Rockburst Forecasting using Composite Modelling for Seismic Sensors Data

Ilia Revin[0000−0002−4459−8724], Vadim A. Potemkin[0000−0002−8019−7282], and Nikolay O. Nikitin

ITMO University, Saint Petersburg, Russia
ierevin@itmo.ru
https://en.itmo.ru/

Abstract. Seismic monitoring is used to ensure the safety of workers in the rock massif. The main security threat is a rockburst, which can be predicted based on the sequence of seismic events. An important task is to develop a mining forecasting model that can take into account the structural heterogeneity of the mountain range and select the necessary forecast horizon depending on monitoring data. In the paper, we propose a flexible approach that combines multiple machine learning models designed to solve various tasks (clustering, time series forecasting) as parts of one composite model. This approach allows for adjustment of the forecast horizon of the model, which enables it to flexibly adapt to rock massifs with different geological structures and seismic monitoring stations. Also, the use of clustering models allows us to take into account the physical and mechanical features of the rockburst formation process. According to experimental results, the resulting composite model showed more accurate results for specific forecast horizons, compared with classical "hierarchical" models and machine learning models. At the same time, the obtained model allows us to interpret the results from the rock mechanics point of view.

Keywords: Rockburst forecasting · Geomonitoring · Clustering · Time series forecasting · Machine learning · Data-driven modeling.

1 Introduction

Rockbursts are a highly complex dynamic phenomenon. The formation of a rockburst is influenced by many factors, such as the physical and mechanical properties of the rock mass, stress state, geological structure, and engineering position [1]. The classical method of rockburst forecasting is the use of various statistical criteria. However, those criteria ignore rock massif's physical and mechanical parameters, and it leads to poor quality of forecast [2].

Many machine learning methods are used in rockburst forecasting [3] and it is an actively developing field of signal processing and machine learning methods for solving rock mechanic problems. However, these models can restore only extremely simple and trivial relationships between seismic events and the probability of a rockburst [4]. Another problem is "class imbalance", which is expressed

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_29

in the fact that rockburst is an extremely rare event, and their number is significantly less than the number of seismic events. Correcting class imbalance leads to a lack of data in the training dataset and the inability to use deep learning models [5]. The principal scheme of the seismic monitoring station and an example of data visualization is shown in Fig. 1. An example of sensor data from a seismic monitoring system is shown in Tab. 1, where the input data consists of timestamps and coordinates of seismic events.

Table 1. Example of input seismic sensor data

Timestamp	X	Y	Z	Energy
00:00:00	0.120	0.200	0.032	1200
00:01:00	0.1284	0.201	0.038	2000
00:02:00	0.1089	0.513	0.011	11000

The sequence of seismic events appears during mining processes in the massif and can be represented as a time series that consists of discrete events. Each event is characterized by a time coordinate and a characteristic describing the "degree of destruction" (e.g., the magnitude of the signal's energy). Clustering methods are used to localize the spatial zone of a potential rockburst, which corresponds to each event with its spatial cluster (Fig. 1).

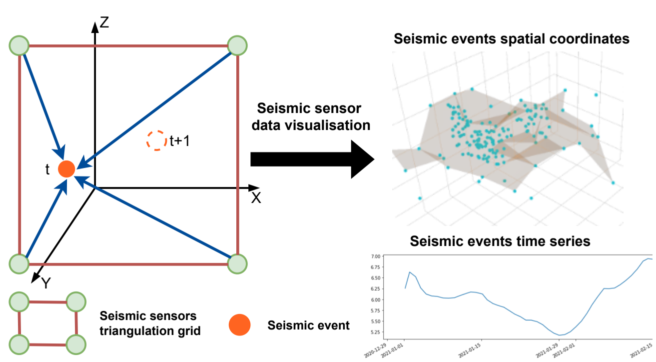

Fig. 1. Principal scheme of the seismic monitoring station, based on the triangulation of seismic events in the rock massif

However, the geological heterogeneity of the rock massif leads to an uneven distribution of rockburst. It leads to the appearance of clusters with an increased or decreased probability of a rockburst.

