

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
Факультет електроніки та інформаційних технологій

Кафедра наноелектроніки

Кваліфікаційна робота бакалавра

**ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ БАГАТОКОМПОНЕНТНИХ
СПЛАВІВ З ВИКОРИСТАННЯМ АЛГОРИТМІВ ШТУЧНОГО ІНТЕЛЕКТУ**

Студент гр. ФЕ-61

М. А. Погорелий

Науковий керівник
к.ф. –м. н. ст. викл.

В. М. Рогоз

Суми – 2020

РЕФЕРАТ

Мета роботи – дослідження методів прогнозування властивостей сплавів, використовуючи алгоритми штучного інтелекту.

Робота присвячена дослідженню можливостей використання алгоритмів штучного інтелекту, базових принципів та інтеграції машинного навчання для отримання твердості та модуля Юнга для багатокomпонентних сплавів.

В даній бакалаврській роботі було проведено наукове дослідження методів аналізування та прогнозування параметрів багатокomпонентних сплавів використовуючи штучний інтелект та принципи машинного навчання. Для виконання завдання були розглянуті основні поняття з побудови штучного інтелекту, машинного навчання, розглянуто було який саме алгоритм машинного навчання буде використаний з розглянутих в загальному випадку. Був проаналізований алгоритм роботи, а також були отримані певні значення для на основі нікель-ніобію бінарного складу та тернарні на основі титану, ванадію, молібдену, ніобію, аргентуму, вольфраму. У результаті порівняння даних моделювання та експериментальних даних були знайдені близькі значення та найкращі значення для твердості та модуля Юнга відносно до твердості, а саме для бінарних це ОЦК $Nb_{24}Ni_{76}$, та ГЦК $Nb_{32}Ni_{68}$ для бінарних, та $V_{40}W_{40}Ag_{20}$ для тернарних сплавів.

. Також були знайдені найкращі значення відносно твердості, та значення модуля Юнга по відношенню до твердості

Робота викладена на 75-х сторінках, зокрема містить 21 рисунки та список цитованої літератури із 37-х джерел.

КЛЮЧОВІ СЛОВА: ШТУЧНИЙ ІНТЕЛЕКТ, МАШИННЕ НАВЧАННЯ, МОДУЛЬ ЮНГА, ТВЕРДІСТЬ, МЕТОД ВІККЕРСА, ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ, СПЛАВИ, КРИСТАЛІЧНА РЕЦІТКА.

ЗМІСТ

Вступ.....	5
1. Літературний огляд.....	8
1.1 Актуальність розвитку.....	8
1.2 Основні принципи машинного навчання	11
1.4 Принципи взаємодії з базами даних.....	19
1.5 Приклади поширених алгоритмів	20
1.6 Покращення методів машинного навчання.....	26
1.7 Приклади застосування машинного навчання та штучного інтелекту для покращення результатів матеріалів.....	27
1.8 Твердість матеріалу. Методи вимірювання	31
1.9 Модуль Юнга.....	32
2 Методи дослідження.....	38
2.1 Засоби розробки	38
2.2 Особливості будови штучного інтелекту для дослідження модуля еластичності та твердості.....	39
2.3 Принцип роботи алгоритмів штучного інтелекту	40
2.3.1. Алгоритм лінійної регресії.....	41
2.3.2. Випадкові ліси (Random forest)	42
2.4 Особливості генерації структур та композитного складу.	43
2.5 Фізична модель для знаходження твердості та модуля пружності	45
2.6 Хід роботи.....	48
3. РЕЗУЛЬТАТИ.....	53
3.1 Твердість та модуль Юнга матеріалу для Ni-Nb с ГЦК та ОЦК структурою.....	53

3.2 Твердість та модуль Юнга матеріалу для матеріалів с ОЦК структури для Ti, V, Nb, Mo, Ag.....	59
4 ТЕХНІКА БЕЗПЕКИ ТА ОХОРОНА ЗДОРОВ'Я.....	63
ВИСНОВКИ.....	70
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	72

ВСТУП

За останні декілька десятиліть наука зазнавала розвитку в багатьох сферах і галузях знань, отримуючи різні дані. У галузі обчислювального матеріалознавства збір цих даних та їх подальший аналіз можливий завдяки розвитку обчислювальних можливостей, створення машинних алгоритмів розрахунку різних типів параметрів, а також успіху в теоретичних базисах. Розвиток в галузях електроніки та машинобудування вже на даному етапі дозволяють отримати велику кількість розрахунків та результатів від експериментальних та теоретичних методів. На сьогоднішній день існують великі бази даних подібних параметрів, які визнані світовою спільнотою, які вдало можна застосовувати у різних галузях, окрім матеріалознавства. На сьогодні у галузі матеріалознавства використовуються алгоритми та принципи автоматичної обробки подібних даних, з мінімальним управлінням людиною. Подібні системи уже на даний час дозволяють розраховувати різні параметри (твердість, модуль пружності, тензор пружності) та вдало досліджувати їх у подальшому [1-2,4,7-8].

Штучний інтелект поширює свій вплив майже на усі галузі життя людини, у тому числі і науку. Завдяки відсутності людських факторів та застосуванню ітераційних методів, точність розрахунків та вдалість використання розрахункових рішень зростає в декілька десятків разів, на відміну від попередніх рішень, які мають більшу поширеність, але менший об'єм.

