**DOI**: 10.63527/1607-8829-2023-1-44-54

*M.M. Korop*

**M. M. Korop**

Institute of Thermoelectricity of the NAS and MES of Ukraine,
1, Nauky str., Chernivtsi, 58029, Ukraine,
*e-mail: mykola.korop@chnu.edu.ua*

# MACHINE LEARNING IN THERMOELECTRIC MATERIALS SCIENCE

*The paper presents machine learning methods and their application in thermoelectric materials science. The results of their application, strong points and application areas are shown. The difficulties that arise in the process of predicting the properties of thermoelectric materials and ways to overcome them were taken into account. Bibl. 30, Fig. 1, Tabl. 2.*

**Key words:** machine learning methods, thermoelectric materials science.

## Introduction

*General characterization of the problem.* Machine learning plays an increasingly important role in the intensification of scientific research and discoveries. New approaches and methods open up the possibility of speeding up the discovery of new promising materials [1 – 3], optimization of device manufacturing technologies [4, 5], calculation of the most economically advantageous solutions under given conditions [6].

As noted in article [7], the property of thermoelectricity is its use in areas that shape scientific and technological progress, therefore it is developing primarily in the leading countries of the world. Thermoelectric materials have great potential; today they serve as the basis for generating, cooling and sensor devices and sensors, which have found application in all spheres of life from medicine to space. However, the selection of the optimal thermoelectric material for given conditions, its search or optimization is not a trivial task, requiring significant time and material costs.

There are a number of studies on the use of machine learning in thermoelectric materials science; they show high efficiency and low cost of use compared to traditional approaches.

*The purpose of the work* is to consider machine learning methods and highlight the results of their application in thermoelectric materials science.

## Supervised machine learning methods in thermoelectric materials science

Machine learning is a field of scientific research that studies the ability of a computer to learn without being pre-programmed [7]. Machine learning algorithms can be divided into two groups: supervised and unsupervised. Each of these groups has its own scope and the algorithm that can give the most accurate result under the given conditions is chosen.

Supervised machine learning is the construction of algorithms capable of finding general patterns and hypotheses using externally obtained training and test data sets. Such an algorithm studies certain patterns in the training data set for further verification on test data sets [8].

Unsupervised machine learning – recognition of patterns and sets of similarities without the involvement of target attributes, all variables that are present in the training data set participate in the analysis, owing to which such algorithms become useful for clustering and associative segmentation [9].

Machine learning models rely on three key components: training data, descriptors, and an algorithm. Training data are sets of examples from which the algorithm tries to obtain chemical trends, descriptors are low-level characteristics of materials (crystal structure, chemical formula, average atomic number, etc.), which allows to "vectorize" a chemical material and make it suitable for further machine learning. learning algorithm – polls training vector data in combination with a certain template [10].

Kamal Choudhary et al. [11] in their work on the search for highly efficient 3D and 2D dimensional thermoelectric materials use supervised learning algorithms, namely: decision trees, random forest of decision trees, k nearest neighbors, multilayer perceptron and gradient amplification techniques.

A decision tree is a popular machine learning algorithm when the task is to classify a data set or perform regression, this solution effectively copes with missing values and possible errors in the data set [12]. A graphical representation can be represented as a tree, where the nodes represent the validation of conditions for the data and the branches represent the possible outcomes of those conditions. Data for use in this method are partitioned into subsets based on entropy, Gini coefficient, etc. In the article by Alrebdi et al. [13], this method is used to predict the thermal conductivity of thermoelectric materials based on $Bi_2Te_3$. Decision trees are often used as the basis for ensemble methods, such as a random forest of decision trees.

The main idea of the random forest ensemble method is to create subsets of training data and train individual decision trees for further averaging of predictions (regression task) or combination (classification task). This approach makes it possible to obtain much higher accuracy of forecasts. The advantages of this algorithm include the evaluation of data structure and dependencies for data analysis, and the disadvantages include a relatively slow learning speed due to the need to create and train a large number of individual trees. In the work of Chen et al. [14] the random forest method is used to discover new M2X3 thermoelectric materials with only composition information.

Gradient amplification is used for the composition of weak models (decision trees), gradient descent is used, which allows minimizing the loss functional. Increasing the accuracy of predictions is achieved by adding new models that correct the errors of previous models. This method is widely used and is included in well-known libraries such as XGBoost, LightGBM, CAT Boost. Sheng et al. [15] used this algorithm to predict the power factor in diamond-like thermoelectric materials.

