N.N. Kiselyova[a][0000-0002-3583-0704], **V.A. Dudarev**[a][0000-0001-7243-9096], **O.V. Senko**[b], **A.A. Dokukin**[a,b], **A.V. Stolyarenko**[a][0000-1111-2222-3333], **Yu.O. Kuznetsova**[a]

Double halides physical properties prediction by means of machine learning methods

[a] Baikov Institute of Metallurgy and Materials Science the Russian Academy of Sciences, Moscow, 119334 Russia

[b] Federal Research Center "Computer Sciences and Control" of the Russian Academy of Sciences, Moscow, 119333 Russia

e-mails: kis@imet.ac.ru; vic@imet.ac.ru

Abstract - The effectiveness of various machine learning methods in predicting the quantitative properties of inorganic compounds is compared. To assess the accuracy, cross-validation in the LOOCV mode was used. It has been shown that the use of algorithms Ridge, Bayesian Ridge, ARD Regression, Elastic net, as well as Extra Trees Regressor; Gradient Boosting Regressor; Hist Gradient Boosting Regressor, based on the methodology of ensembles of machine learning algorithms, from the scikit-learn package, as well as Simple Syndrome Regressor, Convex Linear Regression, and Recursive Regressor, specially designed for solving problems in the field of inorganic materials science, allows you to get the most accurate estimates. Using the selected machine learning programs, we predicted the melting temperature at atmospheric pressure of binary halides of the compositions $ABHal_3$, $ABHal_4$, A_2BHal_4, A_2BHal_5, and A_3BHal_6 (A and B are different elements, Hal is F, Cl, Br, or I) and estimated their crystal lattice parameters at normal conditions. When estimating melting temperatures, the average absolute errors (MAE) determined by the cross-validation method were in the range of 29-52 K, depending on the composition of the halides and the chosen algorithm). The coefficient of multiple determination R^2 for the models used for predicting is not lower than 0.7. When estimating the lattice parameters, the MAE for linear parameters was within 0.0068–0.2120 Å and 0.1209–0.1562˚ for angles. R^2 was above 0.6.

Keywords and phrases: machine learning, halide, melting point, crystal lattice parameters

1. INTRODUCTION

The search and study of halides is associated with the development of new superionic conductors [1], luminescent [2], magnetic [3], acousto-optical [4], nonlinear optical [5], electro-optical [6], and ferroelectric materials [7]. In this regard, for many years, specialists have been trying to find ways to calculate and predict these compounds and evaluate their properties: from quantum mechanical approaches [8, 9] to the development of empirical criteria, for example, [10–12], and recent studies have become a consequence of serious mathematical difficulties in calculating multi-electron systems using quantum mechanical methods.

The development of empirical criteria is based on the assumption that the possibility of formation of inorganic compounds and the type of their crystal structure depend on the size and charge of ions, electronegativity values or ionization potentials. However, the experience of applying such criteria shows that they do not always provide sufficiently acceptable prediction results. To obtain reliable predictions, the criteria for the formation of inorganic compounds should include a wide range of properties of components - chemical elements and simpler compounds: dimensional, thermodynamic, energy parameters, information about the position of elements in the Periodic system, etc. To solve such a complex problem, associated with the enumeration of a huge number of possible combinations of parameter values, we proposed to use special methods for discovering regularities based on machine learning algorithms [13-16].

In recent years, our approach to predicting halides has been widely used. For example, the authors of [17, 18] successfully applied it to predict the compound of the $ABHal_3$ composition with the perovskite crystal structure, and in [19, 20] the crystal lattice parameter of perovskites of the A_2XHal_6 composition was estimated.

This study includes the results of applying machine learning methods to predict lattice parameters and estimate the melting point at atmospheric pressure of double halides of various compositions. Some of these compounds were predicted earlier in our works [14, 15].

2. METHOD OF CALCULATION

For calculations, we used the information-analytical system (IAS) developed by us [21], which includes databases (DB) on the properties of inorganic substances and machine learning programs designed to predict categorical (discrete) properties. To predict the parameters of the crystal lattice and melting points, we used the system ParIS (Parameters of Inorganic Substances) [22] developed by us, which is designed to predict the quantitative properties of substances.

The IAS software is based on a wide class of algorithms of the RECOGNITION multifunctional system developed at the FRC Computing Center of the Russian Academy of Sciences (CCAS) [23], which combines, in addition to the well-known linear machine methods, Fisher's linear discriminant, k-nearest neighbors, support vectors machine, and neural network algorithms, also algorithms developed at the CCAS: machine learning algorithms based on the estimates calculation, deadlock test algorithm, voting algorithms for logical regularities, statistical weighted syndromes, etc. An important subsystem that makes it possible to improve the accuracy of predicting is a set of programs for the collective decision-making of the RECOGNITION system. The IAS data analysis subsystem currently includes 15 machine learning programs and 9 collective decision-making programs.

At present, the ParIS system includes multilevel machine learning programs specially developed by us [24–27], as well as more than 30 programs from the opened distributed scikit-learn package [28, 29]. The software system includes two original regression methods developed at the CCAS: a regression algorithm based on decorrelated ensembles (RADE) and convex linear regression (CLR) algorithm based on the search for optimal convex combinations of weak predictors. In the RADE method, an ensemble of regression trees is constructed based on the condition of minimum squared prediction error, which is achieved by combining the high predictive ability of individual trees with a high prediction variance within the ensemble [25, 27]. The CLR method is based on searching for ensembles of locally optimal convex combinations of simple regression models [24]. Simple averaging or an elastic network model are used as second level aggregating algorithms.

