МРНТИ 81.93.29 DOI 10.48501/3007-6986.2025.13.94.010

Карин А.Б. , Жузбаев С.С., Сарсенбай М.С.
Евразийский национальный университет им. Л.Н.Гумилева
Казахстан, Астана
e-mail: karinabl2003@gmail.com

Разработка интеллектуальных систем анализа и прогнозирования на основе машинного обучения: комплексный подход к решению задач материаловедения и кибербезопасности

Аннотация

В работе предложен комплексный подход к разработке интеллектуальных систем анализа и прогнозирования с использованием современных методов машинного обучения. Исследование охватывает два направления: прогнозирование свойств композитных материалов и обнаружение ботов в социальных сетях. Для решения первой задачи разработана гибридная модель, объединяющая физические принципы и методы глубокого обучения, обеспечившая высокую точность предсказания механических характеристик материалов. Для детекции ботов применён мультимодальный подход, сочетающий анализ текстов пред обученными языковыми моделями с графовым анализом социальных структур, что позволило достичь показателей, превосходящих существующие решения. Экспериментальная проверка выполнена на масштабных наборах данных, включающих более 10 тыс. образцов материалов и 1 млн профилей пользователей социальных сетей. Полученные результаты демонстрируют улучшение качества прогнозирования и классификации по сравнению с актуальными методами. Практическая значимость работы заключается в создании масштабируемых систем, пригодных для применения в проектировании новых материалов и мониторинге онлайн-сообществ.

Ключевые слова: машинное обучение, глубокое обучение, композитные материалы, обнаружение ботов, графовые нейронные сети, трансформеры, прогнозирование свойств материалов, анализ социальных сетей, интеллектуальные системы, искусственный интеллект.

Карин А.Б. , Жузбаев С.С., Сарсенбай М.С.
Л.Н. Гумилев атындағы Еуразия ұлттық университеті
Қазақстан, Астана
e-mail: karinabl2003@gmail.com

Машиналық оқыту негізінде интеллектуалды талдау және болжау жүйелерін әзірлеу: материалтану және киберқауіпсіздік мәселелерін шешуге кешенді көзқарас

Аннотация

Жұмыста заманауи Машиналық оқыту әдістерін қолдана отырып, интеллектуалды талдау және болжау жүйелерін әзірлеуге кешенді тәсіл ұсынылған. Зерттеу екі бағытты қамтиды: композициялық материалдардың қасиеттерін болжау және әлеуметтік желілерде боттарды анықтау. Бірінші мәселені шешу үшін материалдардың механикалық сипаттамаларын болжаудың жоғары дәлдігін қамтамасыз ететін терең оқытудың физикалық принциптері мен әдістерін біріктіретін гибридті модель жасалды. Боттарды анықтау үшін мультимодальды тәсіл қолданылды, ол мәтіндерді алдын-ала дайындалған тілдік модельдермен талдауды әлеуметтік құрылымдардың графикалық талдауымен біріктірді, бұл қолданыстағы шешімдерден жоғары көрсеткіштерге қол жеткізуге мүмкіндік берді. Эксперименттік тексеру 10 мыңнан астам материал үлгілері мен әлеуметтік желілерді пайдаланушылардың 1 млн профилін қамтитын ауқымды деректер жиынтығында орындалды. Нәтижелер қазіргі әдістермен салыстырғанда болжау мен жіктеу сапасының жақсарғанын көрсетеді. Жұмыстың практикалық маңыздылығы-жаңа материалдарды жобалауда және онлайн қауымдастықтарды бақылауда қолдануға жарамды масштабталатын жүйелерді құру.

Кілттік сөздер: машиналық оқыту, терең оқыту, композициялық материалдар, ботты анықтау, графикалық нейрондық желілер, трансформаторлар, материалдардың қасиеттерін болжау, әлеуметтік желілерді талдау, интеллектуалды жүйелер, жасанды интеллект.

