

Article

Application of Machine Learning Methods to Assess Filtration Properties of Host Rocks of Uranium Deposits in Kazakhstan

Yan Kuchin ^{1,2,*} , Ravil Mukhamediev ^{1,2,*} , Nadiya Yunicheva ^{1,3,*}, Adilkhan Symagulov ^{1,2}, Kirill Abramov ¹, Elena Mukhamedieva ¹, Elena Zaitseva ⁴  and Vitaly Levashenko ⁴

¹ Institute of Information and Computational Technologies MSHE RK, Pushkin Str., 125, Almaty 050010, Kazakhstan; asmogulove00@gmail.com (A.S.); abramc@mail.ru (K.A.); muhamedieva@gmail.com (E.M.)

² Institute of Automation and Information Technologies, Satbayev University (KazNRTU), Satpayev Str., 22A, Almaty 050013, Kazakhstan

³ Institute of Automation and Information Technologies, Almaty University of Energy and Communications, Baitursynov Str., 126/1, Almaty 050013, Kazakhstan

⁴ Faculty of Management Science and Informatics, University of Žilina, Univerzitná 8215/1, Žilina 01026, Slovakia; elena.zaitseva@fri.uniza.sk (E.Z.); vitaly.levashenko@fri.uniza.sk (V.L.)

* Correspondence: ykuchin@mail.ru (Y.K.); ravil.muhamedyev@gmail.com (R.M.); naduni@mail.ru (N.Y.)

Abstract: The uranium required for power plants is mainly extracted by two methods in roughly equal amounts: quarries (underground and open pit) and in situ leaching (ISL). Uranium mining by in situ leaching is extremely attractive because it is economical and has a minimal impact on the region's ecology. The effective use of ISL requires, among other things, the accurate assessment of the host rocks' filtration characteristics. An accurate assessment of the filtration properties of the host rocks allows optimizing the mining process and improving the quality of the ore reserve prediction. At the same time, in Kazakhstan, this calculation is still based on methods that were developed more than 50 years ago and, in some cases, produce inaccurate results. According to our estimates, this method provides a prediction of filtration properties with a determination coefficient $R^2 = 0.32$. This paper describes a method of calculating the filtration coefficient of ore-bearing rocks using machine learning methods. The proposed approach was based on nonlinear regression models providing a 20–75% increase in the accuracy of the filtration coefficient assessment compared with the current methodology. The work used different types of machine learning algorithms based on the gradient boosting technique, bagging technique, feed-forward neural networks, support vector machines, etc. The results of logging, core sampling, and hydrogeological studies obtained during the exploration stage of the Inkai deposit were used as the initial data. All used machine learning models demonstrated significantly better results than the old method. This resulted in improved results compared with previous studies. The LightGBM regressor demonstrated the best result ($R^2 = 0.710$).

Keywords: uranium mining; machine learning; regression model; filtration characteristics; boosting; bagging



Citation: Kuchin, Y.; Mukhamediev, R.; Yunicheva, N.; Symagulov, A.; Abramov, K.; Mukhamedieva, E.; Zaitseva, E.; Levashenko, V. Application of Machine Learning Methods to Assess Filtration Properties of Host Rocks of Uranium Deposits in Kazakhstan. *Appl. Sci.* **2023**, *13*, 10958. <https://doi.org/10.3390/app131910958>

Academic Editor: Deqiang Cheng

Received: 15 September 2023

Revised: 29 September 2023

Accepted: 30 September 2023

Published: 4 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Intensive development of nuclear energy can be considered one of the means to combat global warming. For example, nuclear power plants in Europe annually avoid the emission of 700 million tons of CO₂. Uranium mining is carried out in more than 20 countries, but the main uranium reserves are concentrated in Australia, Kazakhstan, and Canada [1]. According to the World Nuclear Association, in 2018, the largest uranium mining companies produced 86% of the world's total uranium production [2], of which NAC Kazatomprom JSC accounted for 21%. There are two main extraction methods used: open pit (underground and open-pit) and in situ leaching (ISL). Open pit (accounting for 45.9% of production) is used with a sufficiently high uranium content, ILS (accounting for

48.3% of production) is used with a fairly low uranium content, and the reserves must be located in highly permeable rocks (most often sands). Approximately 5.8% of uranium is mined as a by-product, such as in gold mining [3] (Figure 1).

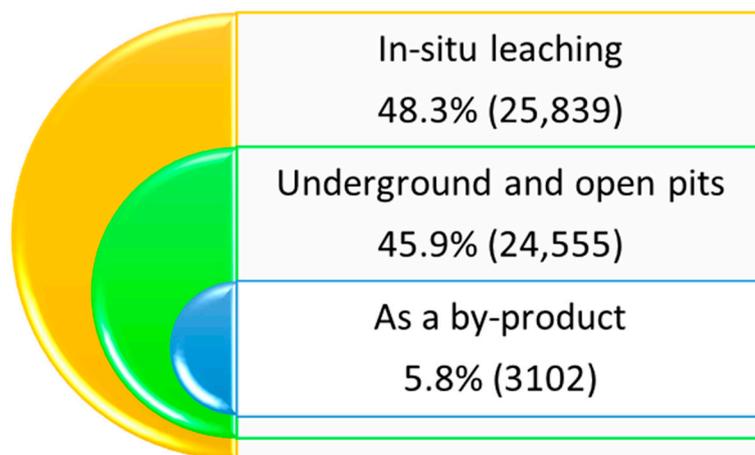


Figure 1. World uranium mining volumes by mining methods.