In this paper, we propose a data-driven automated hybrid modelling approach, which intends to solve the problems described above. The main idea is to combine temporal and spatial sensor data in one feature space and apply

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_29

an automated forecasting model design using a graph-based pipeline representation and evolutionary optimization. We use spatial clustering methods based on machine learning models for spatial and temporal data seismic monitoring combinations. Then, we present the data within each cluster as a discrete time series. The proposed approach is fully automated and has been tested on experimental data, where it has shown its effectiveness compared to similar approaches.

2 Related Works

Classical methods of rockburst forecasting are based mainly on a deterministic empirical approach [6]. This approach cannot be adapted to the uncertain conditions of a complex dynamic system, which is the studied rock massif. In the rock mechanic "hierarchical model" theory, the criterion for the rockburst formation is a violation of the stationarity/quasi-stationarity conditions of the modeled process (for example, the Poisson process). Hierarchical models can usually predict only the total number and time of rockbursts at fixed spatial coordinates.

Short-term prediction methods based on monitoring seismic data can simultaneously predict both the time of the occurrence and the expected location of the rockburst [4]. These methods are primarily focusing on signal processing and filtering. Nevertheless, this approach has proven to be more flexible and scalable than the empirical, numerical, and physical models [7].

Deep learning models have long established themselves as an effective method for modeling various processes using both temporal and spatial data types. In several works [7], the use of such models for the problem of rockburst prediction has shown high efficiency. The reason is the ability to model complex non-linear relationships between factors affecting the probability of rockburst occurrences. However, the disadvantages of this approach include the low interpretability of the model results and the tendency to overfit due to the relatively small size of the datasets.

The hybrid approach has also found its application in rockburst forecasting. This approach combines the capabilities of existing rock mechanic models and machine learning models. Classical models provide a representation of geological heterogeneities in the rock massif. ML models reproduce complex nonlinear connections between inhomogeneities, external factors, and the seismic activity of the rock massif. Accordingly, the combination of classical models that take into account the physical laws and machine learning models capable of modeling complex nonlinear dependencies is the most promising trend of development in this area [8]. Another hybrid approach introduced a hybrid model combining KMeansSMOTE oversampling with Random Forest optimization using Optuna (KMSORF). This model demonstrated high accuracy in predicting rockburst levels in real-world mining projects while addressing challenges like imbalanced datasets. Another work integrated Particle Swarm Optimization (PSO) with neural networks (e.g., BPNN) and ensemble methods like XGBoost, achieving prediction accuracies exceeding 0.9 in practical applications [12]. These develop-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_29

ments highlight how combining classical geomechanical principles with machine learning techniques can enhance prediction precision and robustness under complex geological conditions.

3 Proposed approach

Currently, the identification of data-driven models with a complex heterogeneous structure remains an unsolved problem [9]. The desired mathematical model can be developed using a single machine learning model and a hybrid (composite) approach [10]. The set of clustering models includes HDBSCAN, KMeans and Spectral models. For time series forecasting, such models as Singular Spectrum Analysis (SSA), Random Forest, and XGBoost are used.

The aim of the proposed approach is to predict rockbursts in a technologically disturbed rock massif. We can consider our massif as a discrete dynamical system $X_{next} = F(X_{cur})$, where X_{cur} is the current state of the massif. The discrete-time propagator F is given by the flow map:

$$F(X_{cur}) = X_{cur} + \int_{k\triangle t}^{(k+1)\triangle t} f(x(r))dr \tag{1}$$

Since one of the stages of the model is spatial clustering of seismic events, X_{cur} can be represented as a matrix $X \in \mathbb{R}^{N\times M}$, where N is the number of clusters, and M is the length of the time series of seismic events.

$$X = \begin{bmatrix} x_1(t_1) \dots x_1(t_m) \\ x_2(t_1) \dots x_2(t_m) \\ x_n(t_1) \dots x_n(t_m) \end{bmatrix} \tag{2}$$

The future state of the system X_{next} can be expressed in a similar way:

$$X_{next} = \begin{bmatrix} x_1(t_{m+1}) \dots x_1(t_{m+k}) \\ x_2(t_{m+1}) \dots x_2(t_{m+k}) \\ x_n(t_{m+1}) \dots x_n(t_{m+k}) \end{bmatrix} \tag{3}$$

Here the hyper-parameters of the proposed model are N and K — the length of the forecast.