To assess the accuracy (the ratio of the number of substances for which the belonging to the given classes was correctly recognized to the total number of recognized substances), the IAS used a cross-validation on the material of the training sample. The most accurate algorithms were further used for making a collective decision. At this stage, the most accurate algorithm was also chosen in the IAS, for which examination recognition of a given number of substances was used, randomly selected from training samples and not used in machine learning (at the final stage of predicting, control examples were returned to the training sample). The ensemble of the best algorithms was used further in prediction. The learning quality assessment subsystem in the ParIS system allows you to evaluate the mean absolute errors (MAE) and mean squared errors (MSE) (with cross-validation in the LOOCV - Leave-One-Out Cross-Validation mode), the coefficient of determination R^2 and etc., as well as construct a diagram of the deviations of the calculated values of the parameters from the experimental values for substances, information about which was used in machine learning (see, for example, Fig. 1-3). The most accurate algorithms were further used for prediction. The prediction results for each compound were averaged.

The data for machine learning was extracted from the "Phases" database developed by us [30]. The feature description included the parameters of chemical elements from our "Elements" database [31] and the properties of simple halides from the IAS. The IAS includes a special subsystem that enables generation of different algebraic functions from the original features and selection of the most informative ones for classification. It should be noted that the "Phases" database stores "raw" data extracted from publications, therefore, an important role is played by an expert assessment of information quality, which is usually decided by a specialist preparing a sample for subsequent computer analysis. This task cannot be completely formalized. That's why we proposed to use for this purpose the deviations diagrams analysis of the parameters calculated values the from the experimental ones (see, for example, Fig. 1-3). They make it possible to reduce the time for examination, because the specialist immediately receives information about

the compound composition, the experimental and calculated values, by "clicking" on the point on the diagram that deviates the most from the experimental value. To filter outliers when predicting categorical properties, a special system has also been developed.

3. RESULTS

With the help of IAS, we have predicted compounds with the $ABHal_3$, $ABHal_4$, A_2BHal_4, A_2BHal_5, and A_3BHal_6 compositions, which have not yet been obtained, and also predicted their crystal structure type and space group at room temperature and atmospheric pressure. The lattice parameters and/or melting points of the predicted compounds and those double halides for which no information on these properties was available were estimated using the ParIS system.

For each of the compositions of the $ABHal_3$, $ABHal_4$, A_2BHal_4, A_2BHal_5, and A_3BHal_6 double halides, the problems of prediction the melting point and different lattice parameters were solved separately. The number of examples for machine learning during the melting temperature prediction was as follows: $ABHal_3$ - 241, $ABHal_4$ - 229, A_2BHal_4 - 204, A_2BHal_5 - 141 and A_3BHal_6 - 439. During lattice parameters prediction, the number of examples was: for $ABHal_3$ - 47 examples of compounds with a crystal structure of the perovskite (sp.gr. Pm(-)3m) type, 40 - $CsNiCl_3$ (sp.gr. $P6_3/mmc$), 25 - NH_4CdCl_3 (sp.gr. Pnma), $ABHal_4$ – 12 - scheelite (sp.gr. $I4_1/a$), 20 - $BaSO_4$ (sp.gr. Pnma), 21 - $NaNdF_4$ (sp.gr. P(-)6), A_2BHal_4 – 49 - K_2NiF_4 (sp.gr. I4/mmm), 50 - β-K_2SO_4 (sp.gr. Pnam), A_2BHal_5 – 18 - Cs_2DyCl_5 (sp.gr. Pbnm), 64 - K_2PrCl_5 (sp.gr. Pnma), and A_3BHal_6 – 45 - cryolite (sp.gr. $P2_1/n$), 35 - Rb_3TlF_6 (sp.gr. I4/mmm).

For each group of compounds with a certain space group, a training data sample was formed, which included in the feature description the following chemical elements (A, B and Hal) properties: pseudopotential orbital radius (according to Zunger), ionic radius (according to Shannon), distances to internal and to valence electrons (according to Schubert), ionization energies of the first, second and third electrons, numbers (according to Mendeleev-Pettifor), quantum number, electronegativity (according to Pauling), Miedema chemical potential, melting and boiling points, standard entropy, enthalpy of atomization, thermal conductivity, molar heat capacity, etc. (for the values, see the "Elements" DB [30]). A total of 87 element parameter values were used for each compound in melting temperature prediction problems.

Table 1 shows the examination evaluation results (LOOCV mode) for the most accurate algorithms from the ParIS system in the melting temperatures prediction of double halides [16]. It should be noted that MAE values, determined by the LOOCV method, were in the range of 29–52 K, depending on the halides composition and the chosen algorithm. The R^2 values for the models, used for prediction, were not lower than 0.7.

In most cases, the best performance was obtained using the ETR, GBR, HGBR algorithms from the scikit-learn package [28, 29] and the SAND and RADE methods specially developed for the ParIS system, which is a combination of several methods from the scikit-learn package (BR, GBR or Random Forest Regressor) with the method of clustering the array of chemical element parameters according to the degree of correlation. The efficiency of the ETR, GBR, HGBR, RADE algorithms can be explained by the fact that they are based on the principle of creation of algorithms ensembles that make it possible to compensate disadvantages of one algorithm with the advantages of others, making the final solution more sustainable. The use of such algorithms is especially preferable for solving problems in which training samples are relatively small [16]. The examination recognition results are illustrated in Figure1.