Розвиток матеріалознавства на сьогодні досить сильно залежить від автоматизованих розрахунків, які завдяки готовим алгоритмам або завдяки штучному інтелекту, а також моделюванню його характеристик. На сьогодні існує досить багато існуючих рішень для моделювання параметрів різних типів та різних структур (як приклад quantum espresso). Завдяки подібним рішенням вже на сьогодні можна промоделювати нові сплави різної комбінації та виявити їх окремі риси, які можна застосовувати у матеріалознавстві, а також проводити над отриманими даними подальший аналіз, який як поглибить існуючу інформацію про різні сплави, так і утворити нові.[13]

Для вдалого і більш об'ємного застосування різного типу алгоритмів, бажано реалізувати методи машинного навчання для штучного інтелекту. Завдяки таким рішенням стає можливим забезпечення більш гнучкого аналіз і більш об'ємного

використання системи для розрахунку параметричних даних матеріалознавства та інших існуючих систем. Машинне навчання можна вважати досить важливим пунктом створення штучного інтелекту та його розвитку.

У цій роботі першою задачею стоїть вивчення проблематики подібних систем, вивчення концепцій машинного навчання та автоматизації штучного інтелекту, дослідження вже існуючих алгоритмів штучного інтелекту, можливості застосування, а також обмеження застосування в області матеріалознавства. При дослідженні буде виявлено найбільш точні для нашої задачі алгоритми та методи. Робота буде зосереджена саме на методах машинного навчання, створена з синергії існуючих розрахункових модулів, а також модулів для вилучення даних (matminer + rumatgen), та їх подальший розрахунок даних, таких як модуль пружності та твердість для тернарних сплавів, а також подальше спрощення аналізу нових, або існуючих сплавів. Це дозволить підвищити гнучкість створення та вивчення матеріалів.

Застосування подібних систем дозволить спростити, здешевити вивчення різних типів сплавів, прискорити процеси розрахунку параметрів матеріалів, а також створювати нові бази даних, які у подальшому також можуть надавати дані для обробки іншими алгоритмами штучного інтелекту або для навчання нових систем, за допомогою базисів машинного навчання. В різних галузях можливе використання подібних систем та їх покращення. На відміну від практичних та експериментальних рішень, автоматичні рішення є більш точними та надають можливість отримати більший об'єм даних завдяки багаторазовим ітераційним процесам. Інакше кажучи це дозволить у майбутньому набагато підвищити точність на відміну від експериментальних методів [8].

Метою дослідження у моїй роботі будуть алгоритми штучного інтелекту, його залежне машинне навчання, а також різні можливості застосування розроблених алгоритмів, а саме базові існуючі методи регресії, нейронні мережі а також додаткове дослідження алгоритмів «випадкових лісів», які застосовуються у роботі як основний алгоритм аналізування бінарних та тернарних матеріалів. Буде проведений аналіз обраного методу, та перевірена його точність та релевантність для обраного обраного предмету нашого дослідження. Метою також є порівняння базових методів

дослідження та аналіз методів, на основі яких буде представлений висновок, чи релевантний подібний метод, та чи можливо взагалі використовувати подібні засоби у матеріалознавстві, для дослідження параметрів матеріалів, Фактично можливість застосування розглянутих нижче методів буде об'єктом мого дослідження, тоді як предмет дослідження буде обраний як приклад. Прикладом було обрано аналіз модуля пружності (властивості матеріалу до деформації без втрати властивостей) а також твердість матеріалу (супротив до деформування / руйнування під впливом зовнішніх чинників (контакту)), порівнюючи з методами перевірки на твердість.

1. ЛІТЕРАТУРНИЙ ОГЛЯД

1.1 Актуальність розвитку

На даному етапі існування науки нові матеріали можуть дозволити розробляти нові технологічні програми для подолання різних проблем. Вплив нових відкриттів у матеріалознавстві протягом історії людства, починаючи кам'яною добою и закінчуючи сучасністю, в якій використовуються кремнієві технології, важко піддавати кількісній та якісній оцінці. На сьогодні при підрахунку було виявлено, що розробка матеріалів дала більше двох третин усіх успіхів в галузі обчислень, а також почала розвивати інші існуючі напрямки, такі як накопичення енергії, машинобудування, наплення та їм подібні. Це все можливе лише при впровадженні нових типів матеріалів, які мають неймовірні характеристики. Яскравим прикладом на сьогодні є кремній, якій є фундаментальним матеріалом у процесоробудуванні, машинобудуванні та інших галузях, які використовують розрахункові методи. Загальний час на впровадження іноваційних технологій в будь-який научний сектор займає 15-20 років, або взагалі більше часу, що є досить вагомою проблемою для просування наукових досягнень в будь-якій галузі.

Враховуючи до уваги, що попит на нові сплави стрімко зростає разом з потребами, відносно повільно розвивається розвиток необхідних сплавів, коли як в секторах обчислювальних методів, алгоритмізації та технічних вдосконалень, спостерігаються стрімкі вдосконалення. В галузях біотехніки, енергоефективності, архітектури подібні питання мали певний розгляд.