*K*-nearest neighbors is one of the simplest supervised machine learning algorithms, which is based on the classification of a new object given the classes of *K* nearest neighbors. For example, the Euclidean distance can be used to determine the nearest neighbor. This can lead to slow learning of the algorithm on large data sets due to the need to calculate the distance between all pairs of objects. Gyoung et al. [16] use *K*-nearest-neighbor regression to predict the target value for the input data by interpolating the *K*-nearest-neighbor target values in the training data.

The multilayer perceptron belongs to more advanced machine learning algorithms. It is a neural network consisting of three or more layers, each of which contains several neurons. The algorithm uses backpropagation of error for training, which allows to optimize the weights of the neural network using gradient descent. Thus, each neuron, receiving input data, calculates their internal sum and applies some activation function: sigmoid, hyperbolic tangent, etc., adding nonlinearity to the value of the neuron. Uysal et al. [17] use a multilayer perceptron algorithm to estimate the Seebeck coefficient for a p-type

high-temperature thermoelectric material. This algorithm can serve as a basis for more complex neural networks of the convolutional neural network type.

## Unsupervised machine learning methods in thermoelectric materials science

Unsupervised machine learning algorithms are effective when working with datasets that do not contain labels, allowing for hidden structures to be found. Thus, there is an opportunity to carry out clustering to group materials based on their chemical, mechanical or physical properties, identifying their new classes or properties that are common to a certain group. They are also used in the search for anomalies and defects due to the search for deviations from standard values. One of the main applications is the automatic search for combinations of elements or parameters to obtain optimal materials under given conditions or to carry out their optimization. Jia et al. [18] use *K*-means, Gaussian mixture, DBSCAN, AGNES, Birch to search for promising semi-Heusler thermoelectric materials, and Iwasaki et al. [19] use the LASSO algorithm (Method of Minimization of Absolute Compression and Operator Selection) to identify modern materials based on spin-thermoelectric material.

K-means works by dividing the data set into clusters (groups) and aims to minimize variability within clusters and maximize between them. *K* determines the number of clusters specified by the user. The algorithm uses K arbitrary centroids and places the data record in the closest one. After completing the placement of objects, the centroids calculate their average value in the cluster, this process happens iteratively until the centroids stabilize and equilibrium occurs. Sheng et al. [20] use this method to accelerate the discovery of *Cu-Sn-S* thermoelectric compounds using high-throughput synthesis.

The combination of Gaussian divisions is a statistical model and represents a Gaussian sum, where the skin component contains the mean, dispersion and value. This algorithm is often used to generate elliptic shapes in data clusters. To set the parameters, the method of maximizing likelihood through the EM algorithm (Expectation-Maximization) is used. Shimizu et al. [21] use this method for gate analysis of several target parameters in materials design.

DBSCAN, AGNES and Birch are clustering algorithms used to measure the distance or similarity between data points, their main goal is to detect natural groups in data without the need to explicitly specify the number of clusters.

DBSCAN (Density-Based Spatial Clustering of Applications) works by identifying clusters based on large density variance.

AGNES (Agglomerative Nesting) – gradual unification of objects into hierarchical clusters and visualization using a dendrogram.

Birch – the algorithm is optimized for clustering large data sets, allowing them to be processed quickly.

These methods are often employed to discover new groups of promising materials and are used in a number of works [18, 22].

The LASSO regression method works by $L_1$ regularization, reducing the coefficients of some variables to zero, which facilitates feature selection and helps to highlight the most important features from a large set of features. One drawback follows from this - with signs that are strongly correlated, only one can be taken into account.

When searching for and optimizing a thermoelectric material, the efficiency of the material can be affected by electrical conductivity, thermal conductivity, the Seebeck coefficient, and other parameters. Using this method, you can determine the set that will give the most accurate result. Also,

it provides an opportunity to understand how changing conditions (temperature, pressure, etc.) will affect their properties and to understand the mechanisms of this influence. Wudil et al. [23] use this method to evaluate the performance of materials based on $Bi_2Te_3$.

## Results of the application of machine learning in thermoelectric materials science

In the paper by Gaultois et al. [10] published in 2015, the authors were among the first to develop an electronic system for evaluating the parameters of thermoelectric alloys in real time based on machine learning. In the illustration presented by them it is clear that the best studied thermoelectric materials lie in a certain region of the periodic table (black and blue dots), chalcogenides and p-elements, while orange dots highlight new material alloys proposed by the system based on machine learning and these materials are placed outside the known materials (pure intermetallics).