Table 1. Prediction accuracy assessment results for the double halides melting point [16]

Composition	Algorithm	R^2	MAE, K	MSE, K
$ABHal_4$	ETR	0.887	47	7362
- // -	GBR	0.888	51	7276
- // -	HGBR	0.907	49	6021
- // -	SAND	0.897	46	6716
- // -	BR	0.898	49	6615
A_2BHal_4	ETR	0.905	30	2037

Composition	Algorithm	R^2	MAE, K	MSE, K
- // -	GBR	0.905	34	2030
- // -	HGBR	0.907	34	1992
- // -	RADE	0.919	29	1747
A_2BHal_5	KNR	0.727	52	4947
- // -	GBR	0.719	48	5097
- // -	HGBR	0.701	51	5423
A_3BHal_6	ETR	0.910	35	3399
- // -	GBR	0.900	38	3747
- // -	HGBR	0.906	38	3538
- // -	RADE	0.911	35	3340
$ABHal_3$	ETR	0.898	43	5690
- // -	GBR	0.892	49	5987
- // -	HGBR	0.892	51	5974
- // -	RADE	0.909	42	5052

Notes: ETR - Extra Trees Regressor; GBR - Gradient Boosting Regressor; HGBR - Hist Gradient Boosting Regressor; SAND - Simple Syndrome Regressor; BR - Bagging Regressor; KNR – K-Neighbors Regressor.

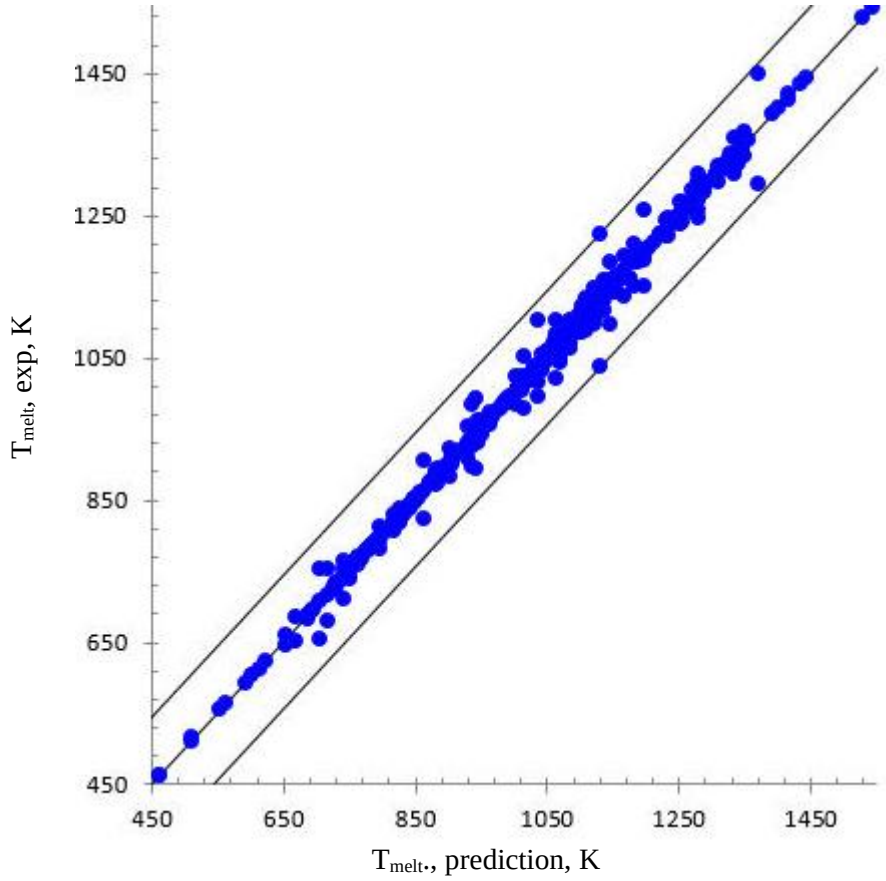

Fig. 1. Diagram of deviations of the predicted melting points of the A_3BHal_6 compounds from the experimental ones for the ETR algorithm [16].

When predicting the crystal lattice parameters, thermophysical properties of simple halides (melting point, standard entropy, isobaric heat capacity, heat and isobaric potential of formation of a simple halide) were also added to the feature description, due to which the number of features increased to 98 values.

Table 2 gives the examination evaluation results (LOOCV mode) of the most accurate algorithms from the ParIS system in the crystal lattice parameters prediction of double halides.

Table 2. Prediction accuracy assessment results for the double halides lattice parameters

