



Machine learning application for prediction of sapphire crystals defects



Yulia Vladimirovna Klunnikova^{a,*}, Maxim Vladimirovich Anikeev^b,
Alexey Vladimirovich Filimonov^c, Ravi Kumar^d

^aInstitute of Nanotechnology, Electronics and Engineering Equipment, Southern Federal University of Russia, Taganrog, 347928, Russia

^bInstitute of Computer Technology and Information Security, Southern Federal University of Russia, Taganrog, 347928, Russia

^cDepartment of Physical Electronics, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, 195251, Russia

^dDepartment of Metallurgical and Materials Engineering, Indian Institute of Technology Madras, Chennai, 600036, India

ARTICLE INFO

Publishing editor: Xuan Xie

Keywords:

Defects
Machine learning
Sapphire crystals

ABSTRACT

We investigate the impact of different numbers of positive and negative examples on machine learning for sapphire crystals defects prediction. We obtain the models of crystal growth parameters influence on the sapphire crystal growth. For example, these models allow predicting the defects that occur due to local overcooling of crucible walls in the thermal node leading to the accelerated crystal growth. We also develop the prediction models for obtained crystal weight, blocks, cracks, bubbles formation, and total defect characteristics. The models were trained on all data sets and later tested for generalization on testing sets, which did not overlap the training set. During training and testing, we find the recall, precision of prediction and analyze the correlation among the features. The results have shown that the precision of neural network method for predicting defects formed by local overcooling of the crucible reached 0.94.

1. Introduction

Machine learning methods are becoming increasingly popular in accelerating the design of new materials by predicting material properties. The minimization of various defects in the crystal structure is extremely important for the improvement and development of modern technologies for the artificial sapphire crystal growth.

Sapphire monocrystals find wide applications in microelectronics, optics, and electronic equipment engineering. They can be used as the substrates for integrated circuits with high resistance to radiation and heat combined with low power consumption. The defects formed in the crystal are one of the key factors affecting the properties of the substrates cut from them and the possibility of their applications in microelectronics and optoelectronics.

Theoretical and experimental studies were carried out with crystal growth by the Kyropoulos method [1–9]. The Kyropoulos method is used for industrial production of sapphire crystals with a diameter of 300 mm and more. S. Kyropoulos has proposed the process, in which the seed is placed into a water-cooled crystal holder and is introduced to the melt in the melting pot. The crystal grows on the seed in the form of a hemisphere and is deepened into the melt. When the growing crystal reaches the pot's surface, the crystal holder is lifted by several millimeters, and this process is repeated until the crystal has grown. Crystallographic orientation, the density and types of

* Corresponding author.

E-mail addresses: yvklunnikova@sfedu.ru (Y.V. Klunnikova), mvanikeev@sfedu.ru (M.V. Anikeev), filalex@inbox.ru (A.V. Filimonov), nvrk@iitm.ac.in (R. Kumar).

<https://doi.org/10.1016/j.jnlest.2020.100029>

Received 13 August 2019; Received in revised form 8 September 2019;

Available online 10 April 2020

1674-862X/© 2020 University of Electronic Science and Technology of China. Publishing Services provided by Elsevier B.V. on behalf of KeAi. This is

an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

point defects, the density of single dislocations, lengths of block boundaries and their misorientations, the level of residual stresses, the heterogeneity, and the chemical purity define the sapphire monocrystal's quality. Features of the crystal growing equipment, the total heating time, the degree of overheating, the seeding time, the speed of crystallization, the heater power output, the speed of the power decay, the voltage and its decay ratio, the crystal growth time, and the output of valid crystals (cracks, blocks, bubbles, residual stresses, etc.) are taken into account to predict the resulting quality of crystals [10]. Nowadays there is a number of different studies of the properties and crystal structure defects (X-ray, optical methods, atomic force microscopy, and others). However, experiments alone do not make it possible to evaluate how crystal obtaining parameters affect their properties. In order to fulfill the aims of the research, it is necessary to estimate the correlation of numerous components and to find their optimal combinations for the optimization of the industrial crystal growth performance.

