

Короп М.М.

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



Короп М.М.

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.

Ключові слова: методи машинного навчання, термоелектричне матеріалознавство.

Вступ

Загальна характеристика проблеми. Машинне навчання відіграє все більш важливе значення в інтенсифікації наукових досліджень та відкриттів. Нові підходи та методи відкривають перед нами можливість пришвидшення відкриття нових перспективних матеріалів [1-3], оптимізації технологій виготовлення приладів [4,5], розрахунок найбільш економічно вигідних рішень за заданих умов [6].

Як зазначається в статті [7], властивістю термоелектрики є її використання у напрямках, що формують науково-технічний прогрес, тому вона розвивається передусім у провідних країнах світу. Термоелектричні матеріали мають великий потенціал, вже сьогодні вони слугують основою для генераторних, охолоджуючих та сенсорних приладів та датчиків, які знайшли застосування у всіх сферах людського життя від медицини до космосу. Проте, вибір оптимального термоелектричного матеріалу для заданих умов, його пошук чи оптимізація, є не тривіальною задачею, яка вимагає значних часових та матеріальних затрат.

Існує ряд досліджень по застосуванню машинного навчання в термоелектричному матеріалознавстві, вони показують високу ефективність та дешевизну використання у порівнянні із традиційними підходами.

Мета роботи є розгляд методів машинного навчання та висвітлити результати їх застосування в термоелектричному матеріалознавстві.

Керовані методи машинного навчання в термоелектричному матеріалознавстві

Машинне навчання є областю досліджень науки, яка вивчає можливості комп'ютера навчатись попередньо не будучи запрограмованим [7]. Алгоритми машинного навчання можна поділити на дві групи: керовані та некеровані. Кожна з цих груп має свою сферу застосування і обирається той алгоритм, який може дати найбільш точний результат за заданих умов.

Кероване машинне навчання – це побудова алгоритмів, здатних знаходити загальні закономірності та гіпотези, використовуючи отримані ззовні навчальні та тестові набори даних. Такий алгоритм вивчає певні шаблони в навчальному наборі даних для подальшої перевірки на тестових наборах даних [8].

Некероване машинне навчання – розпізнавання патернів та множин подібностей без залучення цільових атрибутів, всі змінні, які присутні в навчальному наборі даних беруть участь в аналізі, завдяки цьому такі алгоритми стають в нагоді для кластеризації і асоціативної сегментації [9].

Моделі машинного навчання спираються на три ключові складові: дані для навчання, дескриптори, алгоритм. Дані для навчання – це набори прикладів, з яких алгоритм намагається отримати хімічні тенденції, дескриптори – низькорівневі характеристики матеріалів (кристалічна структура, хімічна формула, середній атомний номер тощо), що дозволяє «векторизувати» хімічний матеріал та зробити його придатним для подальшого машинного навчання, алгоритм навчання – опитує навчально-векторні дані у поєднанні з певним шаблоном [10].

Kamal Choudhary та ін. [11] у своїй роботі з пошуку високоефективних 3D та 2D розмірних термоелектричних матеріалів використовують алгоритми керованого навчання, а саме: дерева рішень, випадковий ліс дерев рішень, k найближчих сусідів, багатошаровий перцептрон та техніки градієнтного підсилення.

Дерево рішень є популярним алгоритмом машинного навчання, коли поставлена задача класифікувати набір даних або провести регресію, дане рішення ефективно справляється із пропущеними значеннями та можливими помилками в наборі даних [12]. Графічне представлення можна зобразити у вигляді дерева, де вузли представляють перевірку умов для даних, а гілки – можливі результати цих умов. Дані для використання в цьому методі розбиваються на підмножини на основі ентропії, коефіцієнта Джинні і т.д. В статті Alrebdі та ін. [13], даний метод застосовується для прогнозування теплової провідності термоелектричних матеріалів на основі Bi_2Te_3 . Часто дерева рішень використовуються у якості базового елемента для ансамблевих методів, наприклад випадкового лісу дерев рішень.