Crystal structure type	Lattice parameter	Algorithm	R^2	Dimension	MAE	MSE
AB_2Hal_4						
K_2NiF_4	*a*	Ridge	0.991	Å	0.0686	0.0020
- // -	- // -	BaR	0.990	Å	0.0389	0.0022
- // -	- // -	EN	0.994	Å	0.0275	0.0014
- // -	*c*	ARD	0.983	Å	0.1416	0.0321
- // -	- // -	TR	0.970	Å	0.2023	0.0562
- // -	- // -	PLS	0.970	Å	0.2120	0.0562
- // -	- // -	EN	0.981	Å	0.1430	0.0788
K_2SO_4	*a*	Ridge	0.970	Å	0.0830	0.0174
- // -	- // -	ARD	0.988	Å	0.0618	0.0071
- // -	- // -	CLR	0.984	Å	0.0671	0.0090
- // -	*b*	RADE	0.940	Å	0.1050	0.0220
- // -	- // -	DTR	0.943	Å	0.1107	0.0210
- // -	- // -	SAND	0.957	Å	0.1026	0.0158
- // -	*c*	Ridge	0.939	Å	0.1735	0.0651
- // -		ARD	0.929	Å	0.2032	0.7757
- // -		TR	0.929	Å	0.1878	0.0759
- // -		EN	0.930	Å	0.1893	0.0749
A_2BHal_5						
Cs_2DyCl_5	*a*	CLR	0.901	Å	0.0729	0.0122
- // -	- // -	PLS	0.898	Å	0.0694	0.0126
- // -	- // -	ETR	0.892	Å	0.0649	0.0133
- // -	*b*	TR	0.662	Å	0.0497	0.0073
- // -	- // -	PLS	0.767	Å	0.0441	0.0051
- // -	- // -	EN	0.730	Å	0.0516	0.0059
- // -	*c*	Ridge	0.963	Å	0.0205	0.0012
- // -	- // -	BaR	0.963	Å	0.0204	0.0012
- // -	- // -	ARD	0.968	Å	0.0215	0.0010
- // -	- // -	DTR	0.979	Å	0.0201	0.0007
K_2PrCl_5	*a*	Ridge	0.998	Å	0.0232	0.0011
- // -	- // -	ARD	0.998	Å	0.0174	0.0008
- // -	- // -	DTR	0.998	Å	0.0247	0.0011
- // -	- // -	ETR	0.998	Å	0.0208	0.0009
- // -	- // -	GBR	0.998	Å	0.0206	0.0008
- // -	*b*	DTR	0.969	Å	0.0465	0.0082
- // -	- // -	GBR	0.972	Å	0.0344	0.0065
- // -	- // -	RF	0.969	Å	0.0529	0.0072
- // -	*c*	ARD	0.999	Å	0.0119	0.0003
- // -	- // -	DTR	0.999	Å	0.0116	0.0002
- // -	- // -	ETR	0.999	Å	0.0104	0.0002
- // -	- // -	GBR	0.999	Å	0.0117	0.0003
A_3BHal_6						
cryolite	*a*	Ridge	0.995	Å	0.0254	0.0026
- // -	- // -	BaR	0.995	Å	0.0265	0.0027
- // -	- // -	ARD	0.995	Å	0.0248	0.0025
- // -	*b*	Ridge	0.998	Å	0.0200	0.0010
- // -	- // -	BaR	0.999	Å	0.0183	0.0009
- // -	- // -	ARD	0.998	Å	0.0205	0.0010
- // -	*c*	Ridge	0.999	Å	0.0273	0.0016
- // -	- // -	BaR	0.999	Å	0.0206	0.0009
- // -	- // -	ARD	0.999	Å	0.0182	0.0009

Crystal structure type	Lattice parameter	Algorithm	R^2	Dimension	MAE	MSE
- // -	β	Ridge	0.696	deg	0.1441	0.0531
- // -	- // -	OMP	0.690	deg	0.1562	0.0541
- // -	- // -	ETR	0.691	deg	0.1209	0.0539
Rb_3TlF_6	*a*	Ridge	0.994	Å	0.0231	0.0011
- // -	- // -	BaR	0.993	Å	0.0234	0.0011
- // -	- // -	EN	0.992	Å	0.0232	0.0012
- // -	*c*	Ridge	0.976	Å	0.0615	0.0095
- // -	- // -	BaR	0.975	Å	0.0635	0.0099
- // -	- // -	ARD	0.975	Å	0.0609	0.0097
- // -	- // -	EN	0.976	Å	0.0587	0.0095
$ABHal_3$						
$CsNiCl_3$	*a*	Ridge	0.996	Å	0.0181	0.0008
- // -	- // -	BaR	0.996	Å	0.0181	0.0008
- // -	- // -	HR	0.997	Å	0.0167	0.0007
- // -	*c*	Ridge	0.978	Å	0.0347	0.0033
- // -	- // -	BaR	0.978	Å	0.0346	0.0034
- // -	- // -	ARD	0.978	Å	0.0341	0.0033
NH_4CdCl_3	*a*	RADE	0.962	Å	0.0694	0.0132
- // -	- // -	BaR	0.953	Å	0.0653	0.0162
- // -	- // -	NuSVR	0.953	Å	0.0668	0.0161
- // -	*b*	Ridge	0.916	Å	0.0404	0.0074
- // -	- // -	TR	0.917	Å	0.0625	0.0074
- // -	- // -	NuSVR	0.931	Å	0.0493	0.0061
- // -	*c*	Ridge	0.980	Å	0.1204	0.0234
- // -	- // -	BaR	0.980	Å	0.1198	0.0238
- // -	- // -	EN	0.981	Å	0.1194	0.0227
perovskite	*a*	Ridge	0.981	Å	0.0272	0.0035
- // -	- // -	OMP	0.981	Å	0.0262	0.0033
- // -	- // -	BaR	0.981	Å	0.0282	0.0035
- // -	- // -	ARD	0.981	Å	0.0277	0.0034
$ABHal_4$						
$BaSO_4$	*a*	ARD	0.995	Å	0.0426	0.0076
- // -	- // -	DTR	0.995	Å	0.0445	0.0078
- // -	- // -	ETR	0.995	Å	0.0445	0.0078
- // -	*b*	OMP	0.999	Å	0.0183	0.0006
- // -	- // -	DTR	0.999	Å	0.0198	0.0007
- // -	- // -	ETR	0.999	Å	0.0194	0.0006
- // -	*c*	ARD	0.999	Å	0.0218	0.0013
- // -	- // -	DTR	0.999	Å	0.0170	0.0009
- // -	- // -	ETR	0.999	Å	0.0140	0.0004
- // -	- // -	GBR	0.999	Å	0.0183	0.0011
$NaNdF_4$	*a*	OMP	0.969	Å	0.0109	0.0002
- // -	- // -	ETR	0.942	Å	0.0129	0.0004
- // -	- // -	SAND	0.924	Å	0.0137	0.0006
- // -	*c*	NuSVR	0.963	Å	0.0129	0.0003
- // -	- // -	ETR	0.935	Å	0.0178	0.0006
- // -	- // -	SAND	0.937	Å	0.0163	0.0006
scheelite	*a*	Ridge	0.999	Å	0.0068	0.0001
- // -	- // -	BaR	0.999	Å	0.0068	0.0001
- // -	- // -	ARD	0.999	Å	0.0069	0.0001