2. Machine learning for materials prediction

The combination of experimental studies of the sapphire crystal growth by the Kyropolous method (see Fig. 1) and machine learning methods allow finding the dependencies of defects level in time and can reduce the number of unsuccessful attempts. The quantity and variety of defects in crystals are described by stochastic functions which depend on different technological parameters, such as the design features of the furnace, the temperature in different areas of the furnace and its variation over time, etc.

Prediction of crystal defects is an important and fundamental problem in materials science. Machine learning for the prediction of materials has been under investigation for several years. Only several works describe the neural network usage for the prediction of crystal properties [11–14]. Data mining tools allow predicting future trends and behavior which allows making decisions. The aim of this contribution is the incorporation of the findings into an overall prediction approach for the computational investigation of possible defects formation during the sapphire crystals growth.

Many studies in the defect prediction use techniques which originated from statistics and machine learning [15]. Such techniques include logistic regression, support vector machine (SVM), classification trees, neural networks, the naive Bayes algorithm, k -nearest neighbour (k NN) algorithm, CN2 algorithm, induction algorithm, adaptive boosting algorithm (AdaBoost), random forest algorithm, and stochastic gradient descent (SGD) algorithm [16].

The development and application concerning data mining algorithms require the use of powerful software tools. There are many available open source data mining tools like the Waikato environment for knowledge analysis (WEKA), Tanagra, the Konstanz information miner (KNIME), and Orange Canvas. In this paper, we use the Orange Canvas framework for defects prediction in sapphire crystals. Orange Canvas is a comprehensive, component-based framework for machine learning and data mining. It provides a platform for experiment selection, predictive modeling, and recommendation systems. Orange Canvas includes a set of components for data preprocessing, feature scoring and filtering, simulation, model evaluation, and exploration techniques. It is intended for both experienced users and researchers in machine learning who want to develop and test their algorithm, with the easy-to-use visual programming environment. Orange Canvas allows easy prototyping of new algorithms and experimental procedures. For explorative data analysis, it provides a visual programming framework with an emphasis on interactions and creative combinations of visual components [16–25].

We investigate the impact of different numbers of positive and negative examples on machine learning for sapphire crystals defects prediction. We obtain the models of crystal growth parameters influence on sapphire crystal growth. Training data are sent to the



Fig. 1. Sapphire crystals obtained (a) by the Kyropolous method and (b) enlarged sapphire crystals.

appropriate classifier (see Fig. 2). Then we compare the results from the classifiers with the real experiment. The main goal of the research was to assess whether the proposed approach may be used on the defects prediction.

3. Sapphire crystals defects prediction results

We present the receiver operating characteristic (ROC) analysis and comparison between the calculation parameters of Orange Canvas (see Fig. 3 and Table 1, where AUC is the area under ROC curve, CA is the classification accuracy, and F1 is the F-score). These models allow predicting, for example, the defects formed as a result of local overcooling of the crucible walls in the thermal node, leading to the accelerated crystal growth. We also receive the prediction models for obtained crystal weight, blocks, crack, bubbles formation, and total defect characteristics. The models obtained for all data sets and they were later used for generalization on a different data set, which does not include the data used on the training stage. During training and testing, we find the recall, precision of prediction and analyze the correlation among the features. The results show that the neural network precision for defects formed as a result of local overcooling of the crucible was 0.94. The neural network determines the current situation as a known state and reproduces its reaction as accurately as possible. The experimental studies of the sapphire crystals growth by the Kyropoulos method describe the dependence of the defects level in time, and neural networks as a machine learning instrument make it possible to derive new dependencies from this data and predict the obtained crystal quality. The precision of SVM and naive Bayes algorithms was 0.857 and 0.801, respectively. The experimental data set has to be extended to specify models, improving the recall and precision for defects prediction.

Data mining is often concerned with the development of predictive models. In order to apply predictive models in practice, they have to be integrated into the decision support systems. The comparison between calculation parameters of Orange Canvas can be applied for the universal expert system development for defects prediction during the sapphire crystals obtaining. The analysis allows the experts to find hidden information in data and improve the efficiency of prediction. The generalized structure of the expert system for defects prediction is presented in Fig. 4. These investigations allow us to improve the expert system for defects prediction in sapphire crystals. We demonstrate the robustness and the predictive power of our method by performing the determination of defects. The designed software is a universal tool for studying the influence of the crystal growth parameters on the quality of sapphire crystals. It can be widely used to estimate and predict the defects of growing crystals.