ISL is a method of developing of sandstone-type uranium deposits without raising ore to the surface by selectively transferring uranium ions into a productive solution directly in the subsurface using a network of injection and extraction wells. Uranium-bearing ore remains underground, unlike traditional mining methods (mines and quarries). ISL is the most economical and environmentally friendly mining method, widely used in Kazakhstan, Uzbekistan, Canada, and Australia. However, to successfully apply the ISL method, it is necessary to solve the following problems:

- Accurate determination of the lithological composition of the host rocks and the depth of permeable and impermeable strata using geophysical methods;
- Assessment of filtration properties of host rocks, for the correct assessment of recoverable reserves and production planning.

Inaccuracies in solving these problems not only prevent the organization of an optimal production process but often lead to serious financial losses. For example, economic losses from incorrect lithological classification in the deposits of Kazakhstan can be estimated to be approximately USD 1 to 4 million per year [4].

Inaccuracies in assessing the filtration properties of host rocks are caused both by inaccuracies in electrical logging and, to a large extent, by the methodology for determining the lithological composition and filtration properties of rocks. When determining the filtration properties of host rocks in a field, the key point is to determine the relationship between the filtration coefficient K_f , determined as a result of hydrogeological studies at the exploratory drilling stage, and electrical logging data, which are subsequently used to calculate the filtration properties of technological wells. However, the accepted methodology, based on analytical methods, has not changed since the end of the last century [5]. At the same time, the correct determination of K_f is necessary for calculating recoverable reserves, predicting production dynamics, and calculating the optimal number of wells and the distance between them (the diameter of a hexagonal cell or the distance between rows of wells).

One of the promising ways to improve the quality of the filtration properties assessment is the use of artificial intelligence [6], more specifically, machine learning methods [7]. Machine learning is used in problems of stratigraphy [8], geological mapping [9], assessment of the prospects of tungsten deposits [10], composition of iron ore deposits [11], and lithology [12–15]. The application of ML for lithological classification is considered in [16].

The purpose of the study was to evaluate the possibilities of using machine learning models for K_f calculation based on well log data in sandstone-type uranium deposits.

This study considered the application of machine learning methods to estimate the filtration characteristics of ore-bearing rocks. The method was based on the use of nonlinear regression models and has shown results 20–75% better than calculations using the existing methodology used in Kazakhstan. The proposed method concerns approximately half of the mined uranium in the world.

The work consists of the following sections. The first section briefly provides general information about uranium and its mining methods and the application of machine learning methods in geophysical research.

In the second section, we provide an overview of the current state of research in the field of determining the permeability of geological formations using machine learning.

In the third section, we describe the research method.

In the fourth, we describe the initial data and the results obtained.

In conclusion, the limitations of the method and directions for further research are discussed.

2. Related Works

The permeability of rocks is an important factor influencing the percentage of hydrocarbon recovery, reservoir management, and carbon dioxide sequestration during oil production. During in situ leaching of uranium, rock permeability is the most important factor in deciding whether to install downhole filters and predicting ore recovery.

To solve the problems of permeability predicting, porosity, and other petrophysical properties of rocks in mining, regression models are often used [11], which allow developing more accurate and robust models than traditional empirical, statistical models [17]. Such models have been studied since the 1990s. For example, in Ref. [18], probably for the first time, a multilayer neural network was used to estimate the porosity of rocks. To assess the permeability of rocks, a hybrid algorithm using neural networks (ANNs) was proposed in [19]. An ANN-based regression analysis was used to obtain a set of relationships between the permeability, porosity, and pore size in [20]. ANNs as a non-linear regression method is used to estimate the porosity and permeability of an oil reservoir based on log data [21,22]. A similar problem is considered in [23], where the authors compared ANN, SVM, and fuzzy neural network models. The authors concluded that ANNs can be used to estimate the permeability of a heterogamous carbonate reservoir based on three parameters: the bulk density, the neutron porosity, and a mobility index introduced by the authors with a mean square error of 0.28. In this paper, the authors also used ANNs to estimate the permeability of an oil-bearing reservoir in the Persian Gulf of Egypt and obtained very high values of the coefficient of determination $R^2 = 96.5\%$.

In Ref. [24], the porosity of oil reservoirs was studied using ANNs based on seismic sounding data. It was possible to study the petrophysical properties in the interwell space and identify zones of bypass sand channels and leaks that were not visible on the structural maps and attribute slices. The high result of the estimation of the porosity of the oil reservoir based on the SVM model was described in [25], where the correlations between the model estimates and real data exceeded 0.96. The authors of the paper stated that machine learning models worked more accurately than traditional estimation methods [26], reaching RMSE = 0.38 and $R^2 = 97\%$ for the SVM-based model when processing log data from carbonate reservoirs of southwestern Iran.

Well logs can be considered as a type of images that can be analyzed using deep learning models when there are enough data. For example, to estimate the permeability in the process of oil production in [27], convolutional neural networks (CNNs) were used, which showed an advantage over ANNs. The use of CNNs to predict the properties of subsurface rocks in the process of drilling wells in real time is discussed in [28]. The authors show that the model is able to distinguish between different rock types such as cemented sandstone, unconsolidated sands, and shales. A CNN-based model was proposed in [29] to accompany the technological process of well construction, namely, to assess the integrity of cement in cased wells.

The deep learning model was used to identify the similarity of geological interlayers, and thus, with higher accuracy, allows estimating cross-well correlation [30]. New Zealand and Norway open datasets were used to tune the model. The accuracy of the model was 0.926, which was significantly higher than the base models based on gradient boosting. A similar problem of estimating the interwell space of an oil and gas reservoir is considered in [31], where a three-dimensional CNN is used, which shows better results compared with ANNs.

As shown above, the use of machine learning methods is very popular in the assessment of permeability, porosity, and interwell space in oil fields, where such methods in many cases show good results. However, the situation is different for uranium deposits.

According to the authors, such studies of rock permeability assessment for sandstone-type uranium deposits have not been previously carried out, with the exception of [6]. Meanwhile, uranium mining at such deposits is carried out by the method of in situ leaching, in which the filtration properties of rocks are critically important. The use of regression models to assess the filtration coefficients is a way to improve the accuracy of the calculation of recoverable ore reserves and optimize the mining process.