Crystal structure type	Lattice parameter	Algorithm	R^2	Dimension	MAE	MSE
- // -	- // -	ETR	0.999	Å	0.0069	0.0001
- // -	*c*	EN	0.993	Å	0.0188	0.0023
- // -	- // -	GBR	0.999	Å	0.0151	0.0005
- // -	- // -	CLR	0.989	Å	0.0351	0.0037

Notes: see Tabl.1; BaR - Bayesian Ridge; ARD - ARD Regression; EN - Elastic net; TR - Tweedie Regressor; PLS - PLS Regression; CLR - Convex with loop reduction and elastic net; DTR - Decision Tree Regressor; RF - Random Forest; OMP - Orthogonal Matching Pursuit; HR - Huber Regressor; svm.NuSVR - NuSVR.

As in solving the previous problems, the most accurate algorithms list included ensemble methods. The Ridge, Bayesian Ridge, ARD Regression, Elastic net algorithms were also the most effective in solving almost all tasks. Analysis of Table 2 shows that MAE values, determined by the LOOCV method for linear parameters, are within 0.0068-0.2120 Å and 0.1209-0.1562˚ for angles, depending on the halides composition and the chosen algorithm. The R^2 coefficients for the models used for predicting were not less than 0.6. Figures 2 and 3 illustrate the prediction accuracy.

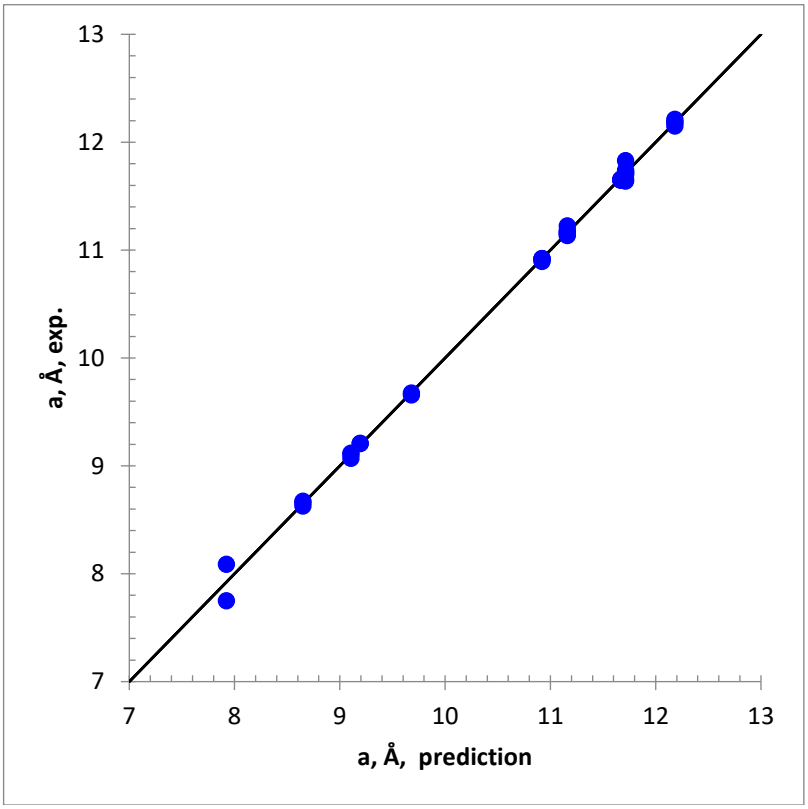

Fig. 2. Diagram of deviations of the predicted parameter *a* of the $BaSO_4$ type crystal lattice for $ABHal_4$ composition compounds from the experimental ones for the ARD algorithm.

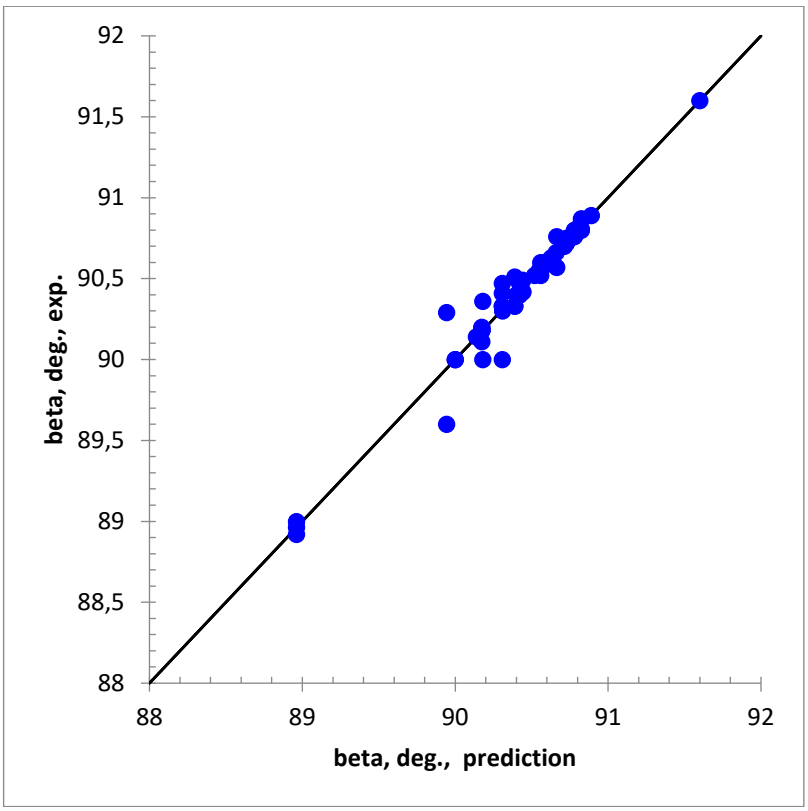

Fig. 3. Diagram of deviations of the predicted crystal lattice angle β of the cryolite type for A_3BHal_6 composition compounds from the experimental ones for the ETR algorithm.

