

УДК 537.32



Анатичук Л.І.

Анатичук Л.І., акад. НАН України^{1,2}

Короп М.М.¹



Короп М.М.

¹ Інститут термоелектрики НАН та МОН України,
вул. Науки, 1, Чернівці, 58029, Україна;

e-mail: anatych@gmail.com,

mykola.korop@chnu.edu.ua

² Чернівецький національний університет
імені Юрія Федьковича,
вул. Коцюбинського 2, Чернівці, 58012, Україна
e-mail: anatych@gmail.com

ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ ТЕРМОЕЛЕКТРИЧНИХ МАТЕРІАЛІВ НА ОСНОВІ Bi_2Te_3

У роботі наводяться приклади оцінки ефективності застосування машинного навчання для прогнозування властивостей термоелектричних матеріалів на основі Bi_2Te_3 . Оглянуто результати їх застосування та способи вибору оптимальних параметрів вхідних даних, описано відмінності та особливості вибору алгоритмів, етапи роботи та навчання машинних моделей, а також критерії оцінки ефективності та валідації отриманих прогнозів.

Ключові слова: методи машинного навчання, термоелектричне матеріалознавство.

Вступ

Загальна характеристика проблеми. Термоелектричні матеріали, які слугують основою для безмашинного перетворення теплової енергії в електричну знаходять все більшу популярність та прикладне застосування. Одним із головних критеріїв для оцінки перспективних матеріалів є коефіцієнт добротності Z запропонований Йоффе, який можна виразити за допомогою формули 1:

$$ZT = \alpha^2 \frac{\sigma T}{\chi}, \quad (1)$$

де α – коефіцієнт Зеєбека, σ – електрична провідність, T – температура, χ – теплова провідність.

Одним із найбільш популярних термоелектричних матеріалів є Bi_2Te_3 високе значення добротності якого досягнуте опираючись на фізичні спостереження та емпіричний підхід. Підвищення Z для цього матеріалу відбулось шляхом використання ізовалентних домішок, завдяки яким досягається зменшення χ (теплопровідності) без істотної зміни σ (електропровідності). Протягом наступних десятиліть подальшого значного росту добротності не відбулось, тому є важливим пошук нових методів та підходів.

У матеріалознавстві все більшу цікавість приділяють машинним методам оптимізації матеріалів та досягнення їх екстремальних значень. Машинне навчання розглядається як підвид штучного інтелекту, який дозволяє проводити узагальнення, інтерполяцію та екстраполяцію вхідних даних, проводити пошуки шаблонів та оперувати інформацією в більш розумний спосіб. Основним завданням машинного навчання при прогнозуванні властивостей термоелектричних матеріалів є знаходження якомога точніших значень із найменшою похибкою, на основі обмеженої кількості вхідних даних, які слугують джерелом інформації, отриманої за допомогою як теоретичних обрахунків, так і експериментальних вимірювань. Тому було поставлене питання вивчити наскільки ефективним є машинне навчання для застосування його у термоелектричному матеріалознавстві та розглянути ряд робіт по його застосуванню.

Мета роботи є вивчення ефективності методів машинного навчання в задачі прогнозування та оптимізації властивостей термоелектричних матеріалів на основі Bi_2Te_3 .

1. Керовані методи машинного навчання Bi_2Te_3

Керовані методи машинного навчання є одними із основних типів машинного навчання, де «керованість» означає, що процес навчання ведеться на основі даних із мітками (абсолютними значеннями) шуканої змінної. Такі методи часто застосовуються для вирішень задач регресії (прогнозування числового значення) або класифікації (призначення даним їхньої категорії). Точність та ефективність таких методів сильно залежить від якості навчальних даних та уникнення ризику перенавчання моделі (адаптувавшись до вхідних даних, модель перестає узагальнювати отриману інформацію та давати точні прогнози для нових даних, які модель раніше не отримувала).

В наукових працях, які присвячені вирішенню задач матеріалознавства, знайшли застосування наступні алгоритми керованого машинного навчання: лінійна регресія, логістична регресія, дерева рішень, випадковий ліс, векторні підтримуючі машини, нейронні мережі. Кожен із алгоритмів має власні переваги та сфери застосування, а одним із основних факторів вибору серед них є досягнення найвищої точності за заданих умов [2].

1.1. Регресивні алгоритми машинного навчання

В сфері машинного навчання, регресивні алгоритми є статистичними методами, які дозволяють отримати числове передбачення для залежної змінної на основі її залежності від однієї і більше незалежних змінних.

Наприклад, при прогнозуванні значення ZT (коефіцієнта термоелектричної добротності), воно (значення ZT) буде виступати залежною змінною, а вхідні дані температури, електропровідності та коефіцієнта Зеєбека виступатимуть в якості незалежних змінних.

Для вирішення регресивних задач існує ряд підходів: лінійна регресія, логістична регресія, поліноміальна, Ridge регресія та Lasso регресія [15]. Такі моделі є відносно простими в навчанні, не вимагають великих обчислювальних потужностей, але містять обмеження в можливості узагальнювати тільки прості залежності.

Лінійні регресивні моделі дозволяють знайти лінійну залежність між x та y на основі вхідного статистичного набору даних. Математичне представлення такої моделі, можна описати за допомогою формули 2.

$$y = b + \sum_{i=1}^n w_i * x_i, \quad (2)$$

де y – прогнозоване значення залежної змінної;

b – зсув (bias) — абсолютне значення, яке дозволяє моделі врахувати зміщення у вихідних значеннях, які не можуть бути пояснені незалежними змінними;

w – коефіцієнт ваг, який вказує наскільки зміна незалежної змінної x пояснює зміну залежної y .