The range of regression models is quite large. They, like other machine learning models, can be roughly divided into classical and modern ones [6].

Although deep learning provides excellent results in many cases [32], its application is possible, as a rule, in the presence of a large amount of data or in the presence of pre-trained models, using the transfer learning technique [7]. According to the authors, there are no such datasets in the public domain. In this regard, for a comparative assessment of the possibilities of solving the problem of predicting the filtration properties of rocks based on logging data by models of different types, it was decided to use several types of regression models based on support vector algorithms, boosting, bagging, and neural networks.

3. Proposed Method

The methodological scheme of the study included the following steps (Figure 2):

- Feature selection and dataset generation;
- Training and tuning machine learning models;
- Evaluation of results using standard quality metrics.

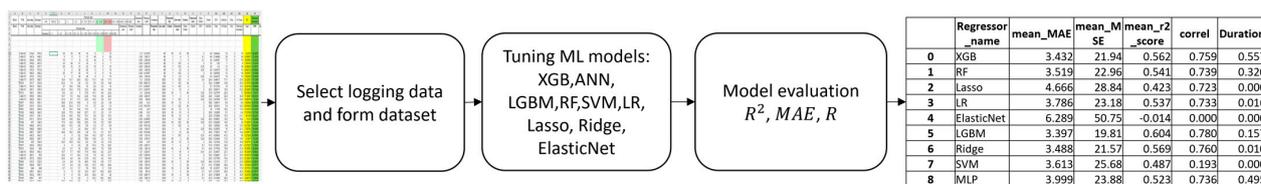


Figure 2. Methodological scheme of the study.

3.1. Data Preparation

The methodology currently used provides for the establishment of a relationship between the filtration properties of host rocks and the apparent resistivity measured at the exploration stage. To do this, resistivity logging data are compared with the results of hydrogeological studies of wells and results of granulometric studies [33] (Figure 3).

As a result, a correspondence table is compiled, intermediate values are obtained by interpolation. The disadvantages of the adopted methodology include the fact that it uses data from only one logging method and cannot be used if the recording quality is insufficient, as well as the fact that it uses the average resistivity value within the selected interval, which cannot be accurately determined for intervals with thickness less than 1.5–2 m, since the distance between the measuring electrodes during logging is 1 m. Given the shortcomings of the existing K_f estimation methodology, we proposed a machine learning model that received basic logging data as the input and generated filtration coefficients as the output. Such a model could be trained on data from exploratory wells

that have actual K_{fpo} (pumped out). The trained model could then be used to calculate K_f of the technological wells.