Умовно розвиток різних галузей, в тому числі і матеріалознавства, враховуючи використання штучного інтелекту, моделювання і взагалі використання різних машинних засобів, можна поділити на 4 типи, які відрізняються підходами у різні роки:

1. Емпірична наука (до 1600р.) – висновки створюються в загальному випадку через експериментальне дослідження.
2. Теоретична наука (1600-1960рр.) – дослідження механічних, термодинамічних та загальних властивостей, враховуючи теоретичні розрахунки та

порівняння з експериментальними даними, для виявлення закономірностей.

3. Комп'ютерна наука (симулювання) (1960-1990pp) – моделювання різних динамічних структур та висновки на основі моделювання.
4. Big data science (прогнозування) (з 1990pp) – прогнозування властивостей матеріалів, їх кластеризація та утворення баз даних для аналізу.

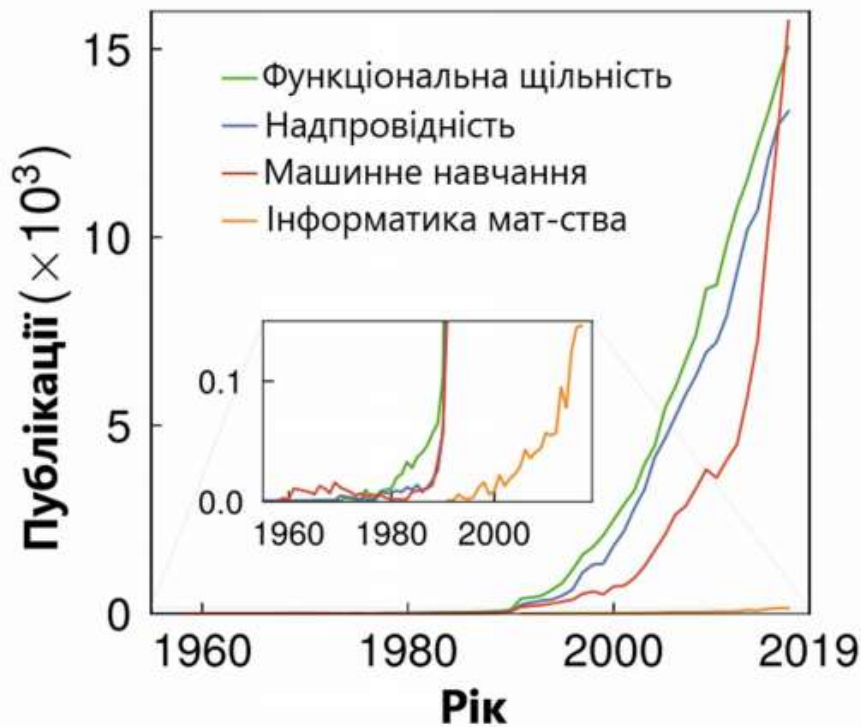


Рис 1.1 – Кількість публікацій по кожній галузі, які застосовуються в матеріалознавстві (на меншій діаграмі – загальна зацікавленість та кількість публікацій, які не пов'язані з дослідженням матеріалознавства) [1]

Вивчення проблематики прогнозування з 2000х років тільки отримало певний розвиток у галузі матеріалознавства. Затримка рішення проблематики пов'язана з малими можливостями механічно-розрахункової частини, що затримало розвиток, але, перші спроби фізичного моделювання матеріалів невеликої складності. На рис.1.1 можна побачити залежність між роками та дослідженнями окремих пов'язаних тем, що доказує цю ситуацію. На даному етапі зацікавленість в цих темах саме у області матеріалознавства має досить великий коефіцієнт, так як вже є певні використовувані рішення для аналізування, які використовуються на сьогодні. [1]

Традиційні підходи теоретичних, обчислювальних та експериментальних видів дослідження матеріалів, які можна узагальнено назвати прямим підходом, застосовують розрахункові властивості, враховуючи структурні данні та дані матеріалів. Досить складним процесом в даній ситуації є пошук більш підходящих матеріалів, в міру затраченого часу як на кожний процес, так і на утворення певного масиву матеріалів (певної бази даних). Можливо застосовувати ті ж самі отриманні завдяки традиційним даним підходи до матеріалів, але на сьогодні це все ще досить складний процес. Тут відіграють роль декілька факторів. По-перше, на сьогоднішній день існує не така велика кількість баз даних, які мають достатню інформацію про матеріали, або кажучи інакше, мають недостатню для прогнозування/аналізу, кількість даних. По-друге, це недосконалість як самої структури баз даних, так і методів аналізу. Частково це можливо виправити завдяки використанню штучного інтелекту, який буде користуватися принципами самонавчання. Використання подібних технологій та підходів на сьогодні дозволяє завдяки обчислювальним наукам керувати, звузити та оптимізувати пошук та аналіз інформації, у тому числі і в матеріалознавстві. [1]

Протягом досліджень утворилась проблематика генерації даних. Масштабне їх генерування не є гарантією її перетворення або аналізу, що може сильніше ускладнити вказані процеси. На рис 1.2. можна побачити пропорційну збережену/згенеровану інформацію та отримані корисні дані, та їх корисність. Тобто, конвертація згенерованих даних в корисну інформацію, а потім і знання є проблемою, з якою можна боротися завдяки тому-ж самому розвитку самонавчання системи. Фактично, система повинна вміти змінювати пріоритети та фільтрувати інформацію, для її вдалого застосування. Саме такою проблематикою займаються з 2000х років галузь інформаційних технологій. Суть полягає в тому, що інформація може бути невірною, не підходящою для певної галузі, що в свою чергу впливає на якість розрахунків, аналізу при наприклад програмуванні певної програми, яка застосовує ці дані. На сьогодні розвиток алгоритмізації в тих сферах досягнув того, що розрив між отриманою и корисною інформацією почав рухатися к мінімізації, що дозволило отримати більш чіткі результати. Виникає логічне твердження що ці ж самі методи можливо застосовувати у матеріалознавстві.