The task of finding function F is simultaneously a data-driven modeling task and a multi-criteria optimization task [11], shown in Eq. 4. Since both spatial and temporal coordinates describe each seismic event, the proposed model consists of various machine learning models most suitable for a particular data type in a single composite model.

$$\hat{U} = \bigcap_{i=1}^{m} \arg\min_{u\in W} f_i(u) \tag{4}$$

where:

- f_i — objective criteria that characterizes the modelling quality;

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_29

– W — a set of possible solutions (search space);
– $\hat{U}$ — a vector of composite model hyperparameters;
– $\bigcap_{i=1}^{m}$ — an intersection of the set of solutions for each of the criteria, and m is the number of criteria used during optimization.

To solve the problem of multi-criteria optimization, we proposed the following criteria:

– **Silhouette criterion**. The silhouette shows how the average distance to the objects in its cluster differs from the average distance to the objects of other clusters. This value is in the range $[-1, 1]$. Values close to -1 correspond to poor (scattered) clusterization, and values close to zero indicate that clusters intersect and overlap.
– **FAR - The proportion of false alarms, and MAR - the proportion of missed alarms.** A sliding window is used to calculate this criterion. Both metrics have identical lower and upper bounds - [0, 1]. The selection of such metrics is based on its applicability for expert use and in order to take into account class imbalance.

In this paper, three values of the detection window width were taken. The short-, medium-, and long-term forecast horizons correspond to the values of six hours, two days, and seven days before and after the appearance of the rockburst.

4 Experiments

In order to evaluate the effectiveness of the proposed approach, experimental comparisons of the composite model with existing approaches were carried out. Obtained composite model consists of three machine learning models. The HDB-SCAN model is used for spatial clusterization of seismic events, KNN model is used to fill in gaps in time series, and SSA model is used to time series forecasting. As a dataset, we used a synthetic dataset that was developed considering key characteristics of real seismic phenomena — seven rockbursts were distributed unevenly in space and time, reproducing the class imbalance problem typical for real monitoring data. The results show that this approach to synthetic data generation allows for effective testing and comparison of various model architectures (hierarchical, LSTM, and the proposed composite model), revealing their strengths and weaknesses across different forecasting horizons. In order to correctly evaluate the work of similar time series prediction models, we carried out preliminary clustering of seismic events using the proposed composite model. Selected cluster is "stable" in time, i.e. they include seismic events during the entire monitoring period.

Fig. 3 shows the solutions considered during optimization. FAR/MAR metric value shows the normalized ratio of the sum of false and missed predicted rockbursts to true rockbursts. The closer the value is to 1, the greater the number of false and missed rockbursts.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_29