Поліноміальні регресивні моделі призначені для пошуку нелінійних залежностей між вхідними та вихідним значеннями. Такі моделі здатні узагальнити більш складні випадки та описуються за допомогою формули 3.

$$y = b + \sum_{i=0}^n w_i * x_i^i, \quad (3)$$

Ризик при використанні поліноміальної регресії полягає у перенавчанні моделі при високих значеннях n , з високою точністю прогнозування на навчальних даних та низькою точністю із новими даними та складністю в поясненні навченої моделі.

Одними із основних причин перенавчання моделі є надто великі значення коефіцієнтів у рівняннях 2 та 3 біля вхідних параметрів (ваг) та невелика кількість навчальних даних. Для вирішення проблеми недостатньої кількості навчальних даних, можуть застосовуватись методи розширення даних, а саме: введення шумів для числових значень, зміна розміру, яскравості для зображень тощо. Тут під терміном «шум» потрібно розуміти випадкові помилки або варіації до оригінальних даних – вони дозволяють забезпечити на вхід моделі навчальні дані, які є максимально наближеними до реальних значень. У випадку перенавчання моделі, у якості запобіжних заходів, є використання регуляризації, яка є методом контролю та додає додаткові обмеження на ваги моделі. Це включає L1 (Lasso), L2 (Ridge) або комбінацію цих регуляризацій.

На кожному етапі навчання моделі, для досягнення максимальної точності, застосовуються функції витрат. Основною їх задачею є оцінка помилки моделі під час навчання, для корегування коефіцієнтів (ваг) так, щоб ця помилка була мінімальною. Для задач регресії часто застосовується середньоквадратична похибка, яка описується формулою 4. На основі значень функції витрат розраховуються градієнти для обчислення нових значень ваг для кожного кроку навчання.

$$\sigma = \frac{\sum_{i=1}^n (x_i - y_i)^2}{n}, \quad (4)$$

де x – точне значення, y – значення прогнозоване моделлю.

Таким чином, L1 та L2 регуляризації зазначені вище виступають, як додаткові члени, які додаються до функції витрат під час навчання моделі. L1 регуляризація додає абсолютні значення ваг для певних дескрипторів, тим самим змінюючи їх на 0 для параметрів, які є незначущими. L2 регуляризація додає квадрати ваг до функції витрат, тим самим «штрафуючи» модель за великі значення ваг.

1.2. Практичний спосіб застосування керованих методів машинного навчання

Організацію процесу навчання моделі штучного інтелекту можна подати у вигляді наступного списку дій.

1. **Збір навчальних даних:** на цьому етапі проводиться збір первинних даних (незалежних змінних), які можуть бути виражені числовими значеннями (кристалічна структура, хімічна формула, середній атомний номер тощо) або категорійними значеннями [3]. Залежні значення (мітки), які представляють результат або клас, що має бути передбачений, отримуються з

теоретичних розрахунків чи експериментальних вимірювань.

2. **Попередня обробка даних:** має на меті покращити якість зібраних даних, знайти та видалити шуми, відсутні чи некоректні дані. Проводиться категоризація та кодування даних або їх нормалізація [4].
3. **Вибір алгоритму моделі:** проводиться на основі підготовлених даних так, щоб досягнути поставленої мети та отримати найвищу точність. Процес вибору містить сильну залежність від поставленого завдання та вхідного набору даних. Оцінка вхідних даних, наявність шумів, кількість функцій, ступінь лінійності між змінними, знаходиться компроміс між складністю та ефективністю моделі, мінімізація можливостей перенавчання моделі.
4. **Навчання вибраної моделі та оптимізація:** коригування (ваг, порогів тощо) для досягнення цільових показників точності та мінімізації похибок.
5. **Оцінка ефективності роботи навченої моделі:** має на меті виявити та скоригувати проблеми, які виникли у процесі навчання, наприклад, точній роботі із навчальними даними та некоректній роботі із новими наборами вхідних даних.
6. **Тестування отриманої моделі:** проводиться для визначення її точності та ефективності.