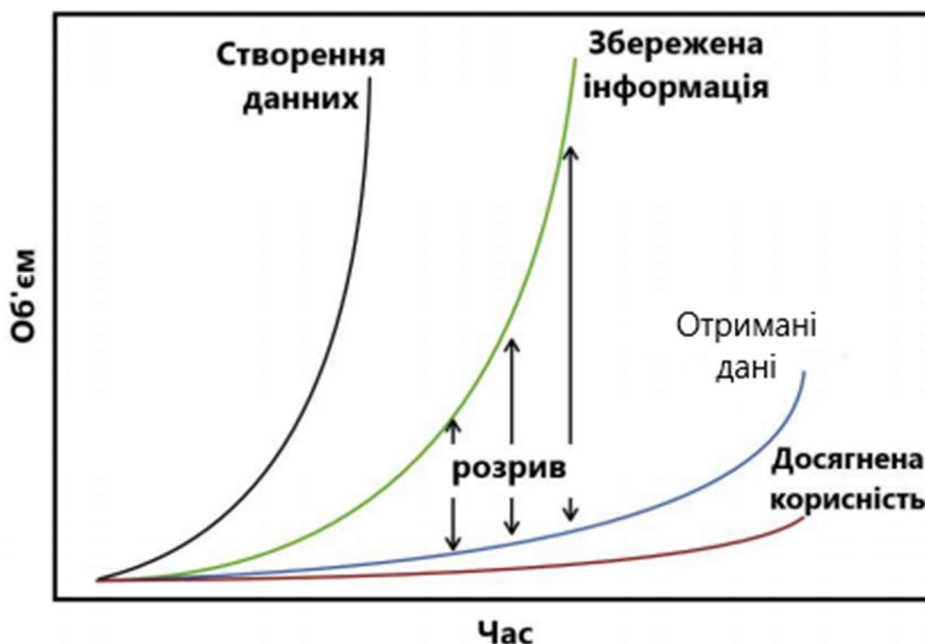


Рис 1.2 – Показаний теоретичний розрив між отриманими даними, використаною інформацією, витягнутими з неї висновками та досягнутою корисністю [1].

1.2 Основні принципи машинного навчання

Суть машинного навчання залежить від певних признаков, на основі яких будується модель, яка підходить до вирішення певної правильно поставленої задачі. У загальному випадку для машинного навчання існує три основних поняття [33], а саме features (особливості / ознаки), task (задача) та model (модель).

Ознаки / особливості фактично являють собою методи підстави або створення інформації для обробки або фільтрації, в залежності від поставленої задачі. Ознаки можуть варіюватися в великому діапазоні в залежності від подальшого їх застосування. Маючи певне уявлення про ознаки, які в машинному навчанні можна назвати «мовою» за якою описуються об'єкти з якими в подальшому будемо взаємодіяти, можуть дозволити не приділяти увагу об'єктам, так як їх інформація буде вичерпною в ознаках

Задача – проблема, яку фактично повинен вирішити штучний інтелект, який є базисом машинного навчання. Це може бути як певна абстрактна проблема, так і більш конкретна, як у нашому випадку, набір даних та властивостей сплавів. Для роботи з задачею потрібно провести класифікацію об'єктів, а саме віднести кожний

об'єкт до певних одного або декількох класів. Ще існує пояснення для задачі, як відображення вихідних даних, або інакше кажучи, моделі [33]. Модель – фактично є результатом, який утворюється при обробці методами машинного навчання. Модель вважається центром машинного навчання, так як вона породжується в результаті навчання на основі оброблених даних.

Алгоритми машинного навчання мають окремий напрямок на оптимізацію виконання певного завдання за допомогою прикладів та/або отриманого попереднього досвіду. Тобто машинне навчання повинно виконувати функції покращення самого алгоритму штучного інтелекту. Види машинного навчання умовно можна розділити на чотири основні категорії, а саме: кероване навчання, некероване навчання, трансферне навчання та частково кероване навчання. [1,5,6].

Контрольоване машинне навчання базується на принципах підгонки, тобто намагається знайти невідому функцію, яка з'єднує відомі входи з невідомими виходами. При контрольованому машинному навчанні присутні дані та «мітки». Алгоритм пробує підбирати шаблони, які зводять дані з мітками. Контрольована частина зароджується під час прогонки. Якщо алгоритм приводить до неправильної мітки, він намагається або змінити сам алгоритм, або підібрати інший шаблон, який приводить до правильного результату.

Відмінність контрольованого машинного навчання від неконтрольованого у тому, що якщо в першому присутні мітки (точки до яких алгоритм повинен в результаті привести), то у другому точного шляху немає. Тобто неконтрольоване навчання приводить до створення умовних кластерів, на основі яких утворюються мітки. Складність у початковому навчанні, так-як немає точного направлення у навчанні [5,9].