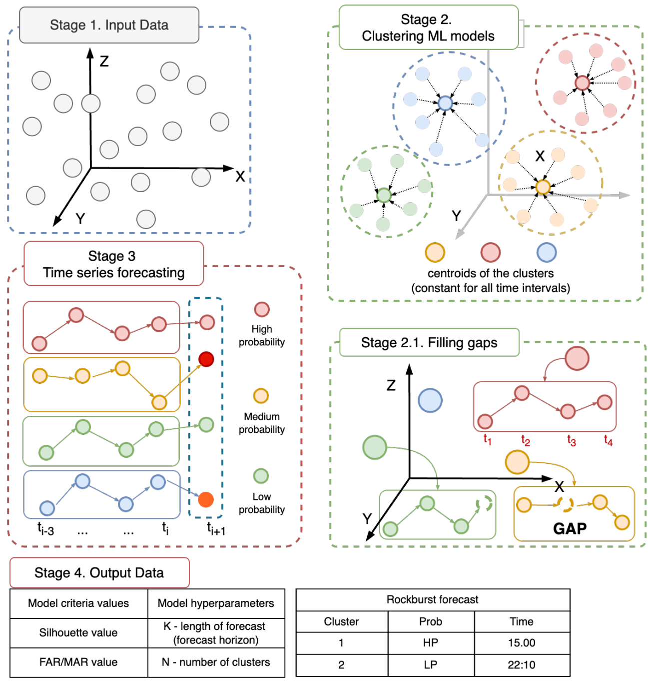

Fig. 2. Proposed approach for rockburst forecasting

For model comparison, we used the "hierarchical model" as the baseline, the LSTM as the DL approach and our composite model. The results of the experiment are shown in Table 2 (*Italics* indicate the best result for each of the forecast horizons, **bold** indicates the best result among all horizons). Composite model showed superior result among all models with long-term forecasts and the best result among all models when using a medium-term forecast. Such results are related to the fact that the use of spatial data in the composite model allows localizing time-stable clusters of seismic events. The probability of rockburst in such clusters increases over time, which explains the effectiveness of the composite model in the medium and long-term forecast horizon. This also explains the effectiveness of the LSTM model in short-term forecasts, because in the absence of formed clusters, such a model allows better modeling of complex nonlinear dependencies that lead to rockbursts.

Table 2. FAR-MAR criteria comparison

Model	Short-term (6 hours)	Mid-term (2 days)	Long-term (7 days)
Hierarchical model	0.993	0.974	0.758
LSTM	*0.417*	0.714	0.688
Composite model	0.533	**0.365**	*0.562*

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_29

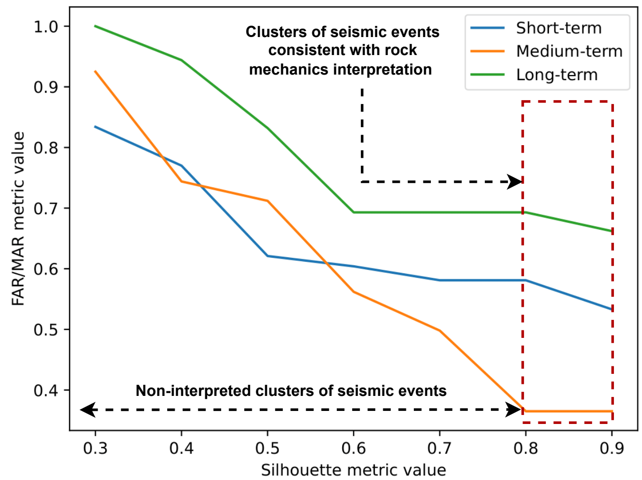

Fig. 3. Comparison of FAR/MAR metrics for three proposed forecast horizons

5 Conclusion

This paper proposes an approach to the problem of rockburst forecasting, based on seismic monitoring data. The idea of the approach is to automatically combine machine learning models based on temporal and spatial data into a single composite model.

The multi-criteria optimization of the proposed composite model has shown its effectiveness. Appropriate values of the forecast horizon and the seismic event silhouette criterion make it possible to obtain results that are superior to hierarchical and DL models.

The proposed model implements a data-driven approach. On the one hand, it simultaneously uses spatial and temporal coordinates of seismic events, using the entire amount of information obtained during seismic monitoring. On the other hand, it reproduces rock mechanics phenomena, such as zones of the stress-strain state of the console, making it more interpretable than other models. The use of a synthetic seismic monitoring dataset in our research serves as an experimental demonstration of the importance of creating directionally generated synthetic multidimensional time series for improving the robustness of machine learning models.