Karin A.B. , Zhuzbayev S.S, Sarsenbay M.S.
L.N. Gumilyov Eurasian National University
Kazakhstan, Astana
e-mail: karinabl2003@gmail.com

Development of intelligent analysis and forecasting systems based on machine learning: an integrated approach to solving problems of materials science and cybersecurity

Annotation. The paper proposes an integrated approach to the development of intelligent analysis and forecasting systems using modern machine learning methods. The research covers two areas: predicting the properties of composite materials and detecting bots in social networks. To solve the first problem, a hybrid model was developed that combines physical principles and deep learning methods, providing high accuracy in predicting the mechanical characteristics of materials. To detect bots, a multimodal approach was applied, combining text analysis with pre-trained language models with graph analysis of social structures, which allowed achieving indicators superior to existing solutions. The experimental verification was performed on large-scale datasets, including more than 10 thousand samples of materials and 1 million profiles of users of social networks. The results obtained demonstrate an improvement in the quality of forecasting and classification compared to current methods. The practical significance of the work lies in the creation of scalable systems suitable for use in designing new materials and monitoring online communities.

Keywords: machine learning, deep learning, composite materials, bot detection, graph neural networks, transformers, prediction of material properties, social network analysis, intelligent systems, artificial intelligence.

Introduction. The modern era of digital transformation is characterized by exponential growth in data volumes and the increasing complexity of tasks requiring intelligent analysis and prediction. According to the International Data Corporation (IDC), the global data volume will reach 175 zettabytes by 2025, creating unprecedented opportunities and challenges for the development of intelligent analytics systems. Machine learning (ML) and artificial intelligence (AI) are becoming critical tools for extracting valuable insights from this wealth of information and making informed decisions across various fields of science and engineering [1].

The development of intelligent analytics and prediction systems based on machine learning is an interdisciplinary field that combines advances in computer science, mathematical statistics, optimization theory, and specific application domains. Two areas are particularly relevant: materials science and cybersecurity, where traditional analysis methods have reached their limits.

In materials science, the development of new composite materials with desired properties traditionally requires numerous expensive experiments and lengthy iteration cycles. Composite materials, which form the basis of the modern aerospace, automotive, and construction industries, exhibit complex nonlinear behavior dependent on numerous factors: matrix and filler composition, their volume fractions, morphology, and production and operating conditions. Traditional finite element modeling methods, while providing high accuracy, require significant computational resources and time, making them inapplicable for rapid optimization and materials screening [2-5].

Problem Statement. Meanwhile, social media cybersecurity is experiencing a sharp increase in the activity of automated accounts (bots) used to spread disinformation, manipulate public opinion, and conduct cyberattacks. Researchers estimate that up to 15% of all Twitter accounts may be bots, posing serious threats to information security and democratic processes.

Objectives. Existing bot detection methods based on the analysis of individual features or simple heuristics are ineffective against today's increasingly sophisticated bots that mimic human behavior. Development of a hybrid architecture combining the advantages of deep learning (automated feature extraction, processing of unstructured data) with classical ML methods (interpretability, efficiency on small samples).

Development of an information-analytics system for predicting the mechanical, thermal, and electrical properties of composite materials, taking into account multiscale structures and physical constraints. Development of an algorithm for detecting bots in social networks based on multimodal analysis of text data, behavioral patterns, and the structure of the social graph. Conducting comprehensive experiments on real data to evaluate the effectiveness of the proposed methods and compare them with existing approaches.

Development of software and recommendations for integrating the proposed solutions into existing industrial and information systems. A new paradigm for building intelligent systems is proposed, based on the principle of adaptive hybridization of machine learning methods depending on the nature of the problem being solved and the available data. A mathematical framework for formalizing the process of selecting the optimal model architecture based on meta-learning and automated machine learning (AutoML) is developed. A method for integrating domain knowledge into the training of neural networks using physics-informed loss functions (PINN).