Основна ідея ансамблевого методу випадкового лісу в створенні підмножин навчальних даних та навчанні окремих дерев рішень для подальшого усереднення прогнозів (завдання регресії) або комбінації (задача класифікації). Даний підхід дозволяє отримати значно вищу точність прогнозів. До переваг даного алгоритму можна віднести оцінку структури та залежності даних для аналізу даних, а до недоліків – відносно повільну швидкість навчання, зумовлену потребою створення та навчання великої кількості окремих дерев. В праці Chen та ін. [14] метод випадкового лісу використовується для відкриття нових M_2X_3 термоелектричних матеріалів тільки з інформацією про склад.

Градієнтне підсилення застосовується для композиції слабких моделей (дерев рішень), використовується градієнтний спуск, що дозволяє мінімізувати функціонал втрат. Підвищення точності передбачень досягається шляхом додавання нових моделей, які виправляють помилки попередніх моделей. Даний метод широко використовується та входить до відомих бібліотек, таких як XGBoost, LightGBM, CAT Boost. Sheng та ін. [15] використали даний алгоритм для прогнозування коефіцієнта потужності в алмазоподібних термоелектричних матеріалах.

K найближчих сусідів є одним із найпростіших алгоритмів керованого машинного навчання, який базується на класифікації нового об'єкта враховуючи класи K найближчих сусідів. Для визначення найближчого сусіда може використовуватися, наприклад, евклідова відстань. Це може призводити до повільного навчання алгоритму на великих наборах даних

через потребу розраховувати відстань між всіма парами об'єктів. Gyoung та ін. [16] використовують регресію K найближчих сусідів для прогнозування цільового значення для вхідних даних шляхом інтерполяції цільових значень K найближчих сусідів в навчальних даних.

До більш просунутих алгоритмів машинного навчання належить багатошаровий перцептрон. Він представляє собою нейронну мережу, яка складається із трьох або більше шарів, де кожен із них містить декілька нейронів. Алгоритм використовує для навчання зворотне поширення помилки, що дозволяє оптимізувати ваги нейронної мережі за допомогою градієнтного спуску. Таким чином, кожен нейрон, приймаючи вхідні дані, обчислює їх внутрішню суму та застосовує деяку функцію активації: сигмоїд, гіперболічний тангенс і тощо, додаючи нелінійність в значенні нейрона. Uysal та ін. [17] використовують алгоритм багатошарового перцептрона для оцінки коефіцієнта Зеебека для високотемпературного термоелектричного матеріалу p -типу. Даний алгоритм може виступати базою для більш складних нейронних мереж по типу згорткових нейронних мереж.

Некеровані методи машинного навчання в термоелектричному матеріалознавстві

Некеровані алгоритми машинного навчання є ефективними при роботі із наборами даних, які не містять міток, що дозволяє знаходити приховані структури. Так, існує можливість провести кластеризацію для групування матеріалів на основі їх хімічних, механічних або фізичних властивостей виявляючи нові їх класи або властивості, які є спільними для певної групи. Вони також знаходять застосування при пошуку аномалій та дефектів завдяки пошуку відхилень від стандартних значень. Одним з головних застосувань є автоматичний пошук комбінацій елементів або параметрів для отримання оптимальних матеріалів при заданих умовах або проведення їхньої оптимізації. Jia та ін. [18] використовують K -середніх, суміш Гауса, DBSCAN, AGNES, Birch для пошуку перспективних напів-Heusler термоелектричних матеріалів, а Iwasaki та ін. [19] використовують алгоритм LASSO (Метод мінімізації абсолютного стиснення та вибору оператора) для ідентифікації сучасних матеріалів, які базуються на спін-термоелектричному матеріалі.

K -середніх працює шляхом розділення набору даних на кластери (групи) і направлений на мінімізацію варіативності всередині кластерів та максимізацію між ними. K визначає кількість кластерів заданих користувачем. Алгоритм використовує K довільних центроїдів і розміщує запис даних в найближчий із них. Після завершення розміщення об'єктів, центроїди перераховують їх середнє значення в кластері, даний процес відбувається ітеративно, поки центроїди не стабілізуються і не наступить рівновага. Sheng та ін. [20] використовують цей метод для прискорення відкриття $Cu-Sn-S$ термоелектричних сполук за допомогою високопродуктивного синтезу.