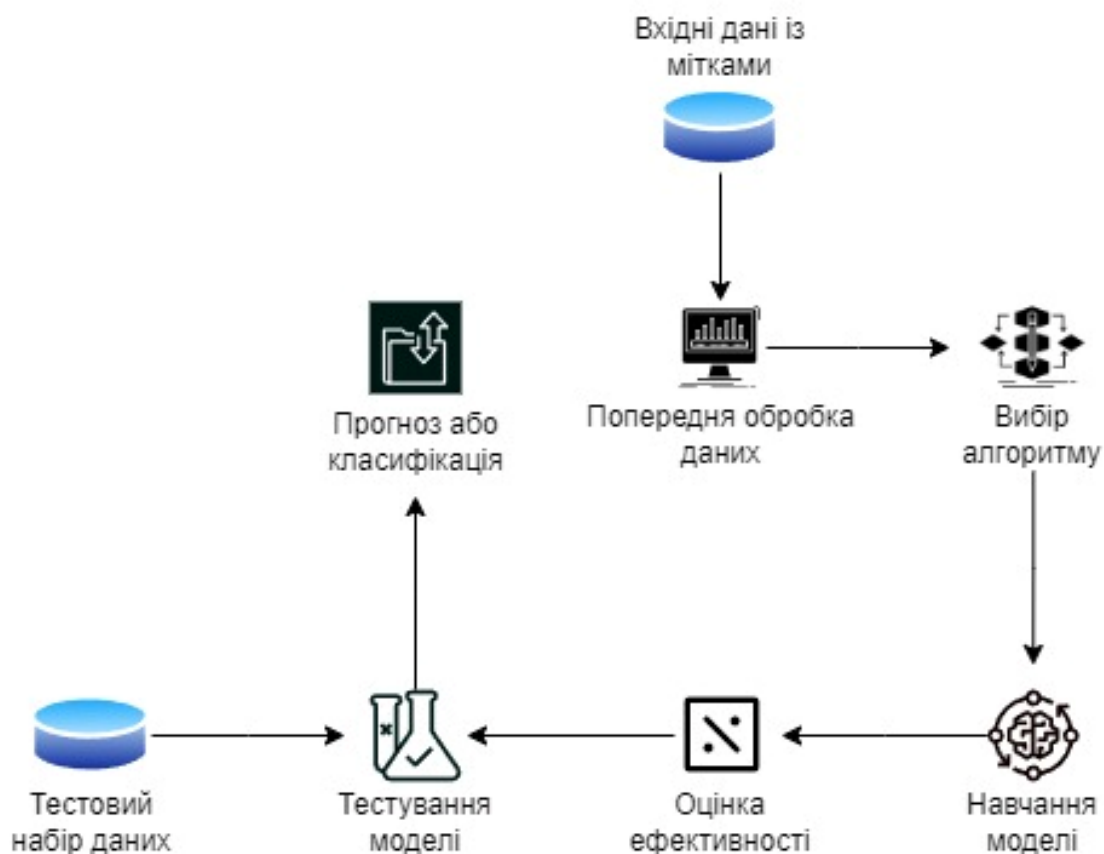


Рис. 1. Схема застосування керованого методу машинного навчання.

Wudil, Y. та ін. [1] у своїй роботі провели роботу по прогнозуванню коефіцієнта добротності матеріалів на основі Bi_2Te_3 із використанням структурних констант ґратки (а та с) та електричних властивостей матеріалів в якості предикторів. Вхідні дані були сформовані на основі ряду експериментальних робіт, звідки були взяті транспортні властивості та параметри

кристалічної ґратки. Також, для узагальнення моделі, результати формувались як для чистого матеріалу, так і для матеріалу з домішками. До уваги були взяті дві інтенсивно досліджувані сполуки *n*-типу $Bi_2Te_{2.7}Se_{0.3}$ та *p*-типу $Bi_{1.5}Sb_{0.5}Te_3$. Фінальний набір даних містив 280 точок даних.

Дескрипторами було вибрано 5 параметрів: коефіцієнт Зеебека, електрична провідність, температура, розміри ґратки (*a*, *c*). Ці параметри було вибрано за допомогою знаходження взаємозалежності цільового значення та вхідних даних, визначеної за допомогою коефіцієнта кореляції Пірсона. Вибір параметрів ґратки є доцільним через залежність структурних параметрів ґратки від методів та способів виготовлення матеріалу.

Parv Katyal та ін. [5] в своєму дослідженні поставили собі за ціль визначити алгоритми та методи машинного навчання, які можуть працювати з обмеженим набором вхідних даних для встановлення взаємозалежностей між коефіцієнтом *ZT* та хімічними і фізичними властивостями сполук Bi_2Te_3 , $CoSb_3$, $Ba_8Ga_{16}Ge_{30}$. Для навчання штучного інтелекту було використано набір даних із 1098 обрахованих точок, в цьому наборі значення *ZT* були отримані за різних температур. Дескрипторами було обрано: температуру, коефіцієнт Зеебека, коефіцієнт потужності, об'єм комірки, опір, загальна маса однієї комірки, середній атомний об'єм, просторова група, елементи симетрії. Набір даних був поділений у довільному порядку у пропорції 80 % – навчальні дані та 20 % – дані для тестування. Для машинного навчання було відібрано метод випадкового лісу, який є ансамблевим методом, полягає у побудові декількох дерев рішень під час тренування та в результаті обчислення середнього, прогнозування окремих дерев [6]. Авторами [5] набір даних було поділено на 500 дерев рішень і обраховано середнє значення результатів всіх дерев.

Zhi-Lei Wang та ін. [7] провели дослідження із прогнозування властивостей екструдованих зразків $Cu_xBi_2Te_{2.85}Se_{0.15}$ за допомогою машинного навчання. Було використано набір даних із семи експериментальних наборів даних та 12 характеристик: склад, відносна щільність, фактор орієнтації, середній розмір зерна, мікроструктурна особливість, характеристики границь зерен, концентрація носіїв заряду, рухливість, коефіцієнт Зеебека, електричний опір, теплопровідність. В якості алгоритму моделювання було вибрано штучну нейронну мережу із одним прихованим шаром, сигмоїдальну функцію було застосовано в якості функції активації, а гіперпараметри розмір вузла прихованого шару та зменшення ваг було оптимізовано за допомогою байєсової оптимізації при швидкості навчання 0.01. Для уникнення перенавчання моделі, вхідні дані були розділені у пропорції 60 % – навчальні та 40 % – тестові із застосуванням 4-х кратної крос-валідації [8]. Крос-валідація допомагає знизити ризик перенавчання моделі, шляхом поділу набору даних на декілька рівних чи практично рівних частин. Після цього, модель тренується на всіх частинах даних, а остання залишається для тестування. Даний процес повторюється декілька разів, при цьому щоразу, у якості тестового набору вибирається інша частина. В результаті обчислюються середні показники та знаходиться загальна оцінка ефективності.