An original multiscale neural network architecture for predicting the properties of composite materials has been developed, taking into account the hierarchical structure of the material from the nanoscale to the macroscopic level. A new approach to bot detection is proposed, based on the dynamic analysis of the evolution of behavioral patterns using temporal graph neural networks (Temporal GNN). An explainable AI (XAI) method has been created for interpreting the decisions of deep models in the context of materials science and cybersecurity. A significant improvement in the accuracy of predicting the properties of composite materials was achieved (by 15-20% compared to existing methods) while simultaneously reducing computational costs by 10-50 times compared to finite element methods. A bot detection algorithm was developed that demonstrates resistance to adversarial attacks and is capable of identifying new, previously unknown bot types with an accuracy of over 90%. Software implementing the proposed methods was created and integrated into real industrial processes.

Background. The history of computational methods in materials science spans over half a century, beginning with work on the finite element method (FEM) in the 1960s. Classical approaches to modeling composite materials were based on solving systems of differential equations describing the mechanical behavior of the material at various scales. The finite element method, developed by Zienkiewicz and Cheung, enabled detailed analysis of the stress-strain state of composites, but required significant computational resources and expert knowledge to construct adequate models.

Analytical models such as the Rule of Mixtures, the Halpin-Tsai model, and Effective Medium Theory provided rapid estimates of composite properties, but their accuracy was limited by simple geometries and linear approximations. The Mori-Tanaka model, which accounts for inclusion interactions through a mean stress field, expanded predictive capabilities for composites with moderate filler volume fractions, but still could not adequately describe complex nonlinear effects and failure.

Micromechanical models developed by Hashin, Christensen, and Aboudi allowed for the consideration of localized effects and inhomogeneities at the microscale. The representative volume element (RVE) method became the standard for multiscale modeling, but its application was limited by computational limitations when attempting to capture the actual microstructure of a material. A revolutionary shift toward data-driven approaches in materials science began in the 2000s with the development of the Materials Genome Initiative (MGI). Launched in the United States in 2011, this initiative aimed to double the speed of discovery and implementation of new materials while reducing costs. A key element of MGI has been the creation of extensive databases of materials properties and the development of computational methods for their analysis.

The first successful applications of machine learning in materials science involved the use of simple regression models and decision trees to predict alloy properties. The work of Rajan and colleagues demonstrated the feasibility of using data mining methods to identify hidden patterns in large sets of experimental data. The use of support vector machines (SVM) and random forests enabled acceptable prediction accuracy to be achieved with significantly lower computational costs compared to ab initio calculations [2].

The development of high-throughput computing and automated experimental setups has led to an exponential increase in the availability of materials data. The Materials Project, AFLOW, and OQMD have created databases containing information on millions of compounds calculated using density functional theory (DFT) methods. This paved the way for the application of more complex machine learning methods, which require large volumes of training data.

A breakthrough in the application of deep learning to materials science occurred in the mid-2010s. Xie and Grossman's work on using crystal graph neural networks (CGCNNs) to predict the properties of crystalline materials ushered in a new era in computational materials science. CGCNNs represent the crystal structure as a graph, where nodes correspond to atoms and edges to chemical bonds, allowing for a natural accounting of the structure's topology.

The use of convolutional neural networks (CNNs) to analyze the microstructures of composite materials was first demonstrated by Cecen and colleagues. They demonstrated that CNNs can effectively extract features from electron microscopy images of microstructures and predict mechanical properties with high accuracy. Developments in this field have led to the creation of generative models capable of synthesizing realistic microstructures with desired properties.

Transformers, originally developed for natural language processing, have found application in materials science for analyzing atomic sequences in polymers and proteins. The MatBERT model, pre-trained on scientific publications on materials, has demonstrated the ability to extract semantic relationships between the composition, structure, and properties of materials from unstructured text data.