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
1	Block	Well	Start dept	End dept	>10	10-5.0	5-2	2-1	1-0.5	Fracture (m)						0.005-0.01	<0.002	Calcareous mass	Comment	Permeability	Lito code	Colour	Characteristic	Class Mode	D50	AlcClay	Clay	<0.1m	K _f	Electrical resistance	Co2	D50
2	Block	Well	Start dept	End dept	>10	10-5.0	5-2	2-1	1-0.5	Fracture (m)						0.005-0.01	<0.002	Calcareous mass	Comment	Permeability	Lito code	Colour	Characteristic	Class Mode	D50	AlcClay	Clay	<0.1m	K _f	Electrical resistance	Co2	D50
3	Block	Well	Start dept	End dept	>10	10-5.0	5-2	2-1	1-0.5	Fracture (m)						0.005-0.01	<0.002	Calcareous mass	Comment	Permeability	Lito code	Colour	Characteristic	Class Mode	D50	AlcClay	Clay	<0.1m	K _f	Electrical resistance	Co2	D50
5	9-304-23	674.8	676.9		20	16	18	14	5	1	14						0.1		90	18	12	58	2	20	0.944444	14	0	15	14.2779	20.4875		
6	3-308-11	647.4	648.3		0	0	13	49	5	29						1.06		95	1	4	2		0.1	49	0.14898	29	0	34	16.8956	13.35284		
7	1-280-12	654.8	655.5		10	7	9	23	34	4	11					0.09		98	1	3	4		0.1	34	0.25087	11	0	15	19.0951	13.47026		
8	1-280-12	678.9	681.5		33	20	17	14	5	2	7					0.09		96	1	123	4		2	33	1.25	7	0	9	7.52415	11.37028		
9	1-280-12	656.5	657.7		20	16	13	22	10	5	13					0.09		99	1	123	4		0.25	22	0.5	13	0	18	18.8821	13.38105		
10	1-280-12	672.3	673.8		0	2	38	40	6	3	10					0.09		99	1	43	4		0.25	40	0.44375	10	0	13	5.84212	9.72843		
11	1-280-12	668.8	668.2		25	17	18	15	6	5	13					0.09		96	1	123	2		2	25	0.89556	13	0	18	19.0507	13.55846		
12	1-280-12	670.5	672.2		0	2	38	40	6	3	10					0.09		99	1	43	4		0.25	40	0.44375	10	0	13	7.06325	10.56393		
13	1-280-7-1	657.9	658.7		15.4	13.1	20.4	16.2	11.1	6.5	17.1					0.1		99.8	1	123	4		0.5	20.4	0.48811	17.1	0	23.6	21.22033	17.12953		
14	*0010	631.7	632.9		52.4	7.4	5.9	9.3	5.5	5.6	14.8					0.09		99.8	1	122	44		2	52.4	2.14895	14.8	0	20.4	27.20779	25.5498		
15	1-280-7-1	661.6	662.5		27	14.5	15.8	14	9.2	5.3	14					0.11		99.8	1	123	2		2	27	0.73732	14	0	19.3	18.41635	12.80838		
16	1-280-13	655.1	656.6		22.3	15.4	17.8	12.2	8.8	6.4	16.9					0.11		99.8	1	12	4		2	22.3	0.690112	16.9	0	23.3	22.28545	18.10476		
17	*0113	638.7	640.3		8.6	2.4	5	37.8	15.1	8.6	22.4					0.1		99.9	18	3	58		0.25	37.8	0.75794	22.4	0	31	20.8891	15.81689		
18	1-280-7-1	659.5	660.4		7	7.6	14.2	24.8	11.8	8.6	22.2					0.09		99.9	1	3	4		0.25	24.8	0.324597	22.2	0	30.8	17.40872	15.23666		
19	1-280-7-1	662.8	664.1		21.4	9.1	11.1	23.2	14.2	5.8	15.1					0.09		99.9	1	123	3		0.25	23.2	0.41556	15.1	0	20.9	19.89783	13.84302		
20	*0037	653.5	655.1		38.6	21.6	16.9	9	4.2	2.6	6.8					0.09		99.9	18	12	8		2	38.6	1.48611	9.8	0	9.4	1.92984	8.27592		
21	*0037	631.5	633.3		45.8	21.4	11.7	6.1	3.7	2.8	7.4					0.18		99.9	1	122	44		2	45.8	1.85514	7.4	0	10.2	5.03026	9.91804		
22	*0061	639.3	640.9		35	18.1	10.9	11.5	7	4.8	12.6					0.19		99.9	18	12	8		2	35	1.176796	12.6	0	17.4	17.44756	12.57837		
23	*0126	640.8	642.3		32.8	17.5	14.8	20.7	5.7	2.3	5.1					0.09		99.9	18	12	8		2	32.8	1.02287	6.1	0	8.4	17.2768	13.75619		
24	*0126	634.7	636.1		23.8	18.2	15.2	10.1	5.9	0.9	15.3					0.09		99.9	18	12	8		2	23.8	0.74839	10.3	0	21.1	18.10501	12.62913		
25	*0127	661.1	661.9		18.9	14.8	20.2	20.6	6.6	5.2	13.6					0.09		99.9	1	123	4		0.25	20.6	0.59901	13.6	0	18.8	11.01003	12.58884		
26	*0128	637	638.5		17.6	14.2	19.8	22.5	5.8	5.5	14.5					0.09		99.9	18	123	58		0.25	22.5	0.542929	14.5	0	20	17.93914	12.34401		
27	*0132	658.8	661.5		49.3	15.1	8.8	8.9	4.3	3.7	9.8					0.09		99.9	1	122	4		2	49.3	1.890265	9.8	0	13.5	16.70612	15.26666		
28	*0133	648.8	649.6		0.6	1.7	2.4	59.4	13.7	6.1	16.4					0.09		99.9	1	3	44		0.25	59.4	0.39764	16	0	22.1	7.78782	10.49595		
29	*0134	628.7	629.7		44.9	18.2	10.9	5.8	5.3	4.1	10.7					0.09		99.9	1	12	44		2	44.9	1.75275	10.7	0	14.8	4.12971	11.25165		
30	*0135	642.5	643.3		0	0	0.2	20.4	14.3	23.7	41.3					0.09		99.9	1	370	2		0.01	41.3	0.088354	41.3	0	65	1.27283	6.78881		
31	*0165	626.8	626.7		1.6	2.4	6.8	59.9	14.6	4.3	11.3					0.09		99.9	18	3	58		0.25	59.9	0.334041	11.3	0	15.8	13.65895	11.8268		
32	*0165	635	636.6		36.6	16.9	13.2	8.6	5.8	5.2	13.6					0.09		99.9	18	12	8		2	36.6	1.213016	13.6	0	18.8	16.57507	12.36799		
33	*0167	634.6	635.9		0	0	0.4	20	51.4	7.8	20.3					0.09		99.9	1	4	2		0.1	51.4	0.163911	20.3	0	28.1	2.17027	7.12043		
34	*0210	624.8	626		0.6	0.1	1.8	26	44.5	7.6	19.8					0.09		99.9	18	4	58		0.1	44.5	0.17018	19.8	0	27.4	15.5666	10.8358		
35	1-280-13	659.5	660.3		15.7	7.7	15.5	17.9	10.8	9.2	23.1					1.37		99.9	1	19	4		0.01	23.1	0.346369	23.1	0	32.3	16.43865	10.88209		
36	1-280-13	662.1	663		36.2	17	14.4	11.2	7.8	8.5	21.1					0.97		99.9	1	12	4		2	36.2	0.48811	21.1	0	29.6	16.58756	11.02876		
37	1-280-13	651.6	652.8		26.6	4.2	8.4	21.9	16.2	9.2	16.4					0.17		99.9	1	12	4		2	26.6	0.377854	16.4	0	22.5	21.23912	14.1451		
38	*0024	631.6	632.5		2.4	2.8	7.8	48.6	13.8	6.8	17.7					0.09		99.9	18	3	58		0.25	48.6	0.310185	17.7	0	24.5	6.92626	12.89148		
39	*0061	658.9	659.7		0.1	0.4	0.4	18.4	41.3	10.9	28.4					0.09		99.9	1	4	44		0.1	41.3	0.138892	28.4	0	39.3	13.54828	9.78977		
40	*0061	648	649		0.7	3.1	3.7	56	18.5	12.9						0.09		99.9	18	3	58		0.25	56	0.307273	12.9	0	17.9	16.75396	11.83951		
41	*0061	659.7	662.3		0	0	0.6	16.3	43.7	10.9	28.4					0.09		99.9	1	4	44		0.1	43.7	0.136728	28.4	0	39.3	13.54828	9.78977		
42	*0125	654.2	655.1		0	0	0.6	28.4	42.4	8.5	22					0.09		99.9	18	4	58		0.1	42.4	0.168886	22	0	30.5	14.3748	9.87544		
43	*0125	659.7	661		0	0	0.2	3	52.3	19.2						0.09		99.9	1	4	4		0.1	52.3	0.116991	29.2	0	44.4	14.40372	9.52591		
44	*0126	624.2	625.6		38.7	12.5	5.5	14	6	1.1	16.1					0.09		99.9	18	12	8		2	38.7	1.104	16.1	0	22.2	22.42884	15.17409		
45	*0127	623.3	624.3		17	0.7	4.5	64.1	13.1	4.4	11.4					0.09		99.9	18	3	58		0.25	64.1	0.332293	11.4	0	15.8	14.40372			

Table 1. Machine learning models.