Комбінація розподілів Гауса є статистичною моделлю і представляє собою суміш Гауса, де кожен компонент містить середнє, дисперсію та вагу. Даний алгоритм часто використовується для відображення еліптичних форм кластерів даних. Для навчання параметрів використовується метод максимізації правдоподібності через ЕМ алгоритм (Expectation-Maximization). Shimizu та ін. [21] використовують даний метод для зворотнього аналізу декількох цільових параметрів при дизайні матеріалів.

DBSCAN, AGNES та Birch є алгоритмами кластеризації, які використовуються для вимірювання відстані або схожості між точками даних, їхня головна мета - виявлення природних

груп у даних без необхідності явно вказувати кількість кластерів.

DBSCAN (Density-Based Spatial Clustering of Applications) працює шляхом визначення кластерів на основі великої дисперсії густини.

AGNES (Agglomerative Nesting) – поступове об'єднання об'єктів в ієрархічні кластери і візуалізацію за допомогою дендрограми.

Birch – алгоритм оптимізований для кластеризації великих наборів даних, дозволяючи швидко їх обробити.

Ці методи часто використовуються саме для відкриття нових груп перспективних матеріалів та застосовуються в ряді робіт [18, 22].

Регресійний метод LASSO працює шляхом L1 регуляції, зведенні коефіцієнтів деяких змінних до нуля, що сприяє відбору ознак, що допомагає виділити з великого набору ознак найбільш важливі. З цього випливає один недолік - при ознаках, що сильно корелюють, до уваги може взятись тільки одна.

При пошуку та оптимізації термоелектричного матеріалу, на ефективність матеріалу можуть впливати як електрична провідність, теплопровідність, коефіцієнт Зеебека, так і інші параметри. За допомогою цього методу можна визначити той набір, який дасть найбільш точний результат. Також, він дає можливість зрозуміти, як зміна умов (температура, тиск і тд.) впливатимуть на їх властивості та розуміння механізмів цього впливу. Wudil та ін. [23] використовують цей метод для оцінки продуктивності матеріалів на основі Bi_2Te_3 .

Результати застосування машинного навчання в термоелектричному матеріалознавстві

В статті Gaultois та ін. [10] опублікованій у 2015 році, автори одними із перших розробили електронну систему для оцінки параметрів термоелектричних сплавів у реальному часі на основі машинного навчання. На представлений ними ілюстрації добре видно, що термоелектричні матеріали, які найкраще досліджені, лежать в певній області періодичної таблиці (чорні та сині точки), халькогеніди та *p*-елементи, а оранжевими точками виділені нові сплави матеріалів, які запропонувала система на основі машинного навчання і ці матеріали розміщені за межами відомих матеріалів (чисті інтерметаліди).

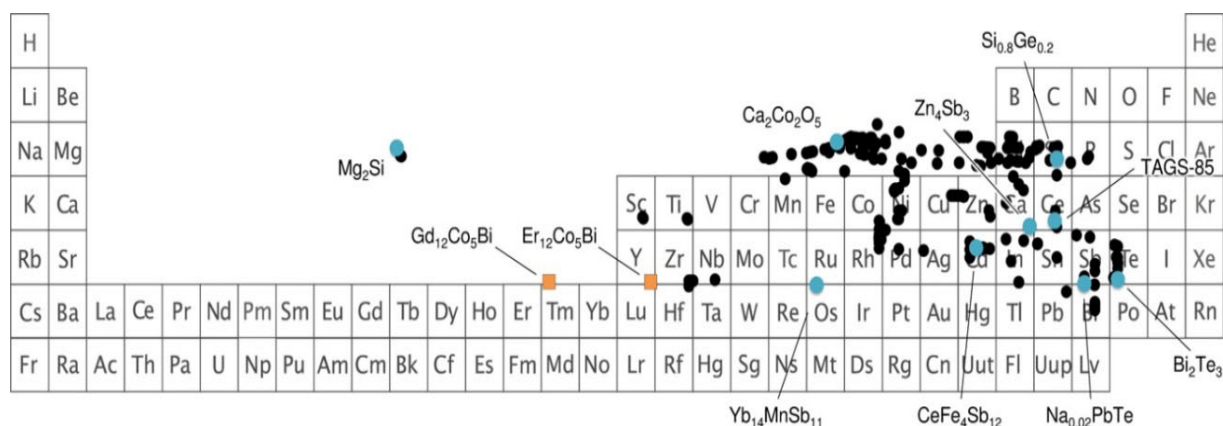


Рис. 1. Періодична таблиця елементів на основі середньозваженого за складом позицій елементів у матеріалі [10].