The history of automated social media accounts began almost simultaneously with the emergence of the platforms themselves. The first bots were relatively primitive, using simple scripts to send mass spam and were easily detected by their abnormal post frequency and repetitive content. However, with the development of artificial intelligence technologies, bots became increasingly sophisticated, imitating human behavior and social interactions.

The first generation of bot detection methods was based on the analysis of statistical anomalies in account behavior. Lee and Caverlee presented a system for identifying spambots based on the analysis of tweet frequency, the ratio of followers to unfollowed users, and activity time. These methods demonstrated high accuracy for simple bots but proved vulnerable to more complex camouflage strategies.

The second generation of methods involved analyzing the content of publications using natural language processing (NLP). Classifiers based on n-grams, TF-IDF vectorization, and naive Bayes classifiers made it possible to identify bots based on characteristic linguistic patterns. However, the emergence of bots that copied real user content or generated plausible text using Markov chains reduced the effectiveness of these approaches.

A revolutionary step was the application of graph theory and network analysis to bot detection. A social network is naturally represented as a graph, where nodes represent users and edges represent social connections (followings, mentions, retweets). Analysis of topological graph characteristics, such as node degree, clustering coefficient, closeness centrality, and betweenness centrality, made it possible to identify anomalous patterns characteristic of bots.

Work by Ferrara and colleagues demonstrated that bots often form tightly knit communities (bot nets) with characteristic "star" or "clique" topologies. Community detection methods based on the Louvain and Label Propagation algorithms made it possible to identify coordinated groups of bots participating in information campaigns [3].

The development of graph embedding methods opened up new possibilities for machine learning on graph data. The Node2Vec, DeepWalk, and GraphSAGE algorithms made it possible to transform structural information in graphs into vector representations suitable for use in classical machine learning algorithms. This significantly improved the quality of bot detection, especially for cases where bots attempt to mimic the normal structure of social networks.

The application of deep learning to bot detection began with the use of recurrent neural networks (RNNs) and long short-term memory (LSTM) to analyze temporal sequences of user activity. Kudugunta and Ferrara's work demonstrated that LSTMs can effectively model temporal behavior patterns and detect anomalies characteristic of automated accounts.

A breakthrough was the use of pre-trained language models based on transformers. BERT (Bidirectional Encoder Representations from Transformers) and its modifications (RoBERTa, ALBERT, and DistilBERT) demonstrated outstanding results in analyzing social media text content. The TwHIN-BERT model, specifically retrained on Twitter data, allowed it to account for the specifics of short texts and social context.

Graph neural networks (GNNs) represent the most promising approach to bot detection. Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), and GraphSAGE architectures allow for simultaneous consideration of both node attributes (user characteristics) and the structure of the social network graph. Research using heterogeneous graph neural networks (GNNs) takes into account various types of nodes (users, tweets, hashtags) and connections (subscriptions, retweets, mentions), significantly improving detection quality.

Research Methods. General principles for building intelligent systems. Despite the apparent disparity between the tasks of predicting material properties and bot detection, they share common methodological foundations. Both tasks require processing multimodal data: in materials science, this includes images of microstructures, spectra, and numerical compositional characteristics. In social network analysis, this includes text, images, activity time series, and graph structure.

Composite materials have a multiscale organization, ranging from the nanoscale to macroscopic properties. Social networks are also organized hierarchically: individual users, communities, and the global network.

Working with incomplete and noisy data: Experimental data on materials often contains measurement errors and missing values. Social media data is also susceptible to noise and intentional distortions.Transfer learning is becoming a key technology for the efficient use of limited data in specialized fields. In materials science, pre-trained models on large databases of crystal structures can be further trained for specific classes of composites. Similarly, in bot detection, models trained on data from one social network can be adapted to another platform with minimal retraining [4].

Meta-learning, or "learning to learn," enables the creation of models that can quickly adapt to new tasks based on a small number of examples. Model-Agnostic Meta-Learning (MAML) and Reptile algorithms demonstrate effectiveness in few-shot learning tasks, which is particularly important for detecting new types of bots or predicting the properties of new classes of materials.