Regression Model	Abbreviation	Method	References
XGBoost	XGB	Ensemble learning method based on the gradient boosted trees algorithm.	[35]
LightGBM	LGBM	Ensemble learning method based on the gradient boosted trees algorithm.	[36–38]
Random forest	RF	Ensemble learning method based on bagging technique.	[39]
Support vector machines	SVM	Method is based on the kernel trick.	[40]
Artificial neural network	ANN	Feed forward neural network.	[41,42]
Linear regression	LR	Modeling impact of independent variables to target variable based on linear approach.	[43]
Lasso regression	Lasso	Based on the use of such a regularization mechanism that not only helps in reducing over-fitting but can help in feature selection.	[44]
Ridge regression	Ridge	A regularization mechanism is used to prevent over-fitting.	[45]
Elastic net	elasticnet	Hybrid of ridge regression and lasso regularization.	[46]

The variety of algorithm types provides a broad search for the appropriate type of models. Looking ahead, we note that the ensemble learning methods, and more precisely the gradient boosting models, generally showed the best result.

To assess the quality of regression models, the following accuracy indicators were used [47]:

Coefficient of determination (R^2);

Mean square error (mean squared error— MSE);

Mean absolute error (mean absolute error— MAE);

Correlation coefficient (R).

The quality measures used to evaluate the regression models are listed in Table 2.

To perform computational experiments, a software system was developed in Python using the numpy, sklearn, matplotlib, cv2, alive_progress, pickle, and tensorflow libraries, which enable the reading and preparing of the initial data, forming a data frame, and applying machine learning models. The computational experiments were performed on a Dell XPS 15 9500 computer equipped with 32 GB of RAM, an Intel(R) Core(TM) i7-10750H processor, and an Nvidia GeForce GTX 1650 Ti discrete graphics card. All equipment is manufactured in China.

Table 2. Evaluation metrics of regression models.

Evaluation Index	Equation	Explanation
Mean absolute error	$MAE = \frac{\sum_{i=1}^n (y^{(i)} - h^{(i)})}{n}$	where n is the sample size; $y^{(i)}$ is the actual value of the target variable for the i -th sample; $h^{(i)}$ is the estimated value for the i -th sample.
Mean squared error	$MSE = \frac{\sum_{i=1}^n (y^{(i)} - h^{(i)})^2}{n}$	

Table 2. Cont.

Evaluation Index	Equation	Explanation
Determination coefficient	$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$ $SS_{res} = \sum_{i=1}^n (y^{(i)} - h^{(i)})^2$ $SS_{tot} = \sum_{i=1}^n (y^{(i)} - \bar{y})^2, \bar{y} = \frac{1}{n} \sum_{i=1}^n y^{(i)}$	
Linear correlation coefficient (or Pearson correlation coefficient)	$R(y, h) = \frac{\sum_{i=1}^n (h_i - \bar{h})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (h_i - \bar{h})^2}},$	where $\bar{h} = \frac{1}{n} \sum_{i=1}^n h_i$.

4. Data and Results

In total, there were 3558 intervals in the collected dataset, with a thickness of 0.5 to 2.2 m. As input parameters, sections of the AR and SP curves with a thickness of 0.5 m, recorded with a step of 0.1 m, as well as the lithological code of the rock, determined by the experts, were used. During the experiments, nine regression models were tested: XGB, RF, Lasso, Linear regression, Elastic Net, LGBM, Ridge, SVM, and ANN (hidden_layer_size = 91) on all combinations of input parameters (AR, SP, AR + SP, AR + lithocode, SP + lithocode, and AR + SP + lithocode). The XGB, RF, and LGBM models did not require data normalization (more precisely, normalization of input parameters can degrade the performance of these models), while other models worked better with normalized data. The normalization was performed using the standard function MinMaxScaler() included in the sklearn package. The results of the experiments are shown in Table 3.

Table 3. Assessments of the performance of the models trained on the exploration wells.

Input Data	Regressor_Name	MAE	MSE	R ²	R	Duration
AR	XGB	3.432	21.94	0.562	0.759	0.557
	RF	3.519	22.96	0.541	0.739	0.320
	Lasso	4.666	28.84	0.423	0.723	0.000
	LR	3.786	23.18	0.537	0.733	0.016
	elasticnet	6.289	50.75	−0.014	0.000	0.000
	LGBM	3.397	19.81	0.604	0.780	0.157
	Ridge	3.488	21.57	0.569	0.760	0.010
	SVM	3.613	25.68	0.487	0.193	0.000
	MLP	3.999	23.88	0.523	0.736	0.495
SP	XGB	6.4580	63.01	0.259	0.172	0.731
	RF	6.5580	75.65	−0.312	0.036	0.357
	Lasso	6.2890	50.75	0.386	0.000	0.000
	LR	6.2570	49.30	0.415	0.277	0.000
	elasticnet	6.2890	50.75	0.386	0.000	0.000
	LGBM	6.2600	50.60	0.389	0.118	0.151
	Ridge	6.4590	55.56	0.289	0.104	0.010
	SVM	6.9040	81.00	−0.419	0.171	0.000
	MLP	7.5120	80.91	−0.217	−0.229	0.569

Table 3. Cont.