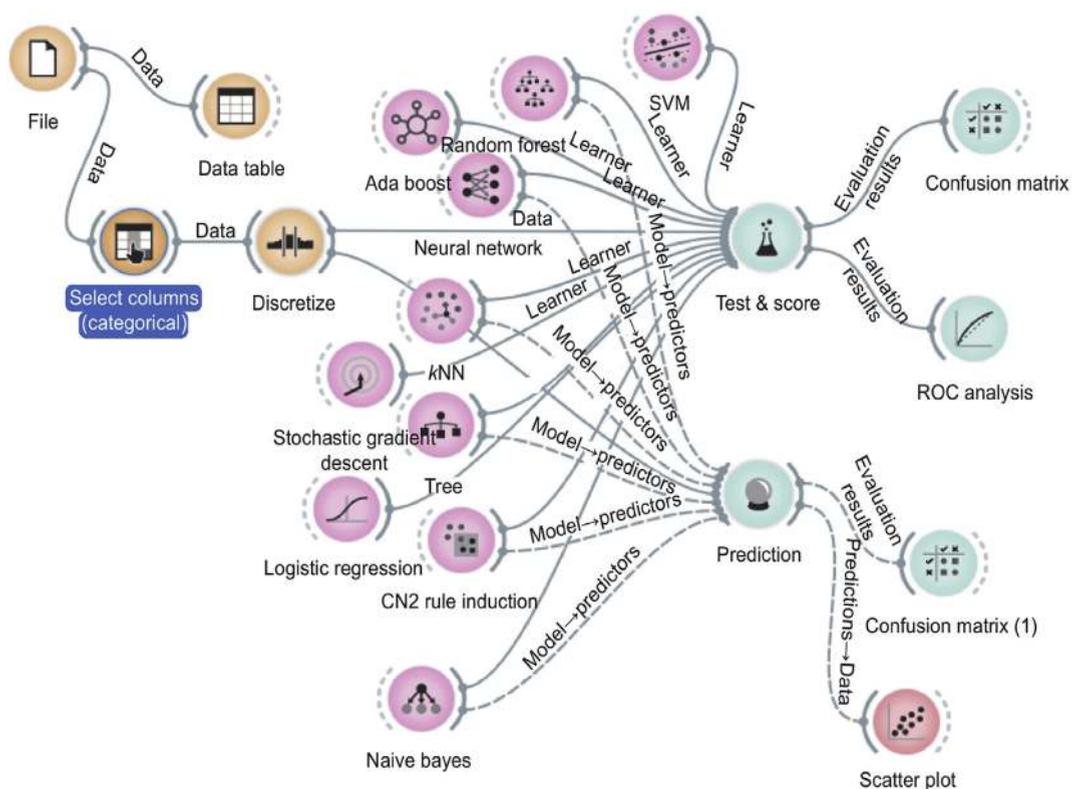


Fig. 2. Illustration for machine learning workflow.

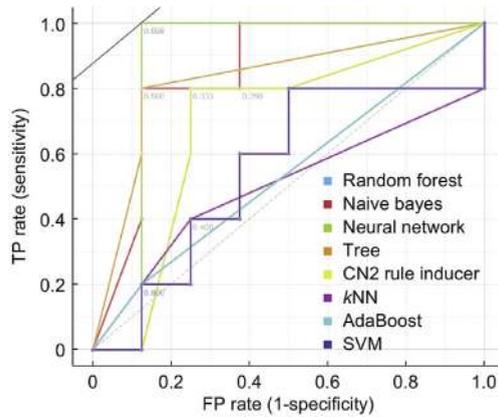


Fig. 3. ROC analysis of calculation parameters in Orange Canvas. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Comparison between calculation parameters of Orange Canvas.