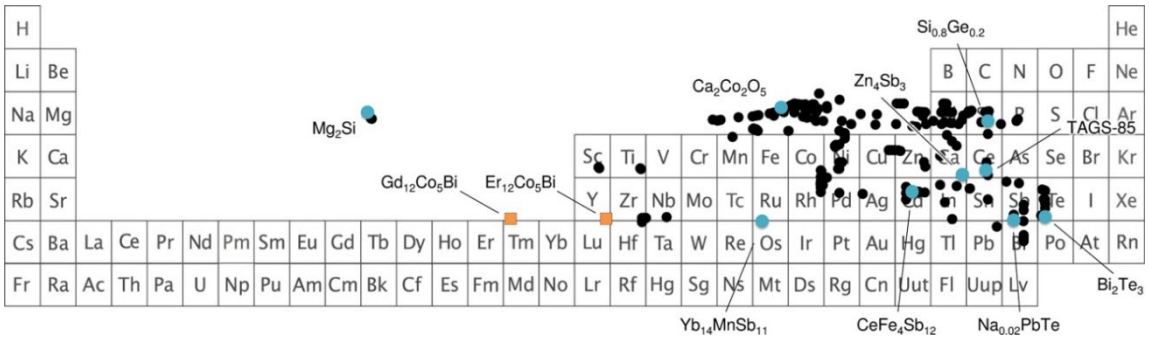


*Fig. 1. The periodic table of elements based on the composition-weighted average of the positions of the elements in the material [10]*

The authors of this paper have developed a machine model that provides recommendations on the suitability of new thermoelectric alloys for further research, but such a model does not return quantitative predictions of thermoelectric properties or accurate identification of thermoelectric alloys. Such a machine learning-based recommendation system looks for empirical, chemical patterns in experimental data on known thermoelectric materials in order to make statistical predictions about the performance of new materials. When checking the obtained results, the system provided results with high accuracy, where the error distribution for various material properties approaches 0. As a result, several interesting thermoelectric alloys were found selected from the list proposed by the algorithm and experimental confirmation of thermoelectric properties [23 – 25] (Table 1).

*Table 1*

*Several promising new thermoelectric compounds selected from the calculated list. P values refer to the model's level of confidence that a given material will exhibit room temperature values of a certain property (for instance, S or ρ) within target values [10]*

| Material | $P_S$ | $P_\rho$ | $P_\kappa$ | $P_{gap}$ | Comments |
|---|---|---|---|---|---|
| $TaPO_5$ and $TaVO_5$ | 0.894 | 0.793 | 0.958 | 0.987 | High polyhedral connectivity and structural superlattices |

*Continuation of table 1*

| | | | | | |
|---|---|---|---|---|---|
| $Tl_9SbTe_6$ | 0.845 | 0.871 | 0.999 | 0.876 | Recently reported to be a good thermoelectric material ($zT \approx 1$ at 600 K) |
| $TaAlO_4$ | 0.893 | 0.703 | 1 | 0.977 | High mass contrast, high polyhedral connectivity ($TaO_6$ octahedra dividing edges and vertices) |
| $SrCrO_3$ | 0.772 | 0.767 | 0.996 | 0.95 | High polyhedral connectivity (3-*D* angular connection of $CrO_6$ octahedra), metallic, when created under high pressure. |
| $TaSbO_4$ | 0.892 | 0.919 | 1 | 0.997 | High polyhedral connectivity: layered, edge-sharing $MO_6$ octahedra |
| *TiCoSb* | 0.981 | 0.714 | 0.958 | 0.833 | *TiCoSb* is not a new compound, but has been studied as a high *zT* material. However, it was not included in the training data. |

Thanks to machine learning, the authors managed to be the first to propose an experimentally viable new compound from a real white chemical space, where no previous characteristics indicated promising chemical processes [10].

Another area where machine learning can find promising application is finding the dependences and parameters of the spin-controlled thermoelectric effect (STE) [19, 26]. Such devices could provide a universal thermoelectric technology with scalable production, but this is hindered by a lack of understanding of the fundamental physics and properties of the materials responsible for the effect. The article by Iwasaki et al. [19, 26] claims the synthesis of a material that helped in the identification of a new STE material with a thermal EMF that is an order of magnitude higher than that of the current generation of devices.