В роботі Qu, R. [16] досліджуються теплопровідні властивості $MnBi_2Te_4$ та супер ґратки $Bi_2Te_3/MnBi_2Te_4$ шляхом застосування глибокої нейронної мережі (DNNP), що використовує дані отримані за допомогою теорії функціоналу густини. Використовуючи розрахунки DFT, набори даних, що містять атомні конфігурації, відповідні енергії та сили та температури (в діапазоні від 200 K до 500 K), були підготовлені для навчання моделі. Для надійності тренування було підготовлено 1200 конфігурацій, розділених рівномірно по зазначеному температурному діапазону, що забезпечує всебічне представлення станового простору системи. DNNP модель була структурована з вбудованим шаром, за яким слідує три прихованих шари, кожен з 160 вузлами. Ця архітектура була обрана для ефективного вловлення складності атомних взаємодій

у досліджуваних матеріалах. Процес тренування був ретельно налаштований, з оптимізацією функцій втрат та гіперпараметрів, щоб звести до мінімуму помилку між прогнозами DNNP та обчисленнями DFT.

Дослідження Agarwal A. та ін. [17] розглядає спосіб передбачення термоелектричного коефіцієнта потужності для матеріалу Bi_2Te_3 під час процесу лазерного плавлення порошкового ложа (PBF-LB), використовуючи методи машинного навчання для адитивного виробництва (AM). Адитивне виробництво (AM) відноситься до групи технологій, які створюють об'єкти, додаючи матеріал шар за шаром, на основі цифрових 3D моделей. На відміну від традиційних методів субтрактивного виробництва, які починають з твердої заготовки, а потім вирізають зайве для створення деталі, AM будує деталі безпосередньо з сировини шар за шаром, що мінімізує відходи та дозволяє створювати складні геометрії, які були б складно або неможливо досягти звичайними методами виробництва. Використовуються ансамблеві методи машинного навчання, для передбачення коефіцієнта потужності Bi_2Te_3 . Із використанням спеціалізованого обладнання було зібрано параметри обробки та дані сенсорів, а саме: потужність лазера, швидкість, проміжок розкладки, товщину шару та фокус. Для обробки зображень було застосовано бібліотеку Python OpenCV для перетворення отриманих зображень сенсорів у значущі ознаки, включаючи текстуру, шорсткість поверхні та статистику інтенсивності пікселів. Було проведено масштабування ознак та розподіл даних (80 % для навчання, 20 % для тестування), щоб забезпечити надійність моделі та узагальнення.

Headley, C. V. та ін. [18] в своїй роботі описується інноваційне поєднання технік машинного навчання (ML) з процесами адитивного виробництва (AM) для оптимізації виготовлення термоелектричних матеріалів. Зокрема, робота звертає увагу на виробництво деталей з *n*-типу $Bi_2Te_{2.7}Se_{0.3}$ за допомогою лазера для плавлення та злиття порошкового матеріалу в тверді 3D об'єкти. Дослідження використовує регресію на основі опорних векторів (SVR) у межах ансамблевого навчання. Використано метод автоматичного завантаження вибірки (bootstrap) для створення ансамблю моделей SVR, техніка, яка передбачає повторну вибірку з навчального набору даних із заміщенням для навчання кількох моделей. Основними дескрипторами для цих моделей є параметри процесу LPBF, такі як потужність лазера (яка змінювалася від 10 Вт до 40 Вт) та швидкість сканування (яка варіювалася від 250 мм/с до 550 мм/с), що безпосередньо впливає на ширину та глибину розплавленої області «басейну». Для оцінки моделі, застосовано середнє значення та стандартне відхилення прогнозів з ансамблю моделей SVR. Ці метрики є важливими для розуміння точності та надійності прогнозів ML щодо розмірів розплавленого «басейну», який утворюється під час процесу LPBF.

2. Некеровані методи машинного навчання Bi_2Te_3

Некеровані методи машинного навчання є окремою категорією алгоритмів, здатних вчитись із вхідних даних без прямого управління та визначених міток [9]. Такі методи знаходять шаблони поведінки у даних та їх взаємозв'язки у автономному режимі, що дозволяє ефективно вирішувати задачі кластеризації чи пошуку асоціацій. Некероване навчання є універсальним методом, так як може знаходити неочевидні зв'язки у даних у складних структурах. До популярних алгоритмів належать: *k*-середніх, ієрархічна кластеризація, алгоритм головних компонент, нейронні мережі. У таблиці 1, представлено наочне порівняння керованих та некерованих методів машинного навчання.

Таблиця 1

Порівняння керованих та некерованих методів машинного навчання.