Method	AUC	CA	F1	Precision	Recall
Neural network	0.889	0.929	0.930	0.940	0.929
SVM	0.578	0.857	0.857	0.857	0.857
Naive Bayes	0.844	0.786	0.789	0.801	0.786
Tree	0.833	0.786	0.779	0.782	0.786
CN2 rule inducer	0.700	0.714	0.714	0.714	0.714
Random forest	0.889	0.714	0.693	0.706	0.714
AdaBoost	0.544	0.643	0.592	0.607	0.643
kNN	0.544	0.643	0.503	0.413	0.643

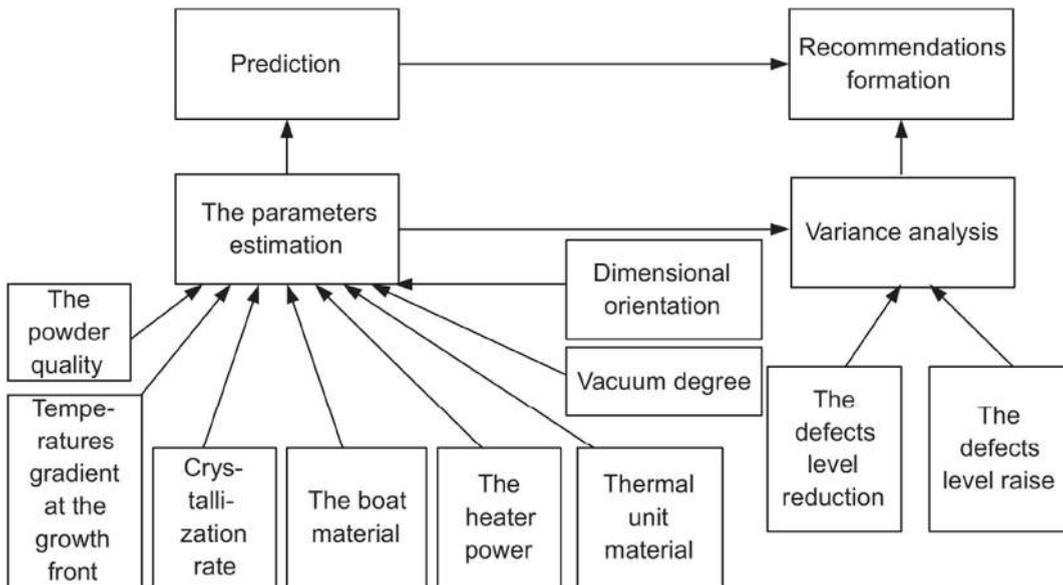


Fig. 4. Generalized structure of the expert system knowledge base.

The class diagrams of the expert system for the sapphire defects prediction can be seen in Fig. 5. In this diagram, the main classes include:

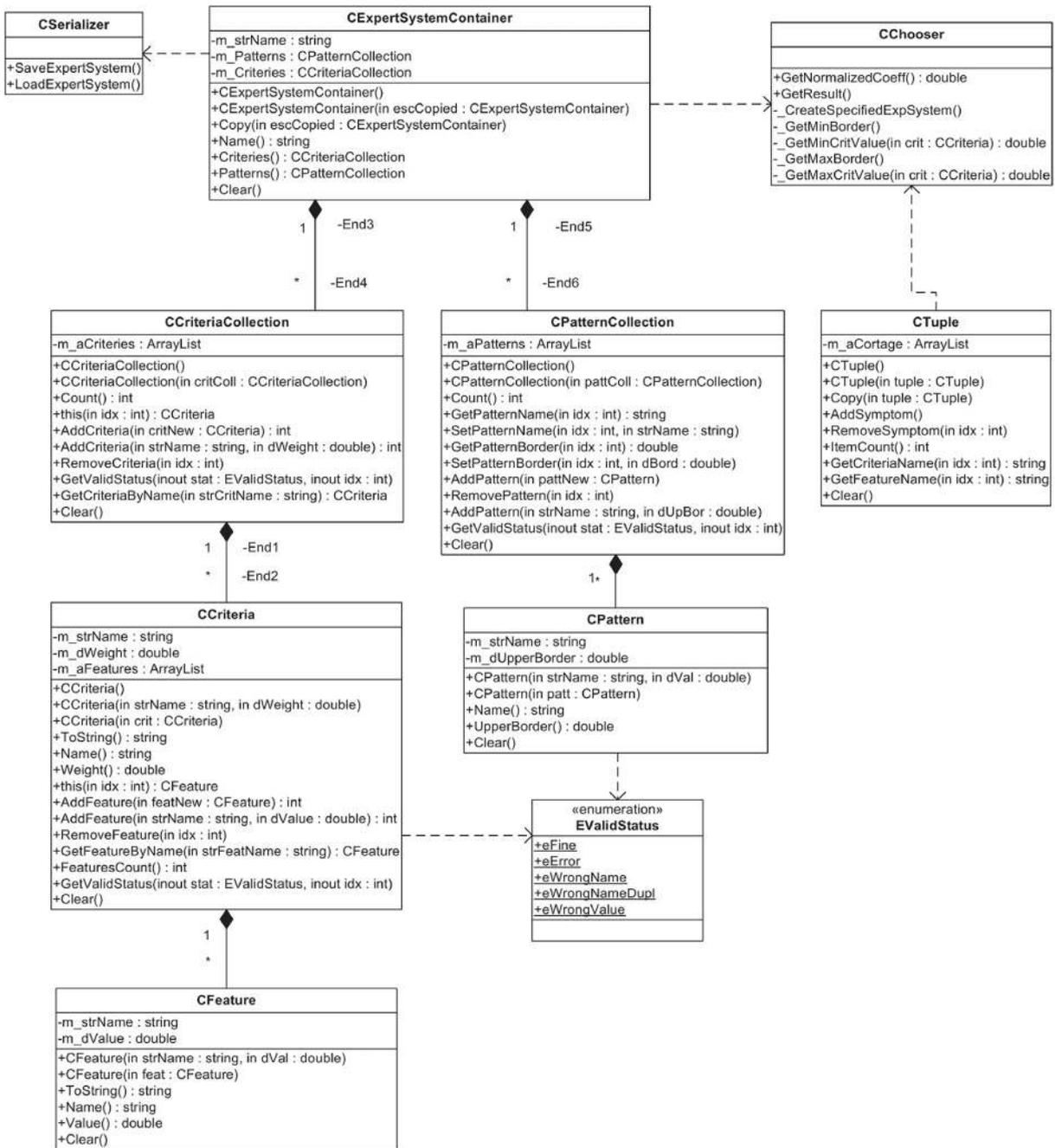


Fig. 5. Class diagram of expert system for the sapphire defects prediction.

- CCriteria—a class that represents a feature, which can be characterized by an object from the subject area. This class includes as attributes the name of the attribute, and the weight of the attribute, which is necessary for evaluating its effect on the resulting output of the calculation. This class also combines a set of possible values for the criterion of the values contained in the collection of objects of the CFeature class.
- CFeature—a class that represents one particular value from a certain set of permissible values related to a specific attribute. The class allows the user to set the name of the value and its numerical rating, defined in the range from 0 to 1.

- CCriteriaCollection—a class that facilitates work with a variety of features that are available in the developed expert system. This class simplifies the search and selection of criteria in the developed system, their addition and deletion, and also includes internal means of checking features on the correctness.
- CPattern—a class of the solution variant that is presented to the user of the expert system. The class includes a field for describing the solution received by the user after the search. Another attribute characterizing a solution is the range of values defined by the expert, which distinguishes the solution from others.
- CPatternCollection—a class that represents a set of solutions specified by an expert. This class includes solutions, sorted in an ascending order, which speeds up and simplifies the search procedure. This class also allows to test for the correctness of the set of solutions.
- CExpertSystemContainer—a class that encapsulates all static information about features and solutions provided by experts.
- CTuple—a class that contains information about the selection of certain values of attributes. This sample is made by the user of the expert system. The accuracy of the solution, as a rule, increases when more features are set by the user. The CChooser instance is passed as an object of this class to find a solution.
- CChooser—a class designed to search for a solution. In the beginning, a coefficient that is scaled to the interval of [0, 1] is determined from the sample transmitted to it. This coefficient is then passed to the CPatternCollection object. The CPatternCollection object determines which of the intervals falls on the number, and then returns the solution.
- CSerializer—a class that allows serializing and deserializing all the data stored in the CExpertSystemContainer. This class is necessary to save the developed expert system to a file on disk and to load it.