Recent works describe the emergence of new, more accurate models, which, in addition to predicting probably new interesting thermoelectric materials for research, also learn to predict the properties of materials with high accuracy, the coefficient of determination $R_2$ = 0.91-0.959 for well-studied materials [27 – 29].

One of the challenges facing improving the accuracy of such machine learning models is the creation of complete and high-quality databases of material properties, as well as access to them [30]. Despite the constant growth of interest in the application of machine learning algorithms for the search and discovery of new materials, today we have very modest tools compared to other areas of its application, namely image processing and industrial production. The degree of freedom (*DoF*) is the number of model variable parameters that are statistically significant. The degree of freedom directly affects the desired size of the required training data sets, so it is common to use models with a limited number of variable parameters and an approximate estimate of properties in the search space.

In the article by Gyoung S. Na et al. [31] a publicly open database of educational data was formed, obtained through a literature search containing the chemical structure and experimentally measured thermoelectric properties of materials. These data are presented in table form (Table 2) and total 5205 experimental values.

*Table 2*

*Description of the collected characteristics in the ESTM data set [31]. The first value is the chemical composition of the collected thermoelectric materials, which must be translated into a digital representation when applied in machine learning models, the second value is the temperature. Other parameters were obtained either experimentally or theoretically calculated*

| Property name | Units of measurement | Range | Average value |
|---|---|---|---|
| Chemical composition | - | - | - |
| Temperature | K | [10, 1275] | 539.28 ± 192.42 |
| Seebeck coefficient | μV/K | [−1174, 1052.4] | 73.18 ± 208.92 |
| Electrical conductivity | S/m | (0.9.47E + 07) | 1.10E + 05 ± 1.47E + 06 |
| Thermal conductivity | W / m * K) | [0.07, 77.16] | 2.25 ± 3.29 |
| Power factor | W / (m * $K^2$) | (0. 7.61E − 03) | 9.92E − 04 ± 1.12E − 03 |
| *ZT* | - | (0. 2.28) | 0.35 ± 0.35 |

As a result, based on the collected database, the authors of [31] developed a machine model that achieved an accuracy of $R_2$ above 0.9 in predicting 5 thermoelectric properties of materials and showed an average absolute error of less than 0.06 when predicting *ZT*. In addition to the publicly available dataset, the authors have developed a method for representing alloys and alloyed materials called System Identifiable Material Description (SIMD). Based on transfer learning using SIMD, it was possible to improve the $R_2$ from 0.13 to 0.71 when extrapolating to predict the *ZT* of materials from unexplored groups to find new high-performance materials.

## Conclusions

1. The study of supervised and unsupervised machine learning methods used in thermoelectric materials science to optimize existing and search for new promising thermoelectric materials is considered.
2. Thanks to new opportunities, it becomes possible to further develop those areas of science that do not contain a perfect fundamental understanding, for example, the spin-driven thermoelectric effect, a deep understanding and determination of the interdependence between parameters that affect the efficiency of a thermoelectric material.
3. The accuracy of the latest models is high. However, their further improvement is limited by the space of well-studied materials and existing databases of experimentally collected thermoelectric properties of materials or theoretically calculated using alternative methods.

*The author expresses his sincere gratitude to his supervisor, Lukyan Ivanovich Anatychuk, for the proposed topic of the work and useful remarks.*

## References

1. Liu Y., Zhao T., Ju W., & Shi S. (2017). Materials discovery and design using machine learning. *Journal of Materiomics*, 3 (3), 159 - 177. Elsevier BV. https://doi.org/10.1016/j.jmat.2017.08.002