Керовані методи машинного навчання	Некеровані методи машинного навчання
Прогнозування числового значення або класифікація вхідних даних із мітками.	Пошук шаблонів та взаємозалежностей у даних на основі даних без будь-яких асоційованих міток.
Потрібна обробка вхідних даних, їх нормалізація, можливе кодування міток, вибір характеристик чи розробки специфічних характеристик для задачі прогнозування.	Основний акцент ставиться на виборі характеристик для окреслення найбільш важливих аспектів даних, відсутнє використання міток та їх обробка.
Вибір моделі відбувається залежно від поставленої задачі (регресія або кластеризація), ключові параметри моделі налаштовуються для мінімізації розходження між фактичними та прогнозованими мітками.	На основі типу розпізнання шаблонів (асоціація, кластеризація, зниження розмірності) вибирається потрібний алгоритм без посилення на вхідні дані із мітками.
В якості параметрів оцінки ефективності навчання моделі використовуються: абсолютна та відносна похибка, прецизійність, середньоквадратична похибка. Вхідні дані розбиваються на навчальні та тренувальні.	Індекс силуета та індекс Девіса-Болдіна для задач кластеризації, а також суб'єктивна оцінка належності кожного відібраного екземпляру до інших об'єктів в класі.
Отримані результати інтерпретуються в контексті поставленої задачі та можливе проведення повторного калібрування моделі на основі оцінки ефективності.	Існує потреба в візуалізації отриманих результатів для кращого інтерпретування та зосереджений на розумінні виявлених групувань та шаблонів.

Узагальнюючи вище наведену таблицю, можна охарактеризувати керовані методи машинного навчання як такі, що працюють із вхідними даними з мітками та дозволяють вирішувати задачі прогнозування та кластеризації. У свою чергу, некеровані методи машинного навчання керують дослідницьким процесом з виявлення шаблонів та прихованих структур взаємозалежностей у вхідних даних без міток.

На даний момент не було знайдено робіт, які використовують методи некерованого машинного навчання для дослідження термоелектричного матеріалу Bi_2Te_3 , проте існує декілька цікавих напрямків їх застосування:

1. Постановка задачі кластеризації для групування зразків матеріалів Bi_2Te_3 за схожістю до певного класу, що дозволить виявити закономірності та кореляції у великих наборах даних щодо впливу методів синтезу, додавання домішок чи наноструктури на продуктивність таких зразків. Таке дослідження допомогло б розробити методологію та рекомендації для покращення процесу синтезу та обробки матеріалу для вирішення поставлених задач [10].

2. Застосування алгоритмів некерованого навчання для видобування ознак мікроскопічних зображень або кристалографічних даних Bi_2Te_3 для пошуку та кластеризації дефектів або кристалічних структур. Результати роботи можна використати, як спосіб підвищення якості контролю за виготовленням термоелектричного матеріалу та оцінки впливу дефектів на ефективність зразків [11].
3. Збір та підготовка набору даних про домішки, які використовуються та традиційно не досліджені з Bi_2Te_3 для пошуку перспективних комбінацій. Завдяки високій продуктивності машинного навчання та специфіці використання некерованих методів існує можливість опрацювати значний об'єм можливих сполук [12].

3. Результати застосування машинного навчання Bi_2Te_3

В статті Wudil та ін. [1] було розроблено п'ять слабких регресійних моделей (регресія ласо, лінійна регресія, регресія дерев рішень, регресія опорних векторів) та одну сильну модель, яка поєднує попередні п'ять за допомогою техніки ансамблю з використанням AdaBoost [13]. Для оцінки ефективності цих моделей було використано коефіцієнти кореляції, середню абсолютну помилку, коефіцієнт детермінованості R^2 та середньоквадратичну помилку. В результаті, моделі регресії дерев рішень та регресії опорних векторів показали високі коефіцієнти кореляції в 99 % та 90.8 % відповідно. Посилені моделі, за допомогою алгоритму AdaBoost показали ще вищі показники 99.5 % та 94 %. Під час валідації було підкреслено, що моделі регресії дерев рішень та регресії опорних векторів із зменшеними значеннями середньої абсолютної похибки та середньоквадратичної похибки є ефективні при оцінці якості матеріалу. На основі цього дослідження зроблено висновок, що впровадження посилених слабких регресивних алгоритмів значно посилює точність прогнозування термоелектричних напівпровідників на основі Bi_2Te_3 .

В іншій роботі Wudil [14] представлено наукове дослідження з використанням машинного навчання для оцінки теплопровідності матеріалів на основі $Bi_2Te_{2.7}Se_{0.3}$. В цій роботі також були використані алгоритми регресії дерева рішень та регресії опорних векторів підсилених за допомогою адаптивного підсилення AdaBoost. Для вибору дескрипторів було знайдено кореляцію між вхідними параметрами набору даних та шуканим значенням, представлено в таблиці 2.

Таблиця 2

Коефіцієнти кореляції між вхідними параметрами та цільовою змінною.

	σ (См/м)	S (мкВ/К)	a (А)	c (А)	K (Вт/(м·К))	T (К)
σ (См/м)	1	-0.71	-0.15	-0.57	0.62	-0.36
S (мкВ/К)	-0.71	1	-0.22	0.33	-0.74	-0.088
a (А)	-0.15	-0.22	1	0.4	-0.11	0.076
c (А)	-0.57	0.33	0.4	1	-0.31	0.25
K (Вт/(м·К))	0.62	-0.74	-0.11	-0.31	1	0.29
T (К)	-0.36	-0.088	0.076	0.25	0.29	1

Модель використовує електричні властивості та структурні параметри ґраток матеріалів, як вхідні характеристики. Ефективність розроблених моделей оцінюється на основі таких параметрів, як коефіцієнт кореляції, середня абсолютна похибка та середньоквадратична похибка. Модель дерева рішень з посиленням AdaBoost, показали коефіцієнт кореляції 99.4 % та коефіцієнт детермінованості R^2 98.8 % у тестовій фазі. Ці моделі також було використано для прогнозування теплопровідності для різних фізичних екземплярів, таких, як сполуки з домішками перехідних металів. Досліджено вплив температури субстрату під час імпульсного лазерного осадження.