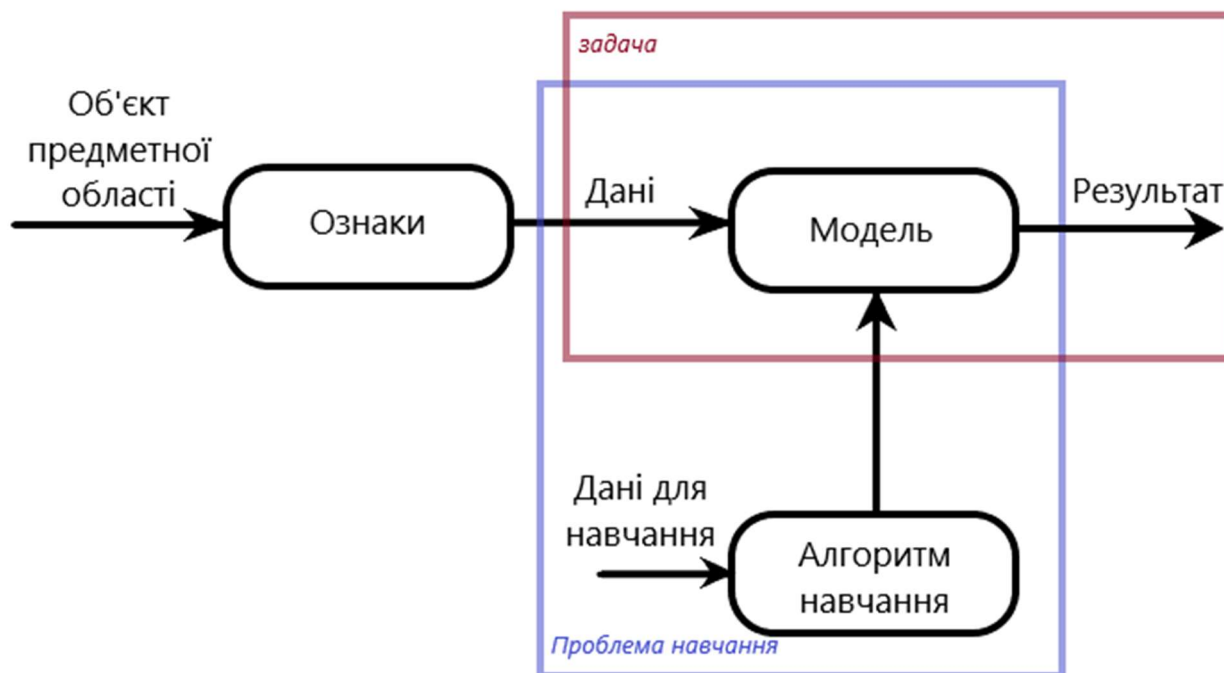


Рис 1.3 – Приклад того, як саме машинне навчання застосовується для рішення задачі. Задача має потребу в підходящому відображенні - моделі - даних, що описуються ознаками, на результати. Отримання такого відображення з даних для навчання і є предметом навчання. [33]

Поєднанням контрольованого та не контрольованого навчання можна назвати напівконтрольоване навчання. У цьому випадку алгоритм забезпечується як не міченими, так і міченими даними. Методи цієї категорії особливо корисні, коли наявні дані неповні та для вивчення уявлень. Слід також відокремити трансферне навчання. Коли наробітки та інформацію яку вже відпрацювала модель машинного навчання, використовуються для інших задач, або для навчання іншої моделі [9].

На рис 1.4 зображений приклад робочого процесу у контрольованому навчанні. Як правило, обирають підмножину відповідної сукупності, для якої відомі значення цільового властивості, або створюють дані, якщо це необхідно. Цей процес супроводжується вибором алгоритму машинного навчання, який буде використовуватися для встановлення бажаної цільової кількості. Більша частина роботи полягає в генеруванні, пошуку та очищенні даних для забезпечення їх послідовності, точності тощо. Також необхідно вирішити, як відобразити властивості системи, тобто вхід для

моделі, в спосіб, відповідний обраному алгоритму. Це означає переведення необробленої інформації в певні функції, які будуть використовуватися як вхідні дані для алгоритму. Після закінчення цього процесу модель навчається шляхом оптимізації її продуктивності, зазвичай вимірюється через якусь функцію витрат. Зазвичай це тягне за собою коригування гіперпараметрів, які контролюють тренувальний процес, структуру та властивості моделі. Дані розділені на різні набори. В ідеалі для оптимізації гіперпараметрів використовується набір даних перевірки, окремо від тестових та навчальних наборів.[9]

Кожен додаток для машинного навчання повинен враховувати аспекти подолання та недосконалості. Причина недосконалення зазвичай криється або в моделі, в якій бракує можливості виразити складність даних, або в особливостях, які не належним чином описують дані. Це неминуче призводить до високої помилки тренувань. З іншого боку, переоснащена модель інтерпретує частину шуму у навчальних даних як відповідну інформацію, тому не в змозі достовірно передбачити нові дані. Зазвичай, переобладнана модель містить більше вільних параметрів, ніж кількість, необхідна для збору складності навчальних даних. Щоб уникнути перевитрат, важливо стежити не лише за похибками навчання, але й за похибками N кількості перевірок. Як тільки помилка перевірки перестане зменшуватися, модель машинного навчання може почати переоцінювати. Ця проблема також обговорюється як компенсація дисперсії упередженості в машинному навчанні. У цьому контексті зміщення - це помилка, заснована на помилкових припущеннях тренувальної моделі, тоді як велика дисперсія - це помилка, що виникає внаслідок занадто великої чутливості до шуму в навчальних даних. [12]