Автори цієї роботи розробили машинну модель, яка дає рекомендації щодо придатності нових термоелектричних сплавів для подальших досліджень, але така модель не повертає ні кількісних прогнозів щодо термоелектричних властивостей, ні точної ідентифікації термоелектричних сплавів. Така система рекомендацій на основі машинного навчання шукає емпіричні, хімічні закономірності в експериментальних даних про відомі термоелектричні матеріали для того, щоб зробити статистичні прогнози щодо ефективності нових матеріалів. При перевірці отриманих результатів, система надавала результати з великою точністю, де розподіл похибок для різних властивостей матеріалів наближається до 0. В результаті, було знайдено декілька цікавих термоелектричних сплавів вибраних із списку запропонованого алгоритмом та експериментальне підтвердження термоелектричних властивостей [23-25] (Таблиця 1).

Таблиця 1

Матеріал	P_S	P_p	P_k	P_{gap}	Коментар
$TaPO_5$ та $TaVO_5$	0.894	0.793	0.958	0.987	Висока багатогранна зв'язність та структурні надрешітки.
Tl_9SbTe_6	0.845	0.871	0.999	0.876	Нещодавно повідомлялося, що є хорошим термоелектричним матеріалом ($zT \approx 1$ при 600 K)
$TaAlO_4$	0.893	0.703	1	0.977	Високий контраст маси, висока поліедральна зв'язність (октаедри TaO_6 , що ділять ребра та вершини)
$SrCrO_3$	0.772	0.767	0.996	0.95	Висока багатогранна зв'язність (3-D кутове з'єднання CrO_6 октаедрів), металічна, коли створена під високим тиском
$TaSbO_4$	0.892	0.919	1	0.997	Висока поліедрична зв'язність: шаруваті, спільні по ребру MO_6 октаедри
$TiCoSb$	0.981	0.714	0.958	0.833	$TiCoSb$ не є новою сполукою, але вивчався як матеріал з високим значенням zT . Однак він не був включений до тренувальних даних.

У таблиці 1 представлено декілька перспективних нових термоелектричних сполук вибраних з обчисленого списку. Значення P відносяться до рівня впевненості моделі, що заданий матеріал продемонструє значення за кімнатної температури для певної властивості (наприклад, S або p) в межах цільових діапазонів [10].

Завдяки машинному навчанню, авторам вдалось першими запропонувати експериментально життєздатну нову сполуку з справжнього білого хімічного простору, де жодні попередні характеристики не вказували на перспективні хімічні процеси [10].

Ще одним напрямком, де машинне навчання може знайти перспективне застосування є знаходження залежностей та параметрів термоелектричного ефекту керованого спіном (СТЕ)

[19, 26]. Такі пристрої могли б дати універсальну термоелектричну технологію з масштабованим виробництвом, однак цьому заважає відсутність розуміння фундаментальної фізики та властивостей матеріалів, відповідальних за ефект. В статті Iwasaki та ін. [19, 26] стверджується про синтез матеріалу, який допоміг в ідентифікації нового матеріалу СТЕ з термоЕРС на порядок вищою, ніж у поточного покоління пристроїв.

Останні роботи описують появу нових, більш точних моделей, які окрім передбачення ймовірно нових цікавих для дослідження термоелектричних матеріалів, також навчаються передбачати властивості матеріалів з великою точністю, коефіцієнт детермінованості $R^2 = 0.91 - 0.959$ для добре вивчених матеріалів [27-29].