В роботі Parv Katyal та ін. [5] результати дослідження підтвердили високу ефективність прогнозування значення ZT для різних сполук (Bi_2Te_3 , $CoSb_3$, $Ba_8Ga_{16}Ge_{30}$, $Ba_8Ga_{18}Ge_{28}$) за допомогою алгоритму випадкового лісу дерев рішень. Оцінка ефективності моделі показала низьке розходження із очікуваним результатом та середньою абсолютною похибкою в 0.0734, що показує перспективність цього методу у процесах оцінки термоелектричного матеріалу. В таблиці 3 представлені результати роботи запропонованої моделі для передбачення ZT за різних температур для групи сполук.

Таблиця 3

Експериментальні та прогнозовані значення ZT на різних температурах для сімейства Телуриду Свинцю, сімейства Антимоніду Кобальту та Германій-базованих Клатратів.

Температура (K)	Експериментальне значення ZT	Передбачене моделлю значення ZT	Хімічна формула матеріалу
400	0.5025	0.5923415	Bi_2Te_3
700	1.392715388	1.4253434	Bi_2Te_3
1000	1.636067789	1.5812441	Bi_2Te_3
600	0.871875767	0.938096442	Bi_2Te_3
300	0.316428584	0.384607488	Bi_2Te_3
300	0.424000225	0.502502554	$CoSb_3$
400	0.668512792	0.578206119	$CoSb_3$
700	1.181566055	1.16347007	$CoSb_3$
700	0.668168	0.7181711	$Ba_8Ga_{16}Ge_{30}$
300	0.01603609	0.067494695	$Ba_8Ga_{18}Ge_{28}$
1000	0.962666667	1.1242445	$Ba_8Ga_{18}Ge_{28}$

В роботі Wang Z. та ін. [7] дослідження показало, що додавання домішок міді до матеріалів $Bi-Te-Se$, покращує їх термоелектричні характеристики. Атоми міді вводяться в міжсіткові місця,

змінюючи мікроструктуру матеріалу та знижуючи концентрацію носіїв заряду. Це призводить до зростання коефіцієнта Зеебека, електричного опору та зниження теплової провідності носіїв. У статті використовуються методи машинного навчання, зокрема моделі штучної нейронної мережі (ANN) та байєсової оптимізації, для прогнозування та оптимізації термоелектричних властивостей цих матеріалів. Хоча модель машинного навчання є перспективною, були відзначені проблеми, пов'язані з перенавчанням, через малий розмір вибірки.

В роботі Qu, R. [16] модель досягла високої точності, як свідчать низькі значення середньоквадратичної похибки 0.15 meV на атом для конфігурацій супер ґратки для прогнозів енергії та сил. Прогнози теплопровідності для $MnBi_2Te_4$ добре збігалися з експериментальними значеннями, підтверджуючи високу ефективність моделі DNNP. Невеликі розбіжності, що були відзначені, були в межах прийнятних меж, підкреслюючи складнощі при виключенні решіткової від електронної теплопровідності в експериментальних вимірах. Супер ґратка $Bi_2Te_3/MnBi_2Te_4$ виявила значно знижену поперечну теплопровідність, що може мати важливе значення для потенційних термоелектричних застосувань. Зокрема, поперечна теплопровідність при 300 K була прогнозована на рівні $0.15 \text{ Вт} \cdot \text{м}^{-1} \cdot \text{К}^{-1}$, значно нижче, ніж чистий $MnBi_2Te_4$ або Bi_2Te_3 , демонструючи вищий термоелектричний потенціал супер ґратки. Подальший аналіз результатів показав, що зниження поперечної теплопровідності в супер ґратці можна пояснити дисперсійними відносинами фононів, зокрема появою заборонених зон та зниженням швидкості фононів. Ці поведінки фононів є ключовими для розуміння механізмів, що керують тепловою передачею у цих складних матеріалах.

Agarwal, A., та ін. [17] Модель класифікатора на основі агрегування «бегінгу» показала найвищу точність у 90 %, що вказує на значущу кореляцію між обраними ознаками та коефіцієнтом потужності термоелектричного матеріалу. Фокус лазера, потужність та швидкість були серед основних параметрів обробки, які впливають на коефіцієнт потужності. Особливості, пов'язані з даними поляриметрії, особливо після розподілу та після плавлення кута поляризації (AoP) та ступеня лінійної поляризації (DoLP), були критично важливими для прогнозування коефіцієнта потужності. Загалом було виготовлено 220 зразків, для аналізу використовувалися 117, в результаті чого було отримано 3157 точок даних для побудови моделей машинного навчання.