Як такі, недоопрацьовані моделі мають високу упередженість, тоді як переобладнані моделі мають велику дисперсію [19]. Перед тим, як модель буде готова до застосувань, її потрібно оцінити за раніше невідомими даними, позначеними як тестовий набір, для оцінки її узагальнення та здатності до екстраполяції. [9]

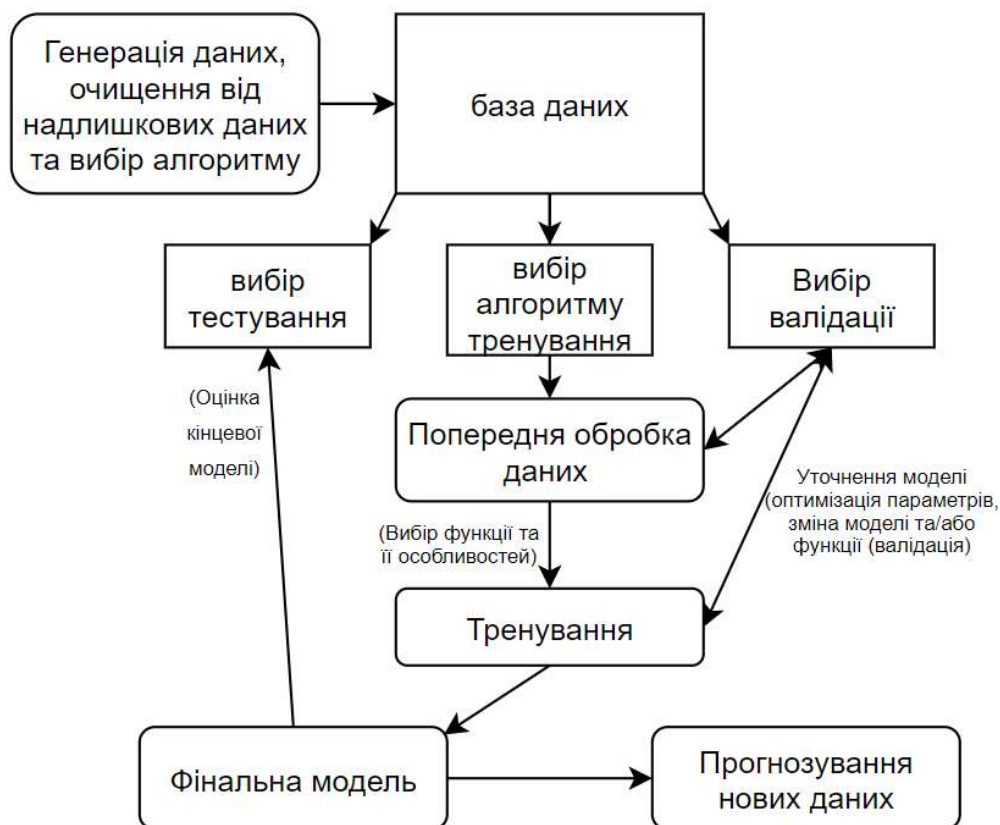


Рис 1.4 – Блок схема, приклад роботи машинного навчання при інтеграції з штучним інтелектом. На схемі вказані основні функції штучного навчання, такі як перевірка даних, вибір алгоритмів тестування, тренування та валідації, які фактично і відіграють роль «навчання».

Основною задачею алгоритму машинного навчання в нашому випадку є представлення отриманих даних у відповідній для задачі формі. Враховуючи особливості даних матеріалознавства, система повинна бути здатною фіксувати всю відповідну інформацію, необхідну для розмежування між різними атомними та кристалічними структурами та середовищами. Сам процес, який може бути наприклад видобутком чи іншою інженерною функцією, може бути як простим, наприклад знаходження атомного числа складових матеріалу / сплаву, але і він може включати в себе складні перетворення даних, наприклад розширені функції радіального розподілу (DFT – теорія функціонала густини [1]), або інші статистичні алгоритми, які можливо застосувати для певної задачі. Залежно від самого алгоритму машинного навчання та штучного інтелекту загалом, кількість ітерацій навчання повністю залежить від

застосованих нами методів. Для методів які, потребують більшого та більш глибокого аналізу, сказане вище може розглядатися як частина від загальної моделі, що вказує на складність базових алгоритмів. Найкращий вибір алгоритмів та методів в нашому випадку буде залежати від різноманітності простору подій (алгоритм повинен бути гнучким до різного типу даних) та цільової кількості даних (тобто алгоритм має можливість оброблювати як можна більший масив даних). Цей напрям буде впливати на якість та швидкість машинного навчання а також на якість виконаної задачі до проведення певних ітераційних методів та перевірок даних.

