boosting + ANN + LSTM) predicts the chemical composition (for example, Cr%) after new parameters are received.

If the prognosis is outside the normal range, a warning is issued before receiving a laboratory analysis.

The model allows you to obtain a continuous forecast in real time, reflecting the dynamics of the process Fig.8.

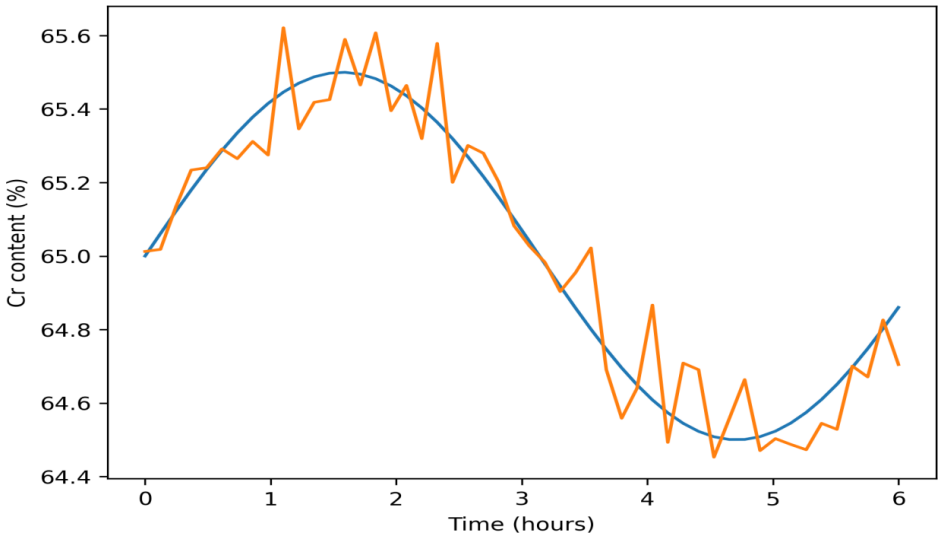


Fig.8 - Control data based on the ensemble model

The results of laboratory control at the enterprise, with a delay of 2 - 4 hours Fig.9.

Unlike traditional laboratory control, the developed system does not wait for laboratory confirmation, but allows you to adjust the technology in advance, reducing the risk of out-of-tolerance production in continuous operation.

The developed model has a number of limitations:

- the formation of the dataset is based on open sources, synthetic data generation, which does not fully reflect the real variability of the industrial process;

- there is no full access to the continuous time series of the enterprise, which limits the ability to account for various anomalies, non-standard melting modes;

- possible distortion of the model' s operation when changing the composition of raw materials, upgrading equipment, changing technological modes.

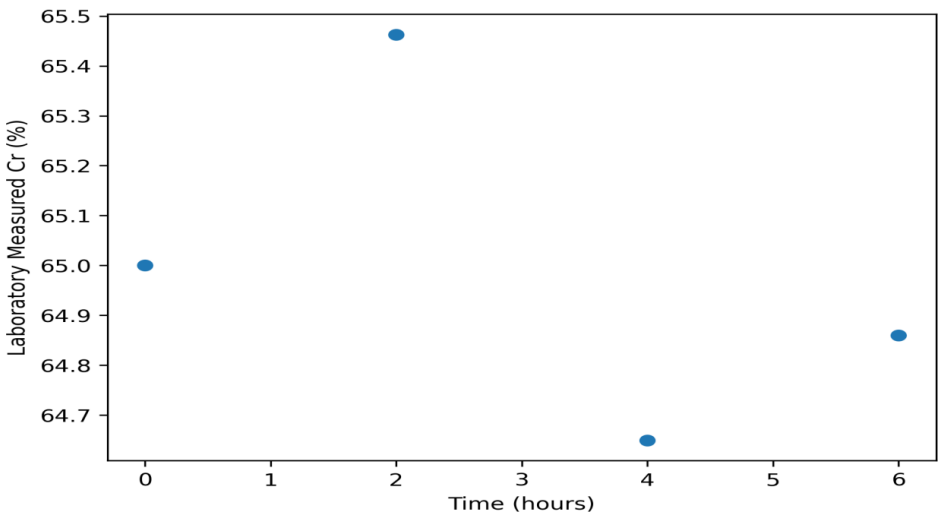


Fig.9 - Laboratory control results

In this regard, it is necessary to refine the algorithms. At the same time, the model does not take into account the physico-chemical equations of the process, relying on statistical dependencies.

One of the promising directions is the integration of the model into the digital twin of the enterprise with the possibility of scenario modeling (”what-if” analysis). It is possible to expand the architecture to other ferroalloys, and introduce probabilistic models with estimates of forecast confidence intervals.

Further research is aimed at combining statistical and physico-chemical models into a hybrid quality management system.

Conclusion. The architecture of an intelligent predictive laboratory control system for the chemical composition of metallurgical products of a mining and metallurgical enterprise has been developed and tested, based on the integration of laboratory (LIMS), technological (SCADA/MES), and real enterprise data.

An ensemble algorithmic approach has been implemented, combining methods of gradient boosting, artificial neural networks, and LSTM models, which made it possible to ensure high accuracy and stability of forecasting.