The algorithms for data collection and analysis are designed to meet the following criteria: Analysis of initial technological data; statistical data processing; modeling of the influence of technological parameters on the quality of crystals; crystal quality prediction

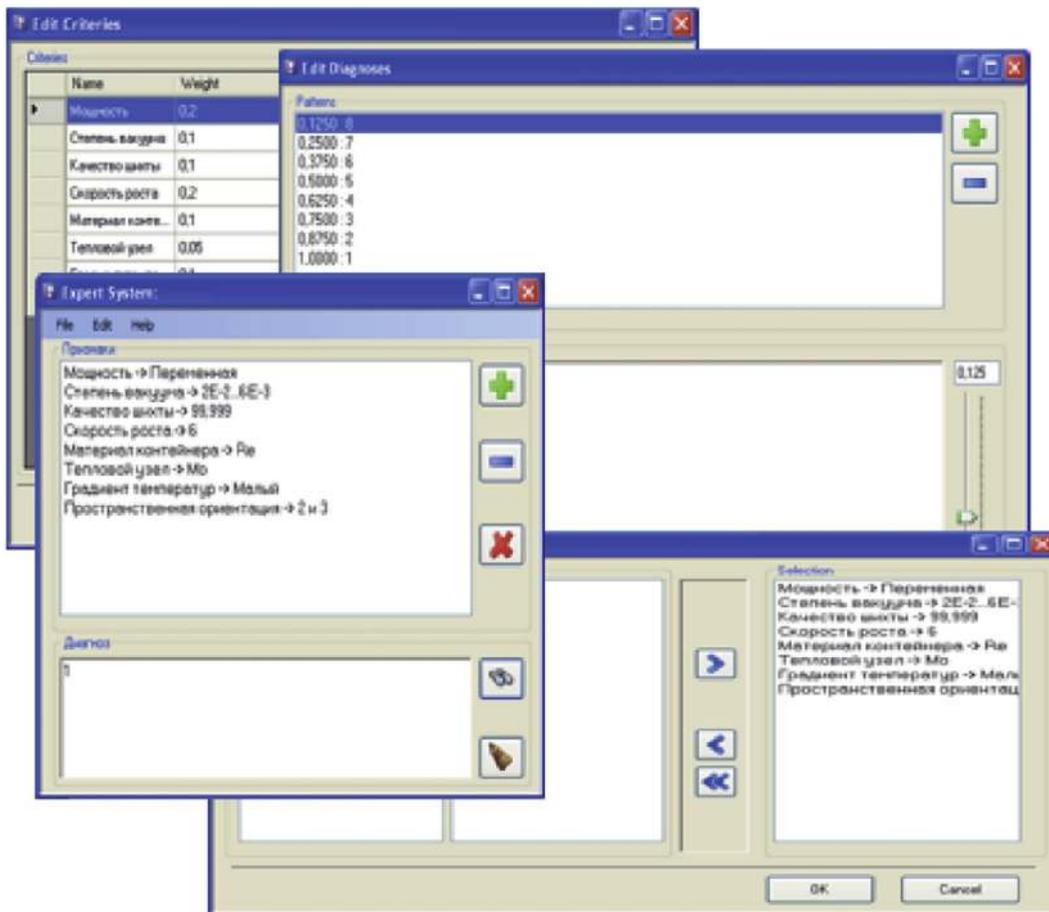


Fig. 6. Expert system software interface.

according to the initial data; decision making; analysis of reasons of possible deviations; model correction based on newly discovered data.

Selected user interface elements of the expert system are shown in Fig. 6.

4. Conclusions

Various machine learning methods are capable of finding optimal conditions for the sapphire monocrystal production with different efficiency.

This research has been conducted as a comparison between data mining tools (the logistic regression, SVM, classification trees, neural networks, naive Bayes algorithm, kNN algorithm, CN2 induction algorithm, AdaBoost, random forest algorithm, and SGD algorithm) for the sapphire crystals defects prediction. We analyzed the fundamental interplay between the availability of materials data and the predictive capability of machine learning models. We studied the various data mining techniques available to predict the crystal defects and to find the best methods for prediction.

The proposed models allow predicting the crystal quality. We obtained an automatic procedure and machine learning use for the fast crystals quality prediction. The industrial application of such methods will heighten the automatization level of production of crystals with a predefined combination of properties that can be important for a particular application in microelectronics and nanoelectronics. Solving these scientific and engineering problems assumes the use of information technologies in crystals production on a new level.