Дескриптори в ідеальному значенні повинні бути відокремленими один від одного (неспорідненими) оскільки велика кількість співвіднесених ознак може перешкоджати ефективності та точності моделі. Коли це станеться, необхідний подальший вибір функцій, щоб обійти проблему розмірності, спростити моделі та покращити їх інтерпретацію, а також ефективність навчання. Наприклад, кілька елементарних властивостей, таких як період і група в періодичній таблиці, потенціал іонізації та ковалентний радіус, можна використовувати як особливості для моделювання міцності та твердості матеріалу або різних коефіцієнтів його стійкості. Однак було показано, що для отримання прийнятної точності в багатьох випадках потрібен лише період та група, за якими можливо узяти властивості матеріалів з баз даних.

Описавши загальні властивості дескрипторів, ми перейдемо до переліку найбільш використовуваних ознак в матеріалознавстві. Без сумніву, найбільш вивченим типом особливостей у цій галузі є ті, що стосуються пристосування потенційних енергетичних поверхонь. Заряд ядра і атомні позиції є достатніми ознаками, оскільки гамільтоніан системи зазвичай повністю визначається цими величинами. На практиці, хоча декартові координати можуть давати однозначну характеристику атомних позицій, вони не роблять відповідного дескриптора, оскільки список координат структури упорядкований довільно, а кількість таких координат змінюється залежно від кількості атомів. Останнє є проблемою, оскільки для більшості моделей машинного навчання потрібна фіксована кількість функцій як вхідних даних. Тому для опису твердих тіл та великих скупчень кількість взаємодіючих сусідів має бути дозволена

змінюватись, не змінюючи розмірність дескриптора. Крім того, багато додатків вимагають, щоб функції були постійними та диференційованими щодо атомних позицій

Перш ніж обговорити різні застосування машинного навчання в матеріалознавстві, потрібно надати огляд різних дескрипторів, алгоритмів та баз даних, що використовуються в інформатиці матеріалів. Також важливою задачею було-б обговорити точність та якість машинного навчання та самого алгоритму. [9]

1.3 Оцінка якості виконаної задачі

Акцентуючи увагу на проблемах машинного навчання потрібно розуміти, що досконало точного або правильного рішення отримати неможливо. Цим машинне навчання сильно відрізняється від інших галузей, які займаються інформатикою, або пов'язаних наук. Як приклад, для порівняння, можна взяти сортування номерів телефону в алфавітному порядку, де можливий один правильний результат. Так, це не є ознакою того, що є єдиний точний алгоритм досягнення одного результату. Можливе використання різних підходів до прикладу вище, наприклад сортування вставками, шаблонування, певні швидкі алгоритми сортування та інше. При порівнянні якості роботи цих алгоритмів ми брали б до уваги їх швидкість і максимальний обсяг оброблюваних даних. Наприклад, можна було б експериментально перевірити алгоритм на реальних даних або проаналізувати його методами теорії обчислювальної складності. Однак ми не враховували б при порівнянні правильність результату, тому що алгоритм, який іноді впорядковує дані, а іноді ні, взагалі не потрібний у якості алгоритму сортування.

В машинному навчанні є діаметрально обернена ситуація. Ми не можемо створити ідеальні фільтри, наприклад для сортування спаму на електронній пошті, або у нашому випадку ми не можемо досить просто знайти наприклад ідеальний матеріал для твердості і в водночас для модуля пружності та ширини забороненої зони. Для цього потрібно використовувати великий об'єм алгоритмів, які повинні модифікуватися під кожний із матеріалів з своїми окремими особливостями. З генерацією нових матеріалів фактично потрібно створювати новий підхід до аналізу матеріалу або нового сплаву, який раніше ніде не моделювався. Використовувати підходи до подібних

матеріалів часто не правильно, через ту ж саму особливість кожного з матеріалів, їх кристалічної будови та інших параметрів (матеріал → кристалічна решітка → провідність). При генерації та подальшій обробці інформації, утворюються великий масив цих даних, який складно піддається обробці і майже зовсім не піддається фільтрації.

В цій роботі, для перевірки масиву даних в буде використовуватися декілька видів перевірки, а саме кросс-валідація, (перехрестна перевірка), та коефіцієнт детермінації. Вони необхідні для того щоб перевірити використовуваний метод на точність використання, що в нашому випадку відіграє ключову роль.

Наприклад, візьмемо за основу прогнозування твердості матеріалу та виявимо наскільки точне виявлення правильних даних при порівнянні. По перше ми одразу можемо виявити так звану ймовірність. Вона розраховується простим чином. Усі отримані правильні дані розділяються на загальний масив даних. Такий результат може вказати на правильно розраховану ймовірність правильного результату.

У статистичному аналізі існує інший метод перевірки правильності машинного навчання. Наприклад у нас є масив даних, який генерується після аналізування. Ми можемо взяти 90% цих даних для навчання, а інші 10% як певний тестовий набір. Якщо перенавчання системи дійсно присутнє, то результати самої першої перевірки будуть кардинально відрізнятися від так званої «навчальної перевірки» (10% даних). Так, це може бути прикладом навчання, але немає повних гарантій того, що з даними може просто пощастити і вони будуть в останній більш точними при обробці. Також слід зазначити що такі дані фактично будуть мати у собі мінімум змістовного наповнення з якого неможливо взяти певний висновок про навчання машинного алгоритму. Але ми можемо провести декілька подібних ітерацій, тобто повторити цей процес, при цьому змінивши сектор «навчальної перевірки».