Table 3 gives a part of the results for the crystal lattice parameters and the melting points prediction for double halides.

Table 3. Results of estimation of the crystal lattice parameters and melting point for halides

Composition	**Crystal structure type**	**Lattice parameters**				**$T_{melt.}$, K [16]**
		a, Å	*b*, Å	*c*, Å	β, град.	
AB_2Hal_4						
Tl_2MgF_4	K_2NiF_4	4.1450		10.6594		890
Rb_2TiF_4	- // -	4.1255		13.8739		1097
Cs_2TiF_4	- // -	4.2017		14.6240		1064
Rb_2VF_4	- // -	4.1953		13.7886		1070
K_2CrF_4	- // -	4.1152		13.0037		1075
Rb_2CrF_4	- // -	4.1824		13.6417		1040
Cs_2CrF_4	- // -	4.2585		14.3918		1014
Tl_2CrF_4	- // -	4.2697		10.4343		889
K_2FeF_4	- // -	4.1242		13.0923		1034
Tl_2MnF_4	- // -	4.3250		10.6557		887
Cs_2FeF_4	- // -	4.2675		14.4804		1002
Tl_2FeF_4	- // -	4.2787		10.5229		872
Tl_2CoF_4	- // -	4.2413		10.4461		870
Cs_2NiF_4	- // -	4.1630		14.4027		1094
Tl_2NiF_4	- // -	4.1742		10.4452		897
Cs_2YbCl_4	- // -	5.4822		17.6030		884
K_2CuCl_4	K_2SO_4	8.7133	7.6048	11.8991		701
K_2BeBr_4	- // -	9.0360	7.1132	12.4638		650
Rb_2BeBr_4	- // -	9.6150	7.3550	12.7933		799
Cs_2BeBr_4	- // -	10.0381	7.4264	12.9454		860

Composition	Crystal structure type	Lattice parameters				$T_{melt.}$, K [16]
		a, Å	*b*, Å	*c*, Å	*β*, град.	
Rb_2FeBr_4	- // -	9.7053	7.5737	13.3458		672
Cs_2FeBr_4	- // -	10.1284	7.6136	13.4978		800
Cs_2CuBr_4	- // -	10.1500	7.6988	13.0209		773
Cs_2ZnI_4	- // -	8.8395	7.4743	13.3383		771
A_2BHal_5						
Cs_2UCl_5	Cs_2DyCl_5	15.2457	9.5246	7.5204		926
K_2PmCl_5	K_2PrCl_5	12.6765	8.6706	7.9480		819
Rb_2PmCl_5	- // -	13.0975	8.7808	8.1673		823
Rb_2DyCl_5	- // -	13.0432	8.4397	8.1482		802
K_2PmBr_5	- // -	13.2989	9.1254	8.4090		775
K_2PuBr_5	- // -	13.2759	9.1597	8.4225		828
Rb_2PmBr_5	- // -	13.6431	9.2924	8.5874		778
Rb_2PuBr_5	- // -	13.6293	9.3453	8.5973		824
Cs_2LaBr_5	- // -	13.9190	9.4030	8.6545		806
Cs_2CeBr_5	- // -	13.8981	9.3706	8.6546		826
Cs_2PrBr_5	- // -	13.8828	9.3283	8.6371		809
Cs_2NdBr_5	- // -	13.8736	9.3004	8.6258		804
Rb_2PuI_5	- // -	14.5247	10.0045	9.2465		789
A_3BHal_6						
Na_3PmBr_6	cryolite	7.3082	7.7866	10.9123	90.5253	900
K_3TmI_6	- // -	8.1767	8.7521	12.0930	90.6897	948
K_3PuF_6	Rb_3TlF_6	6.6412		9.4680		1128
Rb_3PuF_6	- // -	6.7669		9.8963		1185
Ag_3TiF_6	- // -	5.8574		8.2302		1011
Ag_3MnF_6	- // -	5.7917		8.0888		975
Ag_3MoF_6	- // -	5.9219		8.5313		989
Cs_3PmF_6	- // -	6.9390		9.9114		1231
Cs_3PuF_6	- // -	7.0573		10.3984		1220
$ABHal_3$						
$TlTiCl_3$	$CsNiCl_3$	7.0215		5.9398		948
$TlCrCl_3$	- // -	6.9726		6.1256		805
$CsBeBr_3$	- // -	7.3518		6.2052		590
$RbFeBr_3$	- // -	7.3994		6.2874		687
$TlCoBr_3$	- // -	7.2153		6.2500		667
$TlNiBr_3$	- // -	7.1576		6.1487		881
$KHgCl_3$	NH_4CdCl_3	9.0924	4.2561	14.0111		512
$TlZnBr_3$	- // -	9.3234	4.0935	14.8748		656
$KHgBr_3$	- // -	9.5380	4.4482	14.6424		467
$TlHgI_3$	- // -	10.2022	4.6463	15.7729		510
$TlVF_3$	perovskite	4.1834				1093
$TlYbF_3$	- // -	4.5380				1079
$AgVF_3$	- // -	4.0245				1125
$RbEuF_3$	- // -	4.6963				1172
$ABHal_4$						
$RbBCl_4$	$BaSO_4$	10.9798	6.5251	8.8164		638
$AgBCl_4$	- // -	10.2773	6.4391	8.6998		526
$TlBCl_4$	- // -	10.7440	6.5023	8.7954		574
$CsVCl_4$	- // -	11.7073	7.1362	9.3858		878
$TlVCl_4$	- // -	10.8947	7.0517	9.2321		738
$TlCrCl_4$	- // -	10.8947	7.0622	9.2272		814
$RbBBr_4$	- // -	11.3968	6.7326	9.2869		581
$CsBBr_4$	- // -	11.9265	6.8115	9.3434		628
$RbVBr_4$	- // -	11.6411	7.4898	9.8712		791