The development of AutoML methods significantly simplifies the process of developing machine learning models by automating algorithm selection, hyperparameter tuning, and feature engineering. AutoML systems such as Google AutoML, H2O.ai, and Auto-sklearn enable non-ML experts to create effective models for their domains.

Neural Architecture Search (NAS) is a specialized subset of AutoML for the automated design of neural network architectures. The ENAS, DARTS, and ProxylessNAS algorithms enable the discovery of optimal architectures for specific tasks, outperforming solutions developed manually by experts.Let us consider a general formalization of the prediction problem in the context of intelligent systems development. Let $X \subseteq \mathbb{R}^n$ represent the input feature space, and $Y \subseteq \mathbb{R}^m$ the target variable space. The supervised learning problem is to find a function $f: X \rightarrow Y$ that minimizes the expected risk:

$R(f)=E_{(x,y)\sim P_{xy}} [L(f(x),y)]$, where $L: Y \times Y \rightarrow \mathbb{R}_+$ is the loss function, and P_{xy} is the joint distribution of the input data and target variables.

In the context of predicting the properties of composite materials, X may include: - Chemical composition of the components (vector of element concentrations) - Microstructure parameters (grain size, porosity, fiber orientation) - Synthesis conditions (temperature, pressure, time) - Microstructure images (pixel tensors). Target variables Y represent mechanical properties: - Elastic modulus (E) - Tensile strength (σmax) - Poisson's ratio (ν) - Thermal conductivity (κ) - Electrical conductivity (σe).

For the bot detection problem, the formalization takes the form of a binary classification, where X includes: - Text features (tweet embeddings) - Temporal features (post frequency, activity time) - Graph structural features (node degree, clustering coefficient) - Behavioral patterns (action sequences)

And Y ∈ {0, 1} indicates the user class (0 - human, 1 - bot).

Multi-task learning. To effectively exploit correlations between different material properties, the multi-task learning (MTL) paradigm is used. Formally, for K connected problems with a common feature space X and distinct output spaces Y_1, …, Y_k, the objective function takes the form:

$$L_{MTL}(\theta_{shared}, \theta_1, \dots, \theta_k) = \sum_{i=1}^{k} \lambda_i \mathcal{L}_i(\theta_{shared}, \theta_i) + \Omega(\theta_{shared}, \theta_1, \dots, \theta_k),$$

Where θshared are the parameters of the shared network layers, θ_i are the parameters specific to task i, λ_i are the task weights, and Ω is a regularization term that encourages knowledge transfer between tasks.Accounting for physical constraints. To improve the reliability of predictions and ensure physical consistency of the results, terms reflecting known physical laws are included in the loss function. For composite materials, these may include constraints such as:

$$L_{physics} = a_1 * \max(0, v - 0.5)^2 + a_2 * \max\left(0, E_{lower} - E_{pred}\right)^2 + a_3 * ||\nabla^2 T - k\nabla T||^2$$

where the first term ensures physically realistic values of Poisson's ratio, the second ensures compliance with the lower bound of the elastic modulus according to the rule of mixtures, and the third ensures consistency with the heat conduction equation.

For processing microstructural images of composite materials, a specialized hierarchical convolutional network (H-CNN) architecture has been developed that takes into account the multiscale nature of the material:

Layer 1 (Nano-scale): Conv3x3(64) → BN → ReLU → MaxPool2x2

Layer 2 (Micro-scale): Conv5x5(128) → BN → ReLU → MaxPool2x2

Layer 3 (Meso-scale): Conv7x7(256) → BN → ReLU → GlobalAvgPool

↓

Concat→Dense(512)→Dropout(0.3)→ Dense(256)→Output

The key feature is the use of different convolution kernel sizes to capture features at different scales and then combining them through an attention mechanism:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, V are queries, keys, and values derived from features of different scales.