Input Data	Regressor_Name	MAE	MSE	R ²	R	Duration
AR + SP	XGB	3,780	25.86	0.509	0.715	0.406
	RF	4.449	33.01	0.373	0.724	0.490
	Lasso	5.305	38.06	0.277	0.757	0.000
	LR	4.627	33.40	0.365	0.770	0.000
	elasticnet	6.809	60.94	−0.158	0.000	0.000
	LGBM	4.420	30.97	0.412	0.738	0.160
	Ridge	5,770	80.78	−0.535	0.490	0.020
	SVM	6.554	58.63	−0.114	0.178	0.000
	MLP	4,870	34.74	0.340	0.708	0.547
AR + lito code	XGB	2.976	15.93	0.682	0.829	0.709
	RF	2.964	15.77	0.665	0.829	0.344
	Lasso	4.666	28.84	0.423	0.723	0.000
	LR	3.385	18.48	0.631	0.799	0.000
	elasticnet	6.289	50.75	−0.014	0.000	0.000
	LGBM	2.875	14.52	0.710	0.845	0.149
	Ridge	3.196	25.24	0.495	0.728	0.010
	SVM	3.593	23.48	0.531	0.254	0.000
	MLP	3,250	18.12	0.638	0.800	0.559
SP + lito code	XGB	4,801	36.85	0.300	0.551	0.425
	RF	5.013	41.81	0.206	0.535	0.320
	Lasso	6.918	63.00	−0.197	−0.033	0.000
	LR	5.208	39.53	0.249	0.587	0.000
	elasticnet	6.809	60.94	−0.158	0.000	0.000
	LGBM	4,880	36.26	0.311	0.601	0.175
	Ridge	5,883	119.05	−1.262	0.229	0.010
	SVM	6.140	60.65	−0.152	0.235	0.000
	MLP	6.380	62.02	−0.178	−0.271	0.573
AR+SP+ lito code	XGB	3.538	22.084	0.580	0.764	0.453
	RF	4.032	27.889	0.470	0.754	0.496
	Lasso	5.305	38.055	0.277	0.757	0.016
	LR	4.188	26.522	0.496	0.806	0.000
	elasticnet	6.809	60.940	−0.158	0.000	0.000
	LGBM	3.856	25.013	0.525	0.781	0.165
	Ridge	5.637	96.590	−0.835	0.439	0.020
	SVM	6.366	55.389	−0.052	0.154	0.000
	MLP	4.538	31.637	0.399	0.726	0.569
Current method		13.89	192.93	0.32	0.584	

In the last row, for comparison, the results of the calculations according to the currently accepted method are given. It can be seen that the best set of input data was the combination of AR + lithocode, while the use of SP significantly worsened the accuracy of the models. This was probably due to the low quality of the PS curve recording. Since the recorded

potential depended on the difference between the salinity of groundwater and the drilling fluid, for contrast and differentiable recording of SP, it was necessary to strictly adhere to the requirements for the preparation of the drilling fluid, which was often not observed in practice. Because the size of the dataset was small, the training time for all algorithms was less than 1 s.

The LGBM regressor showed the best results when using (AR, Litho code) as the input parameters ($R^2 = 0.710$). At the same time, when we could use only a part of the input parameters, other regression models could also be used. For example, a linear regression model performed well when we could only use SP, and we did not know the lithological code ($R^2 = 0.415$). In cases where we could only use the AR values, the best result again showed LGBM ($R^2 = 0.604$). In addition, linear regression (LR) showed a stable good result. The currently used methodology showed significantly worse results compared with machine learning models ($R^2 = 0.32$). A comparison of the data calculated with accepted methodology with actual data showed that the RMSE was 13.89 and the linear correlation value was 0.584. This low accuracy was probably due to the fact that 0.5 m intervals were used for comparison, while the accepted methodology was designed for intervals with thickness more than 2 m.

5. Conclusions

For efficient and safe uranium mining using the ISL method, it was necessary to determine as accurately as possible the lithological composition of the host rocks and the depth of permeable and impermeable strata, as well as the filtration properties of the host rocks. The necessary dependencies could be determined by comparing data from hydrogeological studies carried out at the exploration stage with logging data. In Kazakhstan, a method for determining the filtration properties was used, developed more than 50 years ago, which often produced incorrect results. Inaccuracies in the assessment of the filtration properties of host rocks led to the incorrect assessment of recoverable reserves and poor production planning. The currently accepted technique was based on the dependence of K_f on the average value of the AR within the boundaries of the selected lithological layer and had the following disadvantages:

- Only the AR curve data were used; if the AR data were poorly recorded, the results would be unreliable.
- When interpreting data from acidified blocks, where the properties of rocks were distorted by the action of acid, the values of the AR turned out to be underestimated, and therefore, the calculation of the filtration properties was not correct.
- Since a downhole tool with a distance between electrodes of 1 m was used to record AR logs in the fields of Kazakhstan, it was possible to reliably measure the average resistivity value only for lithological layers with a thickness of more than 2 m. Therefore, the adopted technique was not suitable for thin intervals (<2 m).

To overcome the shortcomings of the existing approach, we proposed a method for calculating the filtration coefficient based on the use of regression models [48]. The proposed model received electric logging data as an input and the calculated filtration coefficient as an output.

Thanks to a wider range of regression models, it was possible to improve the previously obtained result.

The LGBM regression model with AR and LC as the input variables demonstrated the best results ($R^2 = 0.710$, $R = 0.845$).

At the same time, the proposed method had the following limitations:

- It was not applicable to fields where exploration had been carried out for a long time, and not all the necessary data were available.
- The learning process depended on the lithological code set by the expert, which could be wrong, especially in the case of acidified blocks.