In order to make data mining technics applicable to daily practice, they have to be integrated into the decision support systems. For this, we propose the schema where the predictive modules are developed separately with the Orange Canvas data mining tool, while for the decision support system we suggest to develop the special software. On the base of our study, we propose the software for analyzing the resulting crystals quality, which allows optimizing the process of the crystal growth.

We expect to increase the experimental data in the future so it will give new opportunities for prediction and increasing its accuracy. We plan to recognize crystal images from the furnace chamber and to forecast the conditions' influence on the crystal quality.

Funding

This work was supported by the Russian Foundation for Basic Research Projects under Grant No. 16-52-48016 ИИД_ОМн (R. Kumar and A. V. Filimonov).

References

- [1] E.R. Dobrovinskaya, L.A. Lytvynov, V. Pishchik, *Sapphire: Material, Manufacturing, Applications*, Springer, New York, 2009, pp. 10–23, ch. 2.
- [2] D.I. Cherednichenko, S.P. Malyukov, Y.V. Klunnikova, Heat-physical processes at the sapphire crystals growth by horizontal directed crystallization, in: I. Tartaglia (Ed.), *Sapphire: Structure, Technology and Applications*, Nova Science Publishers, New York, 2013, pp. 101–118.
- [3] S.P. Malyukov, D.I. Cherednichenko, Y.V. Klunnikova, A.V. Sayenko, Investigation of influence of sapphire crystals shape factor on its growth by horizontal directed crystallization, *IOP Conf. Ser. Mater. Sci. Eng.* 475 (1) (2019), pp. 012009:1-4.
- [4] Y.V. Klunnikova, M.V. Anikeev, U. Nackenhorst, A.V. Filimonov, Computational approach to the simulation of sapphire crystals growth by horizontal directed crystallization method, *J. Phys. Conf.* 1236 (1) (Jun. 2019), pp. 012030:1-6.
- [5] Y. Klunnikova, S. Malyukov, A. Sayenko, A.V. Filimonov, An analysis approach to the sapphire crystals growth by horizontal directed crystallization method, *Key Eng. Mater.* 806 (Jun. 2019) 203–208.
- [6] C.-H. Xu, S.-H. Meng, M.-F. Zhang, H.-B. Zuo, G.-G. Wang, Thermal stresses and cracks during the growth of large-sized sapphire with SAPMAC method, *Chin. J. Aeronaut.* 20 (5) (Oct. 2007) 475–480.
- [7] S.A. Udovenko, D.Y. Chernyshov, D.A. Andronikova, A.V. Filimonov, S.B. Vakhrushev, The technique of studying X-ray scattering over wide temperature range in an electric field, *Phys. Solid State* 60 (5) (May 2018) 963–966.
- [8] S.P. Malyukov, Y.V. Klunnikova, M.V. Anikeev, Research of gas bubbles interaction with crystallization front of sapphire melt, in: I.A. Parinov, S.-H. Chang, V.Y. Topolov (Eds.), *Advanced Materials*, Springer, Cham, 2016, pp. 101–113, ch. 8.
- [9] M. Knjazeva, Y. Bronwald, D. Andronikova, et al., Modulated structures in PbHfO₃ crystals at high-pressure-high-temperature conditions, *Defect Diffusion Forum* 386 (Sept. 2018) 149–155.
- [10] S.P. Malyukov, Y.V. Klunnikova, Complex investigations of sapphire crystals production, in: S.-H. Chang, I.A. Parinov, V.Y. Topolov (Eds.), *Advanced Materials*, Springer, Cham, 2014, pp. 55–69.
- [11] T. Xie, J.C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.* 120 (14) (Apr. 2018) 145301, <https://doi.org/10.1103/PhysRevLett.120.145301>.
- [12] M. Azlan Hussain, Review of the applications of neural networks in chemical process control—simulation and online implementation, *Artif. Intell. Eng.* 13 (1) (Jun. 1999) 55–68.
- [13] J.-Z. Chu, P.-F. Tsai, W.-Y. Tsai, et al., An experimental study of model predictive control based on artificial neural networks, in: V. Palade, R.J. Howlett, L. Jain (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems*, Springer, Berlin, 2003, pp. 1296–1302.
- [14] B. Kim, S. Kim, K. Kim, Modelling of plasma etching using a generalized regression neural network, *Vacuum* 71 (4) (Jul. 2003) 497–503.
- [15] D. Rodriguez, R. Ruiz, J.C. Riquelme, R. Harrison, A study of subgroup discovery approaches for defect prediction, *Inf. Software Technol.* 55 (10) (Oct. 2013) 1810–1822.
- [16] K. Rangra, K.L. Bansal, Comparative study of data mining tools, *Int. J. Adv. Res. Comput. Sci. Software Eng.* 4 (6) (Jun. 2014) 216–223.
- [17] A.H. Wahbeh, Q.A. Al-Radaideh, M.N. Al-Kabi, E.M. Al-Shawakfa, A comparison study between data mining tools over some classification methods, *Int. J. Adv. Comput. Sci. Appl.* 8 (2) (2011) 18–26.
- [18] J. Demšar, B. Zupan, Orange: data mining fruitful and fun—a historical perspective, *Informatika* 37 (1) (2013) 55–60.
- [19] S. Neelamegam, E. Ramaraj, Classification algorithm in data mining: an overview, *Int. J. P2P Netw. Trends Technol.* 3 (5) (Oct. 2013) 1–5.
- [20] G. Amala, Orange tool approach for comparative analysis of supervised learning algorithm in classification mining, *J. Anal. Comput.* 13 (1) (Jan. 2019) 1–10.
- [21] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli, Insightful classification of crystal structures using deep learning, *Nat. Commun.* 9 (Jul. 2018) 2775.
- [22] R. Ramprasad, R. Batra, G. Piliand, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, *NPJ Comput. Mater.* 3 (Dec. 2017) 54.