2. Juan Y., Dai Y., Yang Y., & Zhang J. (2021). Accelerating materials discovery using machine learning. *Journal of Materials Science & Technology*, 79, 178 - 190. Elsevier BV. https://doi.org/10.1016/j.jmst.2020.12.010
3. von Lilienfeld O. A. & Burke K. (2020). Retrospective on a decade of machine learning for chemical discovery. *Nature Communications*, 11 (1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41467-020-18556-9
4. Kim E., Huang K., Saunders A., McCallum A., Ceder G. & Olivett, E. (2017). Materials synthesis insights from scientific literature via text extraction and machine learning. *Chemistry of Materials*, 29(21), 9436 - 9444). American Chemical Society (ACS). https://doi.org/10.1021/acs.chemmater.7b03500
5. Cao B., Adutwum L. A., Oliynyk A. O., Luber E. J., Olsen B. C., Mar A. & Buriak J. M. (2018). How to optimize materials and devices via design of experiments and machine learning: demonstration using organic photovoltaics. *ACS Nano*, 12 (8), 7434 - 7444. American Chemical Society (ACS). https://doi.org/10.1021/acsnano.8b04726
6. Queen H. J., J. J., D. T. J., K. V. S. Moses Babu and Thota S.P. (1021), Machine learning-based predictive techno-economic analysis of power system, *IEEE Access*, 9, 123504-123516, 2021, doi: 10.1109/ACCESS.2021.3110774.
7. Mahesh B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research* (*IJSR*)*,* 9(1), 381 - 386.
8. Singh A., Thakur N. and Sharma A. A review of supervised machine learning algorithms (2016). *3rd International Conference on Computing for Sustainable Global Development* (INDIACom), New Delhi, India, 2016, pp. 1310-1315.
9. Alloghani M., Al-Jumeily D., Mustafina J., Hussain A. & Aljaaf A. J. (2019). A systematic review on supervised and unsupervised machine learning algorithms for data science. *Unsupervised and Semi-Supervised Learning* (pp. 3 – 21). Springer International Publishing. https://doi.org/10.1007/978-3-030-22475-2_1
10. Gaultois M. W., Oliynyk A. O., Mar A., Sparks T. D., Mulholland G. J. & Meredig, B. (2016). Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Materials*, 4(5). AIP Publishing. https://doi.org/10.1063/1.4952607
11. Choudhary K., Garrity K. F., & Tavazza F. (2020). Data-driven discovery of 3D and 2D thermoelectric materials. *Journal of Physics: Condensed Matter*, 32 (47), 475501). IOP Publishing. https://doi.org/10.1088/1361-648x/aba06b
12. Somvanshi M., Chavan P., Tambade S. & Shinde S. V. (2016). A review of machine learning techniques using decision tree and support vector machine. *2016 International Conference on Computing, Communication, Control and Automation* (*ICCUBEA*). IEEE. https://doi.org/10.1109/iccubea.2016.7860040
13. Alrebdi T. A., Wudil Y. S., Ahmad U. F., Yakasai F. A., Mohammed J. & Kallas F. H. (2022). Predicting the thermal conductivity of Bi2Te3-based thermoelectric energy materials: A machine learning approach. *International Journal of Thermal Sciences,* 181, 107784. Elsevier BV. https://doi.org/10.1016/j.ijthermalsci.2022.107784
14. Chen D., Jiang F., Fang L., Zhu Y.-B., Ye C.-C. & Liu W.-S. (2022). Machine learning assisted discovering of new M2X3-type thermoelectric materials. *Rare Metals*, 41(5), 1543 – 1553. Springer Science and Business Media LLC. https://doi.org/10.1007/s12598-021-01911-0.