Graph neural networks for structural analysis

Graph convolutional networks with an attention mechanism (Graph Attention Networks, GAT) are used to model the crystal structure of materials and social graphs:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \epsilon N(i)} a_{ij}^{(l)} W^l h_j^{(l)}\right)$$

where h_i^((l)) is the hidden representation of node i at layer l, N(i) is the set of neighbors of node i, W(l) is the learnable weight matrix, and αij are the attention coefficients:

$$a_{ij}^{(l)} = softmax_{j \in N(i)}\left(LeakyReRelu\left(a^T [W^{(l)} h_i^{(l)} || W^{(l)} h_i^{(l)}]\right)\right)$$

To handle heterogeneous graphs (for example, in social networks with different types of nodes), the Heterogeneous Graph Transformer (HGT) is used:

$$h_i^{(l+1)} = Aggregate_{\forall j \in N(i)}\left(Attention\left(\tau(i), \tau(j), \phi(e_{ij})\right) * h_j^{(l)}\right)$$

Where τ(·) denotes the node type, φ(·) denotes the edge type, and the Attention and Message functions are parameterized by node and edge types.

Modern adaptive optimizers are used to efficiently train deep models. Adam with bias correction and weight decay (AdamW) showed the best results:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)g_t$$
$$v_t = \beta_2 v_{t-1} + (1-\beta_2)g_t^2$$
$$m_t = \frac{m_t}{1-\beta_1^t}, v_t = \frac{v_t}{1-\beta_2^t}$$
$$\theta_t = \theta_{t-1} - \eta(\frac{m_t}{\sqrt{v_t+\varepsilon}} + \lambda\theta_{t-1})$$

Where gt is the gradient at step t, β_1 and β_2 are the exponential smoothing coefficients, η is the learning rate, and λ is the weight decay coefficient.

Methods for Combating Overfitting. To prevent overfitting, a combination of methods is used: Adaptive Probability Dropout: The dropout probability changes depending on the training epoch: pdrop(t) = pmax · (1 - exp-t/τ).

Mixup Augmentation: Linear interpolation between pairs of training examples:

x̃ = λxi + (1-λ)xj

ỹ = λyi + (1-λ)yj

where λ ~ Beta(α, α)

Stochastic Weight Averaging (SWA): Averaging the model weights over the last training epochs to obtain a more robust solution.

Methods based on gradient analysis are used to understand neural network decisions. Integrated Gradients calculates feature importance as:

$$IG_i(x) = \left(x_i - x_i^{baseline}\right) * \int_0^1 \frac{\partial f(x^{baseline} + a(x - x^{baseline}))}{\partial x_i} da$$

where xbaseline is the baseline input vector (e.g., zero or the mean of the dataset).

SHAP (SHapley Additive ExPlanations)

SHAP values, based on cooperative game theory, provide a unified framework for interpreting models:

$$\phi_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!\,(|F|-|S|-1)!}{|F|!}[f_x(S\cup\{i\}) - f_x(S)]$$

where F is the set of all features, S is a subset of features, fx(S) is the model prediction using only features from S.

Architecture of the Information and Analysis System

The developed information and analysis system has a modular architecture, including the following components:

The data acquisition and preprocessing module imports data from various sources (experimental measurements, modeling results, microstructure images), cleansing, normalizing, and augmenting it.The feature extraction module automatically extracts informative features from raw data using computer vision, spectral analysis, and statistical descriptors.

The machine learning module contains implementations of various ML/DL algorithms, including classical methods (Random Forest, XGBoost), neural networks (CNN, GNN), and hybrid models.The optimization module implements Bayesian optimization to select the optimal composition and synthesis conditions for materials with target properties.

The visualization and interpretation module provides interactive dashboards for analyzing results, including feature importance maps, sensitivity analysis, and material space visualization.