- [23] A.P. Barók, J. Kermode, N. Bernstein, G. Csányi, Machine learning a general-purpose interatomic potential for silicon, *Phys. Rev. X* 8 (Dec. 2018), pp. 041048:1-32.
- [24] Y. Zhang, C. Ling, A strategy to apply machine learning to small datasets in materials science, *NPJ Comput. Mater.* 4 (May 2018) 25.
- [25] P. Subathra, R. Deepika, K. Yamini, P. Arunprasad, S.K. Vasudevan, A study of open source data mining tools and its applications, *Res. J. Appl. Sci. Eng. Technol.* 10 (10) (Jan. 2015) 1108–1132.



Yulia Vladimirovna Klunnikova was born in Taganrog in 1986. She received the B.S. and M.S. degrees from the Southern Federal University, Taganrog in 2007 and 2009, respectively, both in computer science. She received the Ph.D. degree from the Southern Federal University in 2011. She is currently working as an associate professor with the Institute of Nanotechnology, Electronics and Engineering Equipment, Southern Federal University. Her research interests include crystals growth, treatment, and applications; crystals growth simulation; software for crystal growth.



Maxim Vladimirovich Anikeev was born in Taganrog in 1977. He received the B.S. and M.S. degrees from the Southern Federal University in 1998 and 2000, respectively, both in computer science. He received the Ph.D. degree from the Southern Federal University in 2008. He is currently working as an associate professor with the Institute of Computer Technology and Information Security, Southern Federal University. His research interests include machine learning, computer security, and parallel programming.



Alexey Vladimirovich Filimonov was born in St. Petersburg in 1967. He received his Dr. of Physics and Mathematics from the Peter the Great St. Petersburg Polytechnic University, St. Petersburg in 2013. He is currently working as the Head of Department of Physical Electronics, Peter the Great St. Petersburg Polytechnic University. His research interests include surface physics, spectroscopy and nanostructure physics, X-ray and synchrotron radiation, and physics of nanocomposite materials.



Ravi Kumar was born in Mysore in 1975. He received his Dr. rer. nat. in materials science from the Max Planck Institute for Metals Research (currently, renamed as Max Planck Institute for Intelligent Systems), Stuttgart in 2005. He heads the Laboratory for High Performance Ceramics, Department of Metallurgical and Materials Engineering, Indian Institute of Technology Madras (IIT Madras), Chennai. His research interests include processing and characterization of precursor derived ceramics for structural and functional applications.