15. Sheng Y., Wu Y., Yang J., Lu W., Villars P. & Zhang W. (2020). Active learning for the power factor prediction in diamond-like thermoelectric materials. *Computational Materials*, 6(1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41524-020-00439-8.
16. Na G. S. & Chang H. (2022). A public database of thermoelectric materials and system-identified material representation for data-driven discovery. *Computational Materials,* 8 (1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41524-022-00897-2.
17. Uysal F., Kilinc E., Kurt H., Celik E., Dugenci M. & Sagiroglu S. (2017). Estimating Seebeck coefficient of a p-type high temperature thermoelectric material using bee algorithm multi-layer perception. *Journal of Electronic Materials,* 46(8), 4931 – 4938). Springer Science and Business Media LLC. https://doi.org/10.1007/s11664-017-5497-6
18. Jia X., Deng Y., Bao, X, Yao H., Li S., Li Z., Chen C., Wang X., Mao J., Cao F., Sui J., Wu J., Wang C., Zhang Q. & Liu X. (2022). Unsupervised machine learning for discovery of promising half-Heusler thermoelectric materials. *Computational Materials*, 8 (1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41524-022-00723-9
19. Iwasaki Y., Sawada R., Stanev V., Ishida M., Kirihara A., Omori Y., Someya H., Takeuchi I., Saitoh E., & Yorozu S. (2019). Identification of advanced spin-driven thermoelectric materials via interpretable machine learning. *Computational Materials*, 5 (1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41524-019-0241-9
20. Sheng Y., Deng T., Qiu P., Shi X., Xi J., Han Y. & Yang J. (2021). Accelerating the discovery of *Cu–Sn–S* thermoelectric compounds via high-throughput synthesis, characterization, and machine learning-assisted image analysis. *Chemistry of Materials*, 33(17), 6918 - 6924). American Chemical Society (ACS). https://doi.org/10.1021/acs.chemmater.1c01856
21. Shimizu N. & Kaneko H. (2020). Direct inverse analysis based on Gaussian mixture regression for multiple objective variables in material design. *Materials & Design*, 196, 109168. Elsevier BV. https://doi.org/10.1016/j.matdes.2020.109168
22. Antunes L. M., Vikram Plata J. J., Powell A. V., Butler K. T. & Grau-Crespo R. (2022). Machine learning approaches for accelerating the discovery of thermoelectric materials. *ACS Symposium Series* (pp. 1 – 32). American Chemical Society. https://doi.org/10.1021/bk-2022-1416.ch001
23. Guo Q., Chan M., Kuropatwa B. A., & Kleinke H. (2013). Enhanced thermoelectric properties of variants of $Tl_9SbTe_6$ and $Tl_9BiTe_6$. *Chemistry of Materials,* 25(20), 4097 - 4104). American Chemical Society (ACS). https://doi.org/10.1021/cm402593f
24. Parker D., & Singh D. J. (2013). Alkaline earth lead and tin compounds $Ae_2Pb$, $Ae_2Sn$, $Ae = Ca$, $Sr$, $Ba$, as thermoelectric materials. *Science and Technology of Advanced Materials,* 14(5), 055003. Informa UK Limited. https://doi.org/10.1088/1468-6996/14/5/055003
25. Sun P., Oeschler N., Johnsen S., Iversen B. B. and Steglich F. (2009). Huge thermoelectric power factor: $FeSb_2$ versus $FeAs_2$ and $RuSb_2$, *Appl. Phys. Express* 2, 091102.
26. Iwasaki Y., Takeuchi I., Stanev V., Kusne A. G., Ishida M., Kirihara A., Ihara K., Sawada R., Terashima K., Someya H., Uchida K., Saitoh E, & Yorozu S. (2019). Machine-learning guided discovery of a new thermoelectric material. *Scientific Reports*, 9(1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41598-019-39278-z
27. Jaafreh R., Yoo Seong, K., Kim J.-G. & Hamad K. (2022). A deep learning perspective into the figure-of-merit of thermoelectric materials. *Materials Letters,* 319, 132299. Elsevier BV. https://doi.org/10.1016/j.matlet.2022.132299
28. Li Y., Zhang J., Zhang K., Zhao M., Hu K. & Lin, X. (2022). Large data set-driven machine learning models for accurate prediction of the thermoelectric figure of merit. *ACS Applied*

*Materials & Interfaces*, 14(50), 55517 - 55527). American Chemical Society (ACS). https://doi.org/10.1021/acsami.2c15396

29. Wang T., Zhang C., Snoussi H. & Zhang G. (2019). Machine learning approaches for thermoelectric materials research. Advanced Functional Materials, 1906041. https://doi.org/10.1002/adfm.201906041
30. Na G.S., Chang H. (2022). A public database of thermoelectric materials and system-identified material representation for data-driven discovery. *Comput Mater* 8 (1), 214. Springer Science and Business Media LLC https://doi.org/10.1038/s41524-022-00897-2.

Submitted: 10.01.2023

**Короп М. М.**

Інститут термоелектрики НАН та МОН України,
вул. Науки, 1, Чернівці, 58029, Україна,
*e-mail: mykola.korop@chnu.edu.ua*

## МАШИННЕ НАВЧАННЯ В ТЕРМОЕЛЕКТРИЧНОМУ МАТЕРІАЛОЗНАВСТВІ

*У роботі наводяться методи машинного навчання та їхнє застосування в термоелектричному матеріалознавстві. Показано результати їхнього застосування, сильні сторони та області застосування. Було взято до уваги складнощі, які виникають у процесі прогнозування властивостей термоелектричних матеріалів та способи їх подолання. Бібл. 30, рис. 1, табл. 2.*