Headley, C. V та ін. [18] на початковому етапі використовували 13 ліній сканування, які пізніше розширились до 93 комбінацій параметрів після шести раундів навчання, демонструючи ефективність ітеративного, керованого даними підходу для уточнення параметрів процесу. Машинний метод, виявив параметри процесу LPBF, які призвели до виробництва деталей $Bi_2Te_{2.7}Se_{0.3}$ з щільністю більше ніж 99 % і без тріщин, демонструючи високу точність і якість, які можна досягти. Одним із помітних досягнень була здатність виготовляти термоелектричні деталі з нетиповими геометриями, такими як порожнисті прямокутники та трапеції, з відносною щільністю 98.6 % (± 1 %) та підвищеною термоелектричною ефективністю. Наведені вище форми складно виготовити традиційними методами виробництва, але їх можна досягти через LPBF завдяки точному контролю над параметрами процесу.

Висновки

1. Оцінено ефективність методів машинного навчання для прогнозування властивостей для термоелектричного матеріалу Bi_2Te_3 .
2. Керовані алгоритми машинного навчання, а саме слабкі моделі із підсиленням AdaBoost

регресія дерев рішень та регресія опорних векторів, добре підходять для прогнозування коефіцієнта добротності ZT та оцінки теплопровідності термоелектричних матеріалів на базі Bi_2Te_3 .

3. Оцінка ефективності використання ансамблевого методу, випадкового лісу дерев рішень, показала низьке розходження із очікуваним результатом та середньою абсолютною похибкою в 0.0734 для сполук (Bi_2Te_3 , $CoSb_3$, $Ba_8Ga_{16}Ge_{30}$, $Ba_8Ga_{18}Ge_{28}$).
4. Існує значна складність в кількості існуючої експериментальної вимірної інформації про термоелектричні матеріали, що змушує дослідників працювати з обмеженим набором інформації, що в свою чергу призводить до зниження точності прогнозів.
5. За допомогою машинного навчання можна визначити ряд параметрів для ефективного виготовлення термоелектричних запчастин за допомогою адитивних методів виробництва.

Література

1. Wudil, Y. & Gondal, M.A. (2022). Predicting the thermoelectric energy figure of merit of Bi_2Te_3 -based semiconducting materials: A machine learning approach. *SSRN Electronic Journal. Elsevier BV*. <https://doi.org/10.2139/ssrn.4215166>
2. Burkart, N. & Huber, M.F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70, 245 – 317. AI Access Foundation. <https://doi.org/10.1613/jair.1.12228>
3. 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>
4. Gonzalez Zelaya, C.V. (2019). Towards explaining the effects of data preprocessing on machine learning. *IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE. <https://doi.org/10.1109/icde.2019.00245>
5. Parv Katyal, Madhav Rathi, Piyush Mehra and Amrish K. Panwar (2020). Evaluation of figure of merit of thermoelectric materials using machine learning. *International Journal of Advanced Science and Technology*, 29 (11s), 2858 – 2863. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/23766>
6. Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: random forest. In *Information Computing and Applications* (pp. 246 – 252). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34062-8_32
7. Wang, Z., Yokoyama, Y., Onda, T., Adachi, Y., & Chen, Z. (2019). Improved thermoelectric properties of hot-extruded $Bi\text{-}Te\text{-}Se$ bulk materials with Cu doping and property predictions via machine learning. *Advanced Electronic Materials*, 5 (6). Wiley. <https://doi.org/10.1002/aelm.201900079>
8. A. Ramezan, C., A. Warner, T., & E. Maxwell, A. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing*, 11 (2), 185. MDPI AG. <https://doi.org/10.3390/rs11020185>
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. Na, G. S. (2023). Artificial intelligence for learning material synthesis processes of thermoelectric

- materials. *Chemistry of Materials*, 35(19), 8272 – 8280). American Chemical Society (ACS). <https://doi.org/10.1021/acs.chemmater.3c01834>
11. 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>
 12. 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>
 13. CAO, Y., MIAO, Q.-G., LIU, J.-C., & GAO, L. (2013). Advance and prospects of AdaBoost algorithm. *Acta Automatica Sinica*, 39 (6), 745 – 758. Elsevier BV. [https://doi.org/10.1016/s1874-1029\(13\)60052-x](https://doi.org/10.1016/s1874-1029(13)60052-x)
 14. Wudil Y.S. (2023). Ensemble learning-based investigation of thermal conductivity of *Bi₂Te_{2.7}Se_{0.3}*-based thermoelectric clean energy materials. *Results in Engineering*, 18, 101203. Elsevier BV. <https://doi.org/10.1016/j.rineng.2023.101203>
 15. Wang T., Zhang C., Snoussi H., & Zhang G. (2019). Machine learning approaches for thermoelectric materials research. *Advanced Functional Materials*, 30 (5). Wiley. <https://doi.org/10.1002/adfm.201906041>
 16. Qu, R., Lv, Y., & Lu, Z. (2023). A deep neural network potential to study the thermal conductivity of *MnBi₂Te₄* and *Bi₂Te₃/MnBi₂Te₄* superlattice, *Journal of Electronic Materials*, 52 (7), 4475 – 4483). Springer Science and Business Media LLC. <https://doi.org/10.1007/s11664-023-10403-z>
 17. Agarwal, A., Banerjee, T., Gockel, J., LeBlanc, S., Walker, J., & Middendorf, J. (2023). *Predicting thermoelectric power factor of bismuth telluride during laser powder bed fusion additive manufacturing (Version 1)*. arXiv. <https://doi.org/10.48550/ARXIV.2303.15663>
 18. Headley, C. V., Herrera del Valle, R. J., Ma, J., Balachandran, P., Ponnambalam, V., LeBlanc, S., Kirsch, D., & Martin, J. B. (2024). The development of an augmented machine learning approach for the additive manufacturing of thermoelectric materials. *Journal of Manufacturing Processes*, 116, 165 – 175). Elsevier BV. <https://doi.org/10.1016/j.jmapro.2024.02.045>

Надійшла до редакції: 12.04.2023.