Composition	Crystal structure type	Lattice parameters				$T_{melt.}$, K [16]
		a, Å	*b*, Å	*c*, Å	*β*, град.	
$RbCrBr_4$	- // -	11.6446	7.4970	9.8725		886
$CsCrBr_4$	- // -	12.1692	7.5029	9.8618		806
$AgCeF_4$	$NaNdF_4$	6.3308		3.6853		1037
$AgPrF_4$	- // -	6.3227		3.6580		1014
$AgPmF_4$	- // -	6.2973		3.6294		1083
$AgUF_4$	- // -	6.3153		3.6873		1045
$AgPuF_4$	- // -	6.2940		3.6730		1056
$LiScF_4$	scheelite	5.0507		10.2361		992

4. CONCLUSIONS

The efficiency of using various machine learning algorithms for predicting the inorganic compounds quantitative properties is compared. It is shown that the use of algorithms ensembles gives the most accurate results of examination evaluation in the LOOCV mode. Using the selected machine learning programs, we predicted the melting points at atmospheric pressure for binary halides of the $ABHal_3$, $ABHal_4$, A_2BHal_4, A_2BHal_5, and A_3BHal_6 compositions (A and B are different elements, Hal is F, Cl, Br, or I) and estimated their crystal lattice parameters at normal conditions. To predict the unknown values of the melting point and lattice parameters for the halides, only the properties values of the elements A, B, and Hal were used. The MAE values in the melting points assessment were in the range of 29-52 K depending on the composition of the halides and the selected algorithm. The R^2 multiple determination coefficients for the models, used for prediction, were not lower than 0.7. When estimating the lattice parameters, the MAE values for linear parameters were within 0.0068–0.2120 Å and 0.1209–0.1562˚ for angles, which corresponds to the experimental errors. The R^2 values were above 0.6.

FUNDING

The study was carried out as a part of the state assignments (project nos. 075-01176-23-00 and 0063- 2016-0003).

REFERENCES

1. Schlem R., Bernges T., Li C., Kraft M.A., Minafra N., Zeier W.G. “A lattice dynamical approach for finding the lithium superionic conductor Li_3ErI_6 “, ACS Appl.Energy Mater. **3**(4), 3684 – 3691 (2020).
2. Majher J.D., Gray M.B., Liu T., Holzapfel N.P., Woodward P.M. “Rb_3InCl_6: A Monoclinic Double Perovskite Derivative with Bright Sb^{3+}-Activated Photoluminescence”, Inorg. Chem. **59**(19), 14478 – 14485 (2020).
3. Kurzydlowski D., Jaron T., Ozarowski A., Hill S., Jaglicic Z., Filinchuk Y., Mazej Z., Grochala W. “Local and Cooperative Jahn–Teller Effect and Resultant Magnetic Properties of M_2AgF_4 (M = Na-Cs) Phases”, Inorg. Chem. **55**(21), 11479 - 11489 (2016).
4. Kaidan M. V., Zadorozhna A. V., Andrushchak A. S., Kityk A. V. “Cs_2HgCl_4 crystal as a new material for acoustooptical applications”, Opt. Mater. **22**(3), 263-268 (2003).
5. Gong P., Liang F., Kang L., Lin Z. “Mid-Infrared Nonlinear Optical Halides with Diamond-like Structures: A Theoretical and Experimental Study”, Chem. Mater. **34**(11), 5301 - 5310 (2022).
6. Walters G., Sargent E.H. "Electro-optic Response in Germanium Halide Perovskites", J. Phys. Chem. Lett. **9**(5), 1018-1027 (2018).
7. Zheng W., Wang X., Zhang X., Chen B., Suo H., Xing Z., Wang Y., Wei H.-L., Chen J., Guo Y., Wang F. "Emerging Halide Perovskite Ferroelectrics", Adv. Mater. 2205410/1-21 (2023).
8. Schwarz U., Wagner F., Syassen K., Hillebrecht H. “Effect of pressure on the optical-absorption edges of $CsGeBr_3$ and $CsGeCl_3$”, Phys. Rev. B. **53**(19), 12545/1-4 (1996).
9. Khan N.U., Abdullah, Khan U.A., Tirth V., Al-Humaidi J.Y., Refat M.S., Algahtani A., Zaman A. "Investigation of structural, opto-electronic and thermoelectric properties of titanium based chloro-perovskites $XTiCl_3$ (X = Rb, Cs): a first-principles calculations", RSC Adv. **13**(9), 6199-6209 (2023).
10. Jiang L. Q., Guo J. K., Liu H. B., Zhu M., Zhou X., Wu P., Li C.H. “Prediction of lattice constant in cubic perovskites”, J. Phys. and Chem. Solids. **67**(7), P. 1531-1536 (2006).