**Ключові слова:** методи машинного навчання, термоелектричне матеріалознавство.

## References

1. Liu Y., Zhao T., Ju W., & Shi S. (2017). Materials discovery and design using machine learning. *Journal of Materiomics*, 3 (3), 159 - 177. Elsevier BV. https://doi.org/10.1016/j.jmat.2017.08.002
2. Juan Y., Dai Y., Yang Y., & Zhang J. (2021). Accelerating materials discovery using machine learning. *Journal of Materials Science & Technology*, 79, 178 - 190. Elsevier BV. https://doi.org/10.1016/j.jmst.2020.12.010
3. von Lilienfeld O. A. & Burke K. (2020). Retrospective on a decade of machine learning for chemical discovery. *Nature Communications*, 11 (1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41467-020-18556-9
4. Kim E., Huang K., Saunders A., McCallum A., Ceder G. & Olivett, E. (2017). Materials synthesis insights from scientific literature via text extraction and machine learning. *Chemistry of Materials*, 29(21), 9436 - 9444). American Chemical Society (ACS). https://doi.org/10.1021/acs.chemmater.7b03500

5. Cao B., Adutwum L. A., Oliynyk A. O., Luber E. J., Olsen B. C., Mar A. & Buriak J. M. (2018). How to optimize materials and devices via design of experiments and machine learning: demonstration using organic photovoltaics. *ACS Nano*, 12 (8), 7434 - 7444. American Chemical Society (ACS). https://doi.org/10.1021/acsnano.8b04726
6. Queen H. J., J. J., D. T. J., K. V. S. Moses Babu and Thota S.P. (1021), Machine learning-based predictive techno-economic analysis of power system, *IEEE Access*, 9, 123504-123516, 2021, doi: 10.1109/ACCESS.2021.3110774.
7. Mahesh B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research* (*IJSR*)*,* 9(1), 381 - 386.
8. Singh A., Thakur N. and Sharma A. A review of supervised machine learning algorithms (2016). *3rd International Conference on Computing for Sustainable Global Development* (INDIACom), New Delhi, India, 2016, pp. 1310-1315.
9. Alloghani M., Al-Jumeily D., Mustafina J., Hussain A. & Aljaaf A. J. (2019). A systematic review on supervised and unsupervised machine learning algorithms for data science. *Unsupervised and Semi-Supervised Learning* (pp. 3 – 21). Springer International Publishing. https://doi.org/10.1007/978-3-030-22475-2_1
10. Gaultois M. W., Oliynyk A. O., Mar A., Sparks T. D., Mulholland G. J. & Meredig, B. (2016). Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Materials*, 4(5). AIP Publishing. https://doi.org/10.1063/1.4952607
11. Choudhary K., Garrity K. F., & Tavazza F. (2020). Data-driven discovery of 3D and 2D thermoelectric materials. *Journal of Physics: Condensed Matter*, 32 (47), 475501). IOP Publishing. https://doi.org/10.1088/1361-648x/aba06b
12. Somvanshi M., Chavan P., Tambade S. & Shinde S. V. (2016). A review of machine learning techniques using decision tree and support vector machine. *2016 International Conference on Computing, Communication, Control and Automation* (*ICCUBEA*). IEEE. https://doi.org/10.1109/iccubea.2016.7860040
13. Alrebdi T. A., Wudil Y. S., Ahmad U. F., Yakasai F. A., Mohammed J. & Kallas F. H. (2022). Predicting the thermal conductivity of Bi2Te3-based thermoelectric energy materials: A machine learning approach. *International Journal of Thermal Sciences,* 181, 107784. Elsevier BV. https://doi.org/10.1016/j.ijthermalsci.2022.107784
14. Chen D., Jiang F., Fang L., Zhu Y.-B., Ye C.-C. & Liu W.-S. (2022). Machine learning assisted discovering of new M2X3-type thermoelectric materials. *Rare Metals*, 41(5), 1543 – 1553. Springer Science and Business Media LLC. https://doi.org/10.1007/s12598-021-01911-0.
15. Sheng Y., Wu Y., Yang J., Lu W., Villars P. & Zhang W. (2020). Active learning for the power factor prediction in diamond-like thermoelectric materials. *Computational Materials*, 6(1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41524-020-00439-8.
16. Na G. S. & Chang H. (2022). A public database of thermoelectric materials and system-identified material representation for data-driven discovery. *Computational Materials,* 8 (1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41524-022-00897-2.
17. Uysal F., Kilinc E., Kurt H., Celik E., Dugenci M. & Sagiroglu S. (2017). Estimating Seebeck coefficient of a p-type high temperature thermoelectric material using bee algorithm multi-layer perception. *Journal of Electronic Materials,* 46(8), 4931 – 4938). Springer Science and Business Media LLC. https://doi.org/10.1007/s11664-017-5497-6