Цей процес називається перехресною перевіркою. Ми розбиваємо наші отримані дані на певні окремі сектори, які рівні по кількості даних. З них усі групи крім одної відіграють роль навчання, одна для тестування, як вказувалося вище. Подібне навчання проводиться декілька разів зі зміною навчальної групи. В кінці ми вираховуємо середнє значення якості роботи при тестовому наборі (тобто стандартне відхилення, так як воно вказує на те, чи є значимими параметри)

У цій роботі, для перевірки масиву даних в буде використовуватися декілька видів перевірки, а саме кросс-валідація, (перехрестна перевірка), та коефіцієнт детермінації. Вони необхідні для того щоб перевірити використовуваний метод на точність використання, що в нашому випадку відіграє ключову роль.

1.4 Принципи взаємодії з базами даних

За останній час, використовуючи принципи та структури кодування, було вивчено великий масив даних, велику кількість відомих матеріалів, які с кожним роком експоненційно збільшуються на час об'єму розрахункових даних. Цілю вивчення та створення баз даних є подальше спрощення вивчення матеріалів та їх фізичні та хімічні властивості, наприклад ширина забороненої зони, провідність, та загальні фізичні властивості нових сплавів. Слід зазначити той факт, що використовувати дані з подібних баз даних в їх первозданному вигляді неможливо [4,11].

Машинне навчання в матеріалознавстві здебільшого стосується наглядного навчання. Успіх методів залежить головним чином від кількості та якості доступних даних. Це особливо актуально для конкретних властивостей, які можна визначити лише експериментально дорогим способом З цієї причини бази даних, такі як material project, база даних неорганічної кристалічної структури та інші, які містять інформацію про численні властивості відомих матеріалів, є важливими для успіх матеріалів інформатики [6,11].

Для того, щоб ці бази даних та матеріали для інформатики були відповідними необхідна справедлива обробка даних. Ця процедура включає чотири принципи: гнучкість, доступність, сумісність та відновлюваність. Іншими словами, в незалежності від того, використовує подібні дані вчений, або сама програма штучного інтелекту, усі повинні мати можливість знаходити та отримувати доступ до даних, а також відповідних метаданих, у загальноприйнятому форматі. Це дозволяє застосувати дані для нових цілей.

Негативні результати часто відкидаються від загальної вибірки. але отримані в процесі обробки алгоритмом негативні дані в загальному випадку такі ж важливі для алгоритмів машинного навчання, як і позитивні результати, необхідно по перше

враховувати той факт, що негативний результат та негативні результати – теж результати з яких можна взяти окремий висновок, по друге, ці результати можуть завдяки машинному навчанні допомогти не отримувати у подальшому схожі результати. Усі дані, як позитивні так і негативні у нашому випадку являють собою складові баз даних. У деяких дисциплінах з більш точнішими і тривалими за часом дослідженнями, заснованих на даних (наприклад, хімія) такі бази даних вже існують та використовуються досить давно. У подібному руслі дані, що з'являються як побічний продукт, але не є істотними для публікації, часто залишаються не опублікованими. Це врешті-решт призводить до марної трати ресурсів, оскільки від інших дослідників вимагається повторити роботу. Врешті-решт, кожен викинутий розрахунок буде пропущений у майбутніх алгоритмах машинного навчання [4,11].

Пошук та генерування матеріалів можна розглядати як основну частину робочого процесу штучного інтелекту, так і окремий процес, який можливо використовувати і без інтеграції в штучний інтелект, фактично дозволяючи його використовувати як певний апарат пошуку.

Бази даних як правило повинні перевірятися послідовними підходами, щоб задовільнити кожне дане обмеження, переходили до наступного пункту обробки чи зберігання, тоді як інші – обходять матеріал.

1.5 Приклади поширених алгоритмів

Почнемо з лінійних та ядерних методів регресії та класифікації. Потім ми введемо алгоритми вибору та видобутку змінних, які також значною мірою базуються на лінійних методах. Щодо повністю нелінійних моделей, ми обговорюємо методи, засновані на дереві рішень, як «random forest» та надзвичайно «рандомізовані дерева» та нейронні мережі [6,10-12,19].

Найбільш поширеною з усіх типів алгоритмів, які взагалі використовувалися з самого початку у багатьох галузях фізики, це лінійна регресія. Її підвидів існує досить багато і усі вони фактично базуються на лінійній регресії. В багатьох галузях науки необхідно досліджувати декілька змінних, та їх залежність одна від одної. Прикладом

подібної залежності можна представити закон Ома, у якому вказується залежність сили струму від напруги.

Слід наголосити що просту лінійну регресію в нашому випадку використовувати неможливо через те, що даних які залежать один від одного – досить багато, що призводить до великої середньоквадратичної помилки, що у нашому випадку, при аналізуванні та прогнозуванні бінарних, тернарних та високоентропійних матеріалів, що напряду впливає на якість прогнозування.

В загальному випадку, звичайним методом аналізу та оцінки параметрів є метод найменших квадратів (МНК). Завдяки обширним сферам застосування як самої лінійної регресії так і методу найменших квадратів, він посідає одне з самих головних місць в задачах математичної статистики, які ми у подальшому використаємо у своєму алгоритмі.