We can conclude composite model is a more effective means of predicting rockbursts than classical hierarchical models and is slightly inferior to DL models when using short-term forecasts. The proposed approach based on machine learning and signal processing methods, is an effective forecasting algorithm. However, it can be improved by including more complex forecasting models and new criteria.

The implemented algorithms and examples of their application are available in https://github.com/ITMO-NSS-team/RockBurst.AI repository.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_29

Acknowledgments. This research is financially supported by The Russian Scientific Foundation, Agreement # 24-71-10093, https://rscf.ru/en/project/24-71-10093/

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. F. Gao, Hongpu Kang and Jianzhong Li: Numerical simulation of fault-slip rockbursts using the distinct element method. Tunnelling and Underground Space Technology, **110** (2021) https://doi.org/0.1016/J.TUST.2020.103805
2. Zhang Chao, Gaohan Jin, Liu Chao, Li Shugang, Junhua Xue, Cheng Renhui, Wang Xinglong and Xiangzhen Zeng. Prediction of rockbursts in a typical island working face of a coal mine through microseismic monitoring technology. Tunnelling and Underground Space Technology (2021). https://doi.org/10.1016/J.TUST.2021.103972
3. H. Tian and Ya-dong Xue. Prediction of rockburst classification using Support Vector Machine. IOP Conference Series: Earth and Environmental Science/ **861** (2021).https://doi.org/10.1088/1755- 1315/861/6/062088
4. Ma, T.H., Tang, C.A., Tang, L.X., Zhang, W.D., Wang, L.: Rockburst characteristics and microseismic monitoring of deep-buried tunnels for Jinping II Hydropower Station. Tunnelling and Underground Space Technology **49**, 345–368 (2015). Elsevier
5. Pu, Y., Apel, D., Liu, W., Mitri, H.: Machine learning methods for rockburst prediction-state-of-the-art review. Int. J. Mining Sci. Technol. **29** (2019). https://doi.org/10.1016/j.ijmst.2019.06.009
6. Ochmann, J., Jurczyk, M., Rusin, K., Rulik, S., Bartela, Ł.,Uchman, W., Solution for Post-Mining Sites: Thermo-Economic Analysis of a Large-Scale Integrated Energy Storage System. Energies. (2024) https://doi.org/10.3390/en17081970
7. Zhou, J., Li, X., Mitri, H.S.: Classification of rockburst in underground projects: comparison of ten supervised learning methods. J. Comput. Civ. Eng. **30**(5), 04016003 (2016). ASCE
8. Sarafanov, M., Borisova, Y., Maslyaev, M., Revin, I., Maximov, G., Nikitin, N.: Short-Term River Flood Forecasting Using Composite Models and Automated Machine Learning: The Case Study of Lena River. Water **13**, 3482 (2021). https://doi.org/10.3390/w13243482
9. Nikitin, N., Revin, I., Hvatov, A., Vychuzhanin, P., Kalyuzhnaya, A.: Hybrid and automated machine learning approaches for oil fields development: The case study of Volve field, North Sea. Computers & Geosciences **161**, 105061 (2022). https://doi.org/10.1016/j.cageo.2022.105061
10. Sarafanov, M., Pokrovskii, V., Nikitin, N.: Evolutionary Automated Machine Learning for Multi-Scale Decomposition and Forecasting of Sensor Time Series. In: Proceedings of IEEE CEC (2022). https://doi.org/10.1109/CEC55065.2022.9870347
11. Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Ye, A., Miao, C., Dai, Y.: Multi-Objective Adaptive Surrogate Modeling-Based Optimization for Parameter Estimation of Large, Complex Geophysical Models. Water Resour. Res. **52** (2015). https://doi.org/10.1002/2015WR018230
12. Li, S.; Lu, P.; Liang, W.; Chen, Y.; Da, Q.: Performance Evaluation of Hybrid PSO-BPNN-AdaBoost and PSO-BPNN-XGBoost Models for Rockburst Prediction with Imbalanced Datasets, Appl. Sci.**14** (2024) https://doi.org/10.3390/app142411792

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_29