18. Jia X., Deng Y., Bao, X, Yao H., Li S., Li Z., Chen C., Wang X., Mao J., Cao F., Sui J., Wu J., Wang C., Zhang Q. & Liu X. (2022). Unsupervised machine learning for discovery of promising half-Heusler thermoelectric materials. *Computational Materials*, 8 (1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41524-022-00723-9
19. Iwasaki Y., Sawada R., Stanev V., Ishida M., Kirihara A., Omori Y., Someya H., Takeuchi I., Saitoh E., & Yorozu S. (2019). Identification of advanced spin-driven thermoelectric materials via interpretable machine learning. *Computational Materials*, 5 (1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41524-019-0241-9
20. Sheng Y., Deng T., Qiu P., Shi X., Xi J., Han Y. & Yang J. (2021). Accelerating the discovery of *Cu–Sn–S* thermoelectric compounds via high-throughput synthesis, characterization, and machine learning-assisted image analysis. *Chemistry of Materials*, 33(17), 6918 - 6924). American Chemical Society (ACS). https://doi.org/10.1021/acs.chemmater.1c01856
21. Shimizu N. & Kaneko H. (2020). Direct inverse analysis based on Gaussian mixture regression for multiple objective variables in material design. *Materials & Design*, 196, 109168. Elsevier BV. https://doi.org/10.1016/j.matdes.2020.109168
22. Antunes L. M., Vikram Plata J. J., Powell A. V., Butler K. T. & Grau-Crespo R. (2022). Machine learning approaches for accelerating the discovery of thermoelectric materials. *ACS Symposium Series* (pp. 1 – 32). American Chemical Society. https://doi.org/10.1021/bk-2022-1416.ch001
23. Guo Q., Chan M., Kuropatwa B. A., & Kleinke H. (2013). Enhanced thermoelectric properties of variants of $Tl_9SbTe_6$ and $Tl_9BiTe_6$. *Chemistry of Materials,* 25(20), 4097 - 4104). American Chemical Society (ACS). https://doi.org/10.1021/cm402593f
24. Parker D., & Singh D. J. (2013). Alkaline earth lead and tin compounds $Ae_2Pb$, $Ae_2Sn$, *Ae* = *Ca*, *Sr*, *Ba*, as thermoelectric materials. *Science and Technology of Advanced Materials,* 14(5), 055003. Informa UK Limited. https://doi.org/10.1088/1468-6996/14/5/055003
25. Sun P., Oeschler N., Johnsen S., Iversen B. B. and Steglich F. (2009). Huge thermoelectric power factor: $FeSb_2$ versus $FeAs_2$ and $RuSb_2$, *Appl. Phys. Express* 2, 091102.
26. Iwasaki Y., Takeuchi I., Stanev V., Kusne A. G., Ishida M., Kirihara A., Ihara K., Sawada R., Terashima K., Someya H., Uchida K., Saitoh E, & Yorozu S. (2019). Machine-learning guided discovery of a new thermoelectric material. *Scientific Reports*, 9(1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41598-019-39278-z
27. Jaafreh R., Yoo Seong, K., Kim J.-G. & Hamad K. (2022). A deep learning perspective into the figure-of-merit of thermoelectric materials. *Materials Letters,* 319, 132299. Elsevier BV. https://doi.org/10.1016/j.matlet.2022.132299
28. Li Y., Zhang J., Zhang K., Zhao M., Hu K. & Lin, X. (2022). Large data set-driven machine learning models for accurate prediction of the thermoelectric figure of merit. *ACS Applied Materials & Interfaces*, 14(50), 55517 - 55527). American Chemical Society (ACS). https://doi.org/10.1021/acsami.2c15396
29. Wang T., Zhang C., Snoussi H. & Zhang G. (2019). Machine learning approaches for thermoelectric materials research. Advanced Functional Materials, 1906041. https://doi.org/10.1002/adfm.201906041
30. Na G.S., Chang H. (2022). A public database of thermoelectric materials and system-identified material representation for data-driven discovery. *Comput Mater* 8 (1), 214. Springer Science and Business Media LLC https://doi.org/10.1038/s41524-022-00897-2.

Submitted: 10.01.2023