In this regard, directions for further research are the following:

- Exploring the possibility of transferring the trained model to similar fields for which there are no data required for training;
- Improving the reliability of determining the lithological code of the rock during lithological classification;
- Automatic identification of zones of technological acidification by the characteristic distortion of the AR curve.

Also, as one of the future research directions, it is suggested to use the construction of a forecasting model based on Fuzzy Decision Trees [49].

Author Contributions: Conceptualization, R.M., Y.K. and E.Z.; methodology, R.M.; software, A.S.; validation, Y.K., E.Z., V.L. and A.S.; formal analysis, R.M. and N.Y.; investigation, K.A., V.L., E.M. and Y.K.; resources, Y.K., R.M. and E.M.; data curation K.A., A.S. and V.L.; writing—original draft preparation, R.M., Y.K. and E.Z.; writing—review and editing, Y.K. and N.Y.; visualization, Y.K.; supervision, R.M.; project administration, Y.K.; funding acquisition, N.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP14869110 “Improving the accuracy of solving problems of interpretation of geophysical well research data on uranium deposits using machine learning methods”) and was partially supported by the Slovak Research and Development Agency, Slovakia, under the grant “Risk assessment of environmental disturbance using Earth observation data” (reg.no. SK-UA-21-0037).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors express their sincere gratitude to Viktors Gopejenko and Sergey Stankevich, who actively supported this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Uranium Reserves, Which Countries Have the Largest Reserves? Available online: <https://www.energy.com> (accessed on 28 June 2023).
2. World Nuclear Association. “Recent Uranium Production”, The Nuclear Fuel Report: Expanded Summary—Global Scenarios for Demand and Supply Availability 2019–2040. 2020. Available online: <https://world-nuclear.org/getmedia/b488c502-baf9-4142-8d12-42bab97593c3/nuclear-fuel-report-2019-expanded-summary-final.pdf.aspx> (accessed on 30 June 2023).
3. International Energy Agency (IEA). Key World Energy Statistics. Also Available on Smartphones and Tablets. 2016. Available online: <https://www.ourenergypolicy.org/wp-content/uploads/2016/09/KeyWorld2016.pdf> (accessed on 30 June 2023).
4. Mukhamediev, R.I.; Kuchin, Y.I.; Yakunin, K.O.; Mukhamedieva, E.L.; Kostarev, S.V. Preliminary results of the assessment of lithological classifiers for uranium deposits of the infiltration type. *Cloud Sci.* **2020**, *7*, 258–272.
5. Guidelines for Determining the Coefficient of Filtration of Water-Bearing Rocks by Experimental Pumping, Energoizdat. 1981. Available online: <https://www.geokniga.org/books/17383> (accessed on 15 June 2023).
6. Mukhamediev, R.I.; Popova, Y.; Kuchin, Y.; Zaitseva, E.; Kalimoldayev, A.; Symagulov, A.; Levashenko, V.; Abdoldina, F.; Gopejenko, V.; Yakunin, K.; et al. Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges. *Mathematics* **2022**, *10*, 2552.
7. Mukhamediev, R.I.; Symagulov, A.; Kuchin, Y.; Yakunin, K.; Yelis, M. From classical machine learning to deep neural networks: A simplified scientometric review. *Appl. Sci.* **2021**, *11*, 5541. [CrossRef]
8. Merembayev, T.; Yunussov, R.; Yedilkhan, A. Machine learning algorithms for stratigraphy classification on uranium deposits. *Procedia Comput. Sci.* **2019**, *150*, 46–52. [CrossRef]
9. Cracknell, M.J. Machine Learning for Geological Mapping: Algorithms and Applications. Ph.D. Thesis, University of Tasmania, Hobart, TAS, Australia, 2014.
10. Sun, T.; Li, H.; Wu, K.; Chen, F.; Zhu, Z.; Hu, Z. Data-driven predictive modeling of mineral prospectivity using machine learning and deep learning methods: A case study from southern Jiangxi Province, China. *Minerals* **2020**, *10*, 102. [CrossRef]
11. Dogan, A.; Birant, D.; Kut, A. Multi-target regression for quality prediction in a mining process. In Proceedings of the 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, 11–15 September 2019; pp. 639–644.
12. Deng, C.; Pan, H.; Fang, S.; Konate, A.A.; Qin, R. Support vector machine as an alternative method for lithology classification of crystalline rocks. *J. Geophys. Eng.* **2017**, *14*, 341–349. [CrossRef]