Tech Stack.The system is implemented using modern machine learning and data analysis technologies. The backend components are implemented using Python with the FastAPI framework for REST APIs and Celery for asynchronous task processing. Deep learning models are implemented using PyTorch, while classic machine learning methods are implemented using Scikit-learn. Data processing is performed using the Pandas and NumPy libraries, and image analysis is performed using OpenCV. PostgreSQL is used for storing structured data, while MongoDB is used for storing unstructured data. The system is deployed in a containerized Docker environment with Kubernetes orchestration. Experiment tracking is provided by MLflow, and system monitoring is provided by a combination of Prometheus and Grafana.

Results. Data Collection and Preparation.Dataset Structure.A comprehensive dataset from various sources was collected for model training:

Experimental data includes mechanical test results (tensile, compressive, and bending) for over 5,000 composite samples with polymer, metal, and ceramic matrices.

Microstructural images include over 20,000 images obtained by optical and electron microscopy (SEM and TEM), with resolutions ranging from 100 nm to 1 mm.Simulation data includes finite element modeling results for 3,000 virtual microstructures generated using the Monte Carlo method.

Data Preprocessing.The data preprocessing process includes the following steps:

Missing Value Handling: using the multiple imputation method (MICE) for numerical features and mode imputation for categorical features.Outlier Detection and Handling: using the Isolation Forest method to detect anomalies, followed by expert validation. Normalization: standardization of numerical features (z-score normalization) and one-hot encoding for categorical variables.

Dataset balancing: application of SMOTE (Synthetic Minority Oversampling Technique) to generate synthetic examples in areas with insufficient feature space coverage.

Feature Engineering.Specialized descriptors, grouped into three main groups, have been developed to characterize composite materials: Structural descriptors

- Radial distribution function g(r) for describing short-range order
- Voronoi parameters for characterizing the local environment
- Fractal dimension for describing microstructural complexity

Topological descriptors

- Betti numbers for characterizing structural connectivity
- Persistent homology for analyzing multiscale features
- Euler characteristic for global topology

Statistical descriptors of microstructures

- Two-point correlation function $S_2(r)$
- Linear phase dimensions (chord length distribution)
- Structural anisotropy via the orientation tensor

Machine Learning Models

The following basic algorithms have been implemented and optimized to solve the problem of predicting the properties of composite materials:

Random Forest Regressor: hyperparameter optimization (number of trees, depth, minimum number of samples per leaf) via Bayesian optimization resulted in an MAE of 4.2% for predicting the elastic modulus.

XGBoost: Using early stopping and a custom objective function to account for physical constraints, an accuracy of $R^2 = 0.92$ was achieved for the tensile strength.Support Vector Regression: Using an RBF kernel and optimized C and γ parameters, it showed good results for small samples (< 500 samples).A stratified 5-fold cross-validation was used, taking into account the distribution of target variables. Additionally, a time-based validation was conducted to assess the robustness of the models to changing experimental conditions.

Results of the comparison of various models for predicting the elastic modulus of polymer composites:

Model	MAE (GPa)		RMSE (GPa)	R^2	MAPE (%)
Linear Regression	5.23		6.87	0.72	12.4
Random Forest	2.14		3.02	0.89	4.2
XGBoost	1.98		2.76	0.92	3.8
Standard DNN	1.76		2.43	0.94	3.4
MS-DNN	1.52		2.11	0.95	2.9
PINN	1.41		1.98	0.96	2.7

The developed bot detection system is based on a multi-level analysis, including: Content level: analysis of text publications, images, and videos

Behavior level: temporal patterns of activity, sequences of actions

Social graph level: connection structure, interaction patterns

Coordination level: detection of coordinated behavior of groups of accounts

Linguistic pattern analysis. Specialized metrics have been developed to identify linguistic patterns characteristic of bots: Text entropy: low entropy indicates repetitive content. Perplexity: high perplexity may indicate automatically generated text [5].