Quantitative indicators confirm the effectiveness of the developed solution: the coefficient of determination R2 for the Cr content reached 0.94, the RMS error (RMSE) was 0.72%, the average absolute error (MAE) was 0.48%, significantly lower than the regulatory tolerance (± 1.0%).

For elements with a low concentration (P, S), the absolute error does not exceed 0.01%. An analysis of the sensitivity of the model showed the stability of the model within ± 3-5% without significant degradation of accuracy.

Qualitative indicators indicate that there is no deviation in the forecast, the influence of factors, and the identified dependencies.

The developed system ensures the transition to proactive quality management, reducing decision-making time from 2-4 hours to several minutes, reducing the risk of production outside the standards by 15-25%.

The scientific novelty lies in the creation of a hybrid predictive laboratory control architecture, with the possibility of integration into the digital twin of the enterprise, providing continuous quality monitoring in real time.

The practical significance of the research lies in the possibility of scaling the solution to other types of metallurgical products and other enterprises.

The next stage of the research is the integration of the statistical model with the physico-chemical equations of chromium reduction, which will form a hybrid forecasting system.

Acknowledgement. *The research described in the article was conducted within the framework of*

the initiative project "Digital twin of the product and metallurgical production".
quality management system at all stages of mining

Referenses

1. Li X., Liu X., Li H., Liu R., Zhang Z., Li H., Lyu Q., Wen L. Research on sinter quality prediction system based on Granger causality analysis and stacking integration algorithm// Metals. -2023. -Vol. 13(2): 419. DOI 10.3390/met13020419.

2. Song Y., He S., Wu Q. DRBoost: A learning-based method for steel quality prediction// Symmetry. -2025. -Vol.17(10): 1644. DOI 10.3390/sym17101644.

3. Li Y., Zhang W., Chen X., et al. A model study on raw material chemical composition to predict sinter quality based on GA-RNN// Comput Intell Neurosci. -2022: 3343427. DOI 10.1155/2022/3343427.

4. Thampiriyanon J., Khumkoa S. Machine Learning–Based Prediction of Complex Combination Phases in High-Entropy Alloys//Metals. -2025. -Vol. 15(3): 227. DOI 10.3390/met15030227.

5. Sahoo S., et al. Prediction of Sinter Properties Using a Hyper-Parameter-Tuned Artificial Neural Network //ACS Omega. -2023. -Vol.8(12). -P.11782-11789. DOI 10.1021/acsomega.2c05980.

6. Ma L., Wang M., Peng K. A two-phase soft sensor modeling framework for quality prediction in industrial processes with missing data// Journal of Process Control. -2023. -Vol.129:103061. DOI 10.1016/j.jprocont.2023.103061

7. Zheng R., et al. Prediction of steelmaking process variables using K-medoids and a time-aware LSTM network// Metals. -2024. -Vol.10(12): e32901. DOI 10.3390/met14030456.

8. Yan F., et al. A Survey of Data-driven Soft Sensing in Ironmaking Systems: Research Status and Opportunities// ACS Omega. -2024. -Vol. 9(14). -P.25539-25554. DOI 10.1021/acsomega.4c01254.

9. Luo Y., Zhang X., et al. Data-driven soft sensors in blast furnace ironmaking: a survey// Frontiers of Information Technology & Electronic Engineering. -2023. -Vol.24. -P.327-354. DOI 10.1631/FITEE.2200366.

10. Molly R.R., Barton I.F. The process mineralogy of leaching sandstone-hosted uranium-vanadium ores// Minerals Engineering. -2022. -Vol.187:107811. DOI 10.1016/j.mineng.2022.107811.

11. Fudyma K., Nowak A., Kowalski P. Development of Digital Twins for Continuous Processes: Concept Description of Virtual Mass Balance Based on the Tennessee Eastman Process// Processes. -2025. -Vol.13(2):337. DOI 10.3390/pr13020337.

12. Zhang R., et al. State of the art in applications of machine learning in steelmaking process modeling// International Journal of Minerals, Metallurgy and Materials. -2023. -Vol.30. -P.2055-2075. DOI 10.1007/s12613-023-2646-1.

13. Eurasian Resources Group (ERG). Sustainable Development report. -2023. URL: https://www.eurasianresources.lu/uploads/1/files/ERG%20SDR%202023%20(ENG)%20(1).pdf. - Date of access: 17.01.2026.

14. U.S. Geological Survey. Mineral Commodity Summaries 2024: Chromium. -2024. URL: https://pubs.usgs.gov/periodicals/mcs2024/mcs2024-chromium.pdf. - Date of access: 17.01.2026.

Information about the authors

Akisheva L. - Nazarbayev Intellectual School, Astana, Kazakhstan, e-mail: lenaraakisheva@gmail.com;

Utegenov A. - student, K. Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan, e-mail: ali.utegenov@icloud.com;

Kyzken A. - student, K. Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan, e-mail: raptorblob76@gmail.com.

Информация об авторах

Акишева Л. - Назарбаев интеллектуальная школа, Астана, Казахстан, e-mail: lenaraakisheva@gmail.com;

Утегенов А. - студент, Казахский университет технологии и бизнеса им. К.Кулажанова, Астана, Казахстан, e-mail: ali.utegenov@icloud.com;

Кызкен А. - студент, Казахский университет технологии и бизнеса им. К.Кулажанова, Астана, Казахстан, e-mail: raptorblob76@gmail.com.