11. Li C., Lu X., Ding W., Feng L., Gao Y., Guo Z. "Formability of ABX_3 (X = F, Cl, Br, I) halide perovskites", Acta Crystallogr. B. **64**(6), 702-707 (2008).
12. Nalbandyan V., Novikova A. "Structural chemistry of A_2MX_4 compounds (X = O, F) with isolated tetrahedral anions: search for the densest structure types", Acta Crystallogr. B. **68**(3), 227-239 (2012).
13. Kiselyova N.N., Savitskii E.M. "Prediction of compounds with composition $ABHal_4$ in systems of halides", Zh. Neorgan. Khimii, **27**(11), 2880-2884 (1982) [in Russian].
14. Kiseleva N.N. "Prediction of New Compounds in Systems of Monovalent and Divalent Metal Halides", Russ. J. Inorg. Chem. **59**(5), 496-502 (2014).
15. Kiselyova N.N., Stolyarenko A.V., Ryazanov V.V., Sen'ko O.V., Dokukin A.A., Podbel'skii V.V. "Prediction of New Compounds in the AHal–$BHal_3$ Systems", Russ. J. Inorg. Chem. **59**(12), 1462-1471 (2014).
16. Kiselyova N.N.,Dudarev V.A.,Stolyarenko A.V.,Dokukin A.A.,Sen'ko O.V., Kuznetsova Yu.O. "Prediction of the melting points of binary halides", Russ. J. Inorg. Chem. **68**(5), 555-560 (2023).
17. Pilania G., Balachandran P.V., Kim C., Lookman T. "Finding New Perovskite Halides via Machine Learning", Front. Mater. **3** (Article 19), 1-7 (2016).
18. Gomez–Peralta J.I., Bokhimi X. "Discovering new perovskites with artificial intelligence", J. Solid State Chem. **285**, 121253/1-6 (2020).
19. Zhang Y., Xu X. "Machine learning lattice constants for cubic perovskite A_2XY_6 compounds", J. Solid State Chem. **291**, 121558/1-10 (2020).
20. Alade I.O., Olumegbon I.A., Bagudu A. "Lattice constant prediction of A_2XY_6 cubic crystals (A = K, Cs, Rb, Tl; X = tetravalent cation; Y = F, Cl, Br, I) using computational intelligence approach", J. Appl. Phys. **127**(1), 015303/1-9 (2020).
21. Kiselyova N.N., Stolyarenko A.V., Ryazanov V.V., Senko O.V., Dokukin A.A., Podbel'skii V.V. "A system for computer-assisted design of inorganic compounds based on computer training", Pattern Recognition and Image Analysis. **21**(1), 88–94 (2011).
22. Dudarev V.A., Kiselyova N.N., Stolyarenko A.V., Dokukin A.A., Senko O.V., Ryazanov V.V., Vashchenko E.A., Vitushko M.A., Pereverzev-Orlov V.S. "An Information System for Inorganic Substances Physical Properties Prediction Based on Machine Learning Methods", CEUR Workshop Proceedings (CEUR-WS.org), v. 2790. Supplementary Proceedings of the XXII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2020), 89 – 102 (2020). http://ceur-ws.org/Vol-2790/paper09.pdf
23. Zhuravlev Yu.I., Ryazanov V.V., Sen'ko O.V. «RASPOZNAVANIYE». Matematicheskiye metody. Programmnaya sistema. Prakticheskiye primeneniya ["Recognition". Mathematical methods. Program system. Practical applications]. Moscow, FAZIS Publ., 2006, 176 p. [in Russian].
24. Senko O.V., Dokukin A.A., Kiselyova N.N., Khomutov N.Yu. "Two-Stage Method for Constructing Linear Regressions Using Optimal Convex Combinations", Dokl. Mathematics. **97**(2), 113-114 (2018).
25. Zhuravlev Yu.I., Senko O.V., Dokukin A.A., Kiselyova N.N., Saenko I.A. "Two-Level Regression Method Using Ensembles of Trees with Optimal Divergence", Dokl. Mathematics. **104**(1), P.212-215 (2021).
26. Vashchenko E.A., Vitushko M.A., Dudarev V.A., Kiselyova N.N., Pereverzev-Orlov V.S. "On the possibility of predicting the parameter values of multicomponent inorganic compounds", Informatsionnye processy. **19**(4), 415-432 (2019). [in Russian]
27. Senko O.V., Dokukin .A.A., Kiselyova N.N., Dudarev V.A., Kuznetsova Yu. O. "New Two-Level Ensemble Method and Its Application to Chemical Compounds Properties Prediction", Lobachevskii Journal of Mathematics. **44**(1), 188–197 (2023).
28. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. "Scikit-learn: Machine Learning in Python", J. Machine Learning Research. **12**(Oct.), 2825 - 2830 (2011).
29. Scikit-learn module: https://github.com/scikit-learn/scikit-learn (visited on 14.08.2023)
30. DB "Phases": https://phase.imet-db.ru/ (visited on 14.08.2023).
31. DB "Elements": https://phase.imet-db.ru/elements/ (visited on 14.08.2023).

Responses to reviewers' comments

Reviewer 1

Remark 1:

This task cannot be completely formalized, however, we proposed to use for this purpose the deviations diagrams analysis of the parameters calculated values the from the experimental ones (see, for example, Fig. 1-3), which makes it possible to reduce the time for examination, because the specialist immediately receives information about the compound composition, the experimental and calculated values, by “clicking” on the point on the diagram that deviates the most from the experimental value.

Answer:

Corrected

Remark 2:

About references style (flavour): the authors use references to "anchor"/"seminal" research papers even when they speak of software. The example of this is scikit-learn, for which the authors use the reference to a paper as old as of 2011, whilst scikit-learn is a very popular software that has a well-established website and a code repository in Git. It would feel more appropriate to cite not only a seminal paper but also a well-established community website or/and a code repository for such popular tools.

Answer:

Reference was added.

Reviewer 2

Remark 1:

Большой успех авторов работы – предсказание характеристик и параметров для новых соединений, однако, хотелось бы увидеть экспериментальное подтверждение прогноза.

Answer:

В приложении к нашей статье [16] проведено сравнение результатов прогноза температур плавления галогенидов с известными экспериментальными данными для некоторых прогнозированных составов.

Reviewer 3

No remarks.