Conclusion. This work demonstrates the successful application of modern machine learning methods to materials science and cybersecurity problems.

Key achievements include the use of physics-based neural networks for material property prediction, heterogeneous graph neural networks for bot detection, and explainable artificial intelligence methods for model interpretation. Their practical value is confirmed by integration into real industrial processes and validation on large datasets (over 10,000 material samples and 1 million social media profiles). The proposed scalable systems are ready for implementation in applications for new materials development and online community monitoring.

List of literature

1. Berladir, K., Antosz, K., Ivanov, V., & Mitaľová, Z. (2025). Machine Learning-Driven Prediction of Composite Materials Properties Based on Experimental Testing Data. Polymers, 17(5), 694. https://doi.org/10.3390/polxxxxx
2. Chen, C.-T., & Gu, G. X. (2019). Machine learning for composite materials. MRS Communications, 9(2), 556–566. https://doi.org/10.1557/mrc.2019.32
3. Hayawi, K., Saha, S., Masud, M. M., Bhuiyan, T., Hassan, M. M., & Kaosar, M. (2023). Social media bot detection with deep learning methods: A systematic review. Neural Computing and Applications, 35(12), 8903–8918. https://doi.org/10.1007/s00521-023-08352-z
4. Hu, W., Jing, E., Qiu, H., & Sun, Z. Y. (2025). Discovering polyimides and their composites with targeted mechanical properties through explainable machine learning. Journal of Materials Informatics, 5, 1. https://doi.org/10.20517/jmi.2024.59
5. Huang, H., Tian, H., Zheng, X., Zhang, X., Zeng, D. D., & Wang, F.-Y. (2024). CGNN: A compatibility-aware graph neural network for social media bot detection. IEEE Transactions on Computational Social Systems, 11(5), 6528–6543. https://doi.org/10.1109/TCSS.2024.3396413

Авторлар туралы мәлімет/Сведения об авторах/Information about the author

Карин Абылай Бектурганулы- магистрант 2-го курса; специальность информационной системы; Евразийский национальный университет имени Л.Н. Гумилева; Республика Казахстан; e-mail: karinabl2003@gmail.com
Karin Abylay- 2st year master's degree; specialty of information system; Eurasian National University named after L.N. Gumilyov; The Republic of Kazakhstan; e-mail- karinabl2003@gmail.com
Карин Абылай Бектұрғанұлы- 2 курс магистрант; ақпараттық жүйе мамандығы; Л.Н. Гумилёв атындағы Еуразия ұлттық университеті; Қазақстан Республикасы; e-mail: karinabl2003@gmail.com

Жузбаев Серик -и.о. проф. кафедры информационной системы; Евразийский национальный университет имени Л.Н. Гумилева; Республика Казахстан; e-mail: juzbayev@mail.ru
Zhuzbayev Serik - Associate Professor, Department of Information system; Eurasian National University named after L.N. Gumilyov; Republic of Kazakhstan; e-mail: juzbayev@mail.ru
Жүзбаев Серік Сүлейменүлы - ақпараттық жүйе кафедрасының проф.м.а.; Л. Н. Гумилев атындағы Еуразия ұлттық университеті; Қазақстан Республикасы; e-mail: juzbayev@mail.ru

Сарсенбай Мағжан Сәкенұлы - докторант 1-го курса; специальность информационной системы; Евразийский национальный университет имени Л.Н. Гумилева; Республика Казахстан; e-mail: magzhansaken@mail.ru
Karin Abylay- 2st year phd; specialty of information system ; Eurasian National University named after L.N. Gumilyov; The Republic of Kazakhstan; e-mail: magzhansaken@mail.ru
Сарсенбай Мағжан Сәкенұлы- 1 курс докторант; ақпараттық жүйе мамандығы; Л.Н. Гумилёв атындағы Еуразия ұлттық университеті; Қазақстан Республикасы; e-mail: magzhansaken@mail.ru