Одним із викликів, який стоїть перед покращенням точності таких моделей машинного навчання, є створення повних та якісних баз даних властивостей матеріалів, а також доступу до них [30]. Не зважаючи на постійне зростання інтересу до застосування алгоритмів машинного навчання для пошуку та відкриття нових матеріалів, на сьогодні ми володіємо дуже скромними інструментами у порівнянні із іншими сферами його застосування, а саме обробки зображень та промислового виробництва. Ступінь свободи (DoF) – це кількість змінних параметрів моделі, які є статистично важливими. Ступінь свободи напряму впливає на бажаний розмір потрібних навчальних наборів даних, тому частим є використання моделей із обмеженою кількістю змінних параметрів та приблизною оцінкою властивостей, в шуканому просторі.

В статті Gyoung S. Na та ін. [31] було сформовано публічно відкриту базу навчальних даних отриманих за допомогою пошуку по літературі, яка містить хімічну структуру та експериментально виміряні термоелектричні властивості матеріалів. Ці дані представлені у вигляді таблиці 2 і налічують 5205 експериментальних значень.

Таблиця 2

Назва властивості	Одиниці виміру	Діапазон	Середнє значення
Хімічний склад	-	-	-
Температура	К	[10, 1275]	539.28 ± 192.42
Коефіцієнт Зеєбека	мкВ/К	[-1174, 1052.4]	73.18 ± 208.92
Електрична провідність	См/м	(0, $9.47E + 07$)	$1.10E + 05 \pm 1.47E + 06$
Теплова провідність	Вт / (м·К)	[0.07, 77.16]	2.25 ± 3.29
Коефіцієнт потужності	Вт / (м·К ²)	(0, $7.61E - 03$)	$9.92E - 04 \pm 1.12E - 03$
<i>ZT</i>	-	(0, 2.28)	0.35 ± 0.35

В таблиці 2 представлено опис зібраних характеристик у наборі даних ESTM [31]. Перше значення – це хімічний склад зібраних термоелектричних матеріалів, який при застосуванні у моделях машинного навчання повинен бути переведений до цифрового представлення, друге значення – це температура. Інші параметри були отримані або експериментально, або теоретично обраховані.

В результаті, на основі зібраної бази даних, авторами [31] було розроблено машинну модель, яка досягнула точності R^2 вище 0.9 в передбаченні 5 термоелектричних властивостей матеріалів і показала середню абсолютну похибку менше ніж 0.06 при прогнозуванні ZT . Окрім публічно доступного набору даних, автори розробили метод для представлення сплавів та легованих матеріалів, що називається системно-ідентифікованим описом матеріалу (SIMD). На основі перенесення навчання із використанням SIMD, вдалось покращити показник R^2 із 0.13 до 0.71 при екстраполяції для прогнозування ZT матеріалів із не досліджених груп для пошуку нових високоефективних матеріалів.

Висновки

1. Розглянуто дослідження керованих та некерованих методів машинного навчання, які застосовуються в термоелектричному матеріалознавстві для оптимізації існуючих та пошуку нових перспективних термоелектричних матеріалів.
2. Завдяки новим можливостям, стає доступним подальший розвиток тих сфер науки, що не містять досконалого фундаментального розуміння, наприклад, термоелектричний ефект керований спіном, глибокого розуміння та визначення взаємозалежності між параметрами, які впливають на ефективність термоелектричного матеріалу.
3. Точність останніх моделей є високою. Проте, їх подальше покращення обмежується простором добре вивчених матеріалів та існуючими базами експериментально зібраних термоелектричних властивостей матеріалів або теоретично обчисленими за допомогою альтернативних методів.

Висловлюю щире подяку своєму науковому керівнику, Анатичуку Лук'яну Івановичу, за запропоновану тему роботи та корисні зауваження.

Література

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. (2021), 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 Bi_2Te_3 -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. & Sagioglu 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 Tl₉SbTe₆ and Tl₉BiTe₆. *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>.

Надійшла до редакції: 10.01.2023.

Korop M.M.

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.

Keywords: machine learning methods, thermoelectric materials science.

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 Bi_2Te_3 -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 Ti_9SbTe_6 and Ti_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 , $\text{Ae} = \text{Ca}, \text{Sr}, \text{Ba}$, 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., Kirihaara A., Ihara K., Sawada R., Terashima K., Someya H., Uchida K., Saitoh E., & Yoroze 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.