Anatychuk L.I., Acad. NAS Ukraine^{1,2}

Korop M.M.¹

¹ Institute of Thermoelectricity of the NAS and MES of Ukraine,

1 Nauky str., Chernivtsi, 58029, Ukraine;

e-mail: anatych@gmail.com, mykola.korop@chnu.edu.ua

² Yuriy Fedkovych Chernivtsi National University, 2 Kotsiubynskyi str.,

Chernivtsi, 58000, Ukraine

e-mail: anatych@gmail.com

APPLICATION OF MACHINE LEARNING TO PREDICT THE PROPERTIES OF Bi_2Te_3 - BASED THERMOELECTRIC MATERIALS

The paper provides examples of assessing the effectiveness of machine learning for predicting the properties of Bi_2Te_3 -based thermoelectric materials. The results of their application and methods for selecting optimal input data parameters are considered, the differences and features of choosing algorithms, the stages of work and training machine models, as well as the criteria for assessing the effectiveness and validation of the obtained forecasts are described.

Key words: machine learning methods, thermoelectric materials science.

References

1. Wudil, Y. & Gondal, M.A. (2022). Predicting the thermoelectric energy figure of merit of Bi_2Te_3 -based semiconducting materials: A machine learning approach. *SSRN Electronic Journal*. Elsevier BV. <https://doi.org/10.2139/ssrn.4215166>
2. Burkart, N. & Huber, M.F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70, 245 – 317. AI Access Foundation. <https://doi.org/10.1613/jair.1.12228>
3. 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>
4. Gonzalez Zelaya, C.V. (2019). Towards explaining the effects of data preprocessing on machine learning. *IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE. <https://doi.org/10.1109/icde.2019.00245>
5. Parv Katyal, Madhav Rathi, Piyush Mehra and Amrish K. Panwar (2020). Evaluation of figure of merit of thermoelectric materials using machine learning. *International Journal of Advanced Science and Technology*, 29 (11s), 2858 – 2863. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/23766>
6. Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: random forest. In *Information Computing and Applications* (pp. 246 – 252). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34062-8_32
7. Wang, Z., Yokoyama, Y., Onda, T., Adachi, Y., & Chen, Z. (2019). Improved thermoelectric properties of hot-extruded $Bi-Te-Se$ bulk materials with Cu doping and property predictions via machine learning. *Advanced Electronic Materials*, 5 (6). Wiley. <https://doi.org/10.1002/aelm.201900079>
8. A. Ramezan, C., A. Warner, T., & E. Maxwell, A. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing*, 11 (2), 185. MDPI AG. <https://doi.org/10.3390/rs11020185>
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. Na, G. S. (2023). Artificial intelligence for learning material synthesis processes of thermoelectric materials. *Chemistry of Materials*, 35(19), 8272 – 8280). American Chemical Society (ACS). <https://doi.org/10.1021/acs.chemmater.3c01834>

11. 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>
12. 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>
13. CAO, Y., MIAO, Q.-G., LIU, J.-C., & GAO, L. (2013). Advance and prospects of AdaBoost algorithm. *Acta Automatica Sinica*, 39 (6), 745 – 758. Elsevier BV. [https://doi.org/10.1016/s1874-1029\(13\)60052-x](https://doi.org/10.1016/s1874-1029(13)60052-x)
14. Wudil Y.S. (2023). Ensemble learning-based investigation of thermal conductivity of *Bi₂Te_{2.7}Se_{0.3}*-based thermoelectric clean energy materials. *Results in Engineering*, 18, 101203. Elsevier BV. <https://doi.org/10.1016/j.rineng.2023.101203>
15. Wang T., Zhang C., Snoussi H., & Zhang G. (2019). Machine learning approaches for thermoelectric materials research. *Advanced Functional Materials*, 30 (5). Wiley. <https://doi.org/10.1002/adfm.201906041>
16. Qu, R., Lv, Y., & Lu, Z. (2023). A deep neural network potential to study the thermal conductivity of *MnBi₂Te₄* and *Bi₂Te₃/MnBi₂Te₄* superlattice, *Journal of Electronic Materials*, 52 (7), 4475 – 4483). Springer Science and Business Media LLC. <https://doi.org/10.1007/s11664-023-10403-z>
17. Agarwal, A., Banerjee, T., Gockel, J., LeBlanc, S., Walker, J., & Middendorf, J. (2023). *Predicting thermoelectric power factor of bismuth telluride during laser powder bed fusion additive manufacturing (Version 1)*. arXiv. <https://doi.org/10.48550/ARXIV.2303.15663>
18. Headley, C. V., Herrera del Valle, R. J., Ma, J., Balachandran, P., Ponnambalam, V., LeBlanc, S., Kirsch, D., & Martin, J. B. (2024). The development of an augmented machine learning approach for the additive manufacturing of thermoelectric materials. *Journal of Manufacturing Processes*, 116, 165 – 175). Elsevier BV. <https://doi.org/10.1016/j.jmapro.2024.02.045>

Submitted: 12.04.2023.