13. Kumar, C.; Chatterjee, S.; Oommen, T.; Guha, A. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *86*, 102006.
14. Cracknell, M.J.; Reading, A.M. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* **2014**, *63*, 22–33.
15. Harris, J.R.; Grunsky, E.C. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Comput. Geosci.* **2015**, *80*, 9–25. [[CrossRef](#)]
16. Kuchin, Y.I.; Mukhamediev, R.I.; Yakunin, K.O. One method of generating synthetic data to assess the upper limit of machine learning algorithms performance. *Cogent Eng.* **2020**, *7*, 1718821. [[CrossRef](#)]
17. Khan, H.; Srivastav, A.; Kumar Mishra, A.; Anh Tran, T. Machine learning methods for estimating permeability of a reservoir International. *J. Syst. Assur. Eng. Manag.* **2022**, *13*, 2118–2131.
18. Wong, P.M.; Henderson, D.J.; Brooks, L.J. Reservoir permeability determination from well log data using artificial neural networks: An example from the Ravva field, offshore India. In Proceedings of the Petroleum Development under Challenging Environments (with Special Emphasis on Gas), Kuala Lumpur, Malaysia, 14–16 April 1997; pp. 149–155.
19. Matinkia, M.; Hashami, R.; Mehrad, M.; Hajsaedi, M.R.; Velayati, A. Prediction of permeability from well logs using a new hybrid machine learning algorithm. *Petroleum* **2023**, *9*, 108–123. [[CrossRef](#)]
20. Rezaee, M.R.; Jafari, A.; Kazemzadeh, E. Relationships between permeability, porosity and pore throat size in carbonate rocks using regression analysis and neural networks. *J. Geophys. Eng.* **2006**, *3*, 370–376. [[CrossRef](#)]
21. Lim, J.S. Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea. *J. Pet. Sci. Eng.* **2005**, *49*, 182–192. [[CrossRef](#)]
22. Antoniuk, V.; Bezrodna, I.; Petrokushyn, O. Multiple regressions and ann techniques to predict permeability from pore structure for terrigenous reservoirs, west-shebelynska area Monitoring. *Eur. Assoc. Geosci. Eng.* **2019**, *2019*, 1–5.
23. Elkhatny, S.; Mahmoud, M.; Tariq, Z.; Abdulraheem, A. New insights into the prediction of heterogeneous carbonate reservoir permeability from well logs using artificial intelligence network. *Neural Comput. Appl.* **2018**, *30*, 2673–2683. [[CrossRef](#)]
24. Fajana, A.O.; Ayuk, M.A.; Enikanselu, P.A. Application of multilayer perceptron neural network and seismic multi-attribute transforms in reservoir characterization of Pennay field, Niger Delta. *J. Pet. Explor. Prod. Technol.* **2019**, *9*, 31–49. [[CrossRef](#)]
25. Ahmadi, M.A.; Ahmadi, M.R.; Hosseini, S.M.; Ebadi, M. Connectionist model predicts the porosity and permeability of petroleum reservoirs by means of petro-physical logs: Application of artificial intelligence. *J. Pet. Sci. Eng.* **2014**, *123*, 183–200. [[CrossRef](#)]
26. Talebkeikhah, M.; Sadeghtabaghi, Z.; Shabani, M. A comparison of machine learning approaches for prediction of permeability using well log data in the hydrocarbon reservoirs. *J. Hum. Earth Future* **2021**, *2*, 82–99. [[CrossRef](#)]
27. Zhong, Z.; Carr, T.R.; Wu, X.; Wang, G. Application of a convolutional neural network in permeability prediction: A case study in the Jacksonburg-Stringtown oil field, West Virginia, USA. *Geophysics* **2019**, *84*, 363–373. [[CrossRef](#)]
28. Kanfar, R.; Shaikh, O.; Yousefzadeh, M.; Mukerji, T. Real-time well log prediction from drilling data using deep learning. In Proceedings of the International Petroleum Technology Conference, Dhahran, Saudi Arabia, 13 January 2020.
29. Viggen, E.M.; Merciu, I.A.; Løvstakken, L.; Måsøy, S.E. Automatic interpretation of cement evaluation logs from cased boreholes using supervised deep neural networks. *J. Pet. Sci. Eng.* **2020**, *195*, 107539. [[CrossRef](#)]
30. Romanenkova, E.; Rogulina, A.; Shakirov, A.; Stulov, N.; Zaytsev, A.; Ismailova, L.; Kovalev, D.; Katterbauer, K.; AlShehri, A. Similarity learning for wells based on logging data. *J. Pet. Sci. Eng.* **2022**, *215*, 110690. [[CrossRef](#)]
31. Du, S.; Wang, R.; Wei, C.; Wang, Y.; Zhou, Y.; Wang, J.; Song, H. The connectivity evaluation among wells in reservoir utilizing machine learning methods. *IEEE Access* **2020**, *8*, 47209–47219. [[CrossRef](#)]
32. Sudakov, O.; Burnaev, E.; Koroteev, D. Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks. *Comput. Geosci.* **2019**, *127*, 91–98. [[CrossRef](#)]
33. Physical Basis for Determining the Lithological and Filtration Properties of Rocks of the Productive Horizon. 2018. Available online: https://ozlib.com/832945/tehnika/fizicheskie_osnovy_metoda (accessed on 10 July 2023).
34. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
35. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13 August 2016; pp. 785–794.
36. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.
37. Al Daoud, E. Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *Int. J. Comput. Inf. Eng.* **2019**, *13*, 6–10.
38. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [[CrossRef](#)]
39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
40. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
41. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [[CrossRef](#)]
42. Galushkin, A.I. *Theory of Neural Networks*; Hotline-Telecom; Springer: Berlin/Heidelberg, Germany, 2010; p. 496.

43. Yu, H.F.; Huang, F.L.; Lin, C.J. Dual coordinate descent methods for logistic regression and maximum entropy models. *Mach. Learn.* **2011**, *85*, 41–75. [[CrossRef](#)]
44. Santosa, F.; Symes, W. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comput.* **1986**, *7*, 1307–1330. [[CrossRef](#)]
45. Tikhonov, A.N.; Goncharsky, A.V.; Stepanov, V.V.; Yagola, A.G.; Tikhonov, A.N.; Goncharsky, A.V.; Yagola, A.G. *Numerical Methods for the Approximate Solution of Ill-Posed Problems on Compact Sets*; Springer: Dordrecht, The Netherlands, 1995; pp. 65–79.
46. Hoerl, A.E.; Kennard, R.W. Ridge regression: Applications to nonorthogonal problems. *Technometrics* **1970**, *12*, 69–82. [[CrossRef](#)]
47. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [[CrossRef](#)]
48. Mukhamediev, R.I.; Kuchin, Y.; Amirgaliyev, Y.; Yunicheva, N.; Muhamedijeva, E. Estimation of Filtration Properties of Host Rocks in Sandstone-Type Uranium Deposits Using Machine Learning Methods. *IEEE Access* **2022**, *10*, 18855–18872. [[CrossRef](#)]
49. Zaitseva, E.; Levashenko, V. Construction of a reliability structure function based on uncertain data. *IEEE Trans. Reliab.* **2016**, *65*, 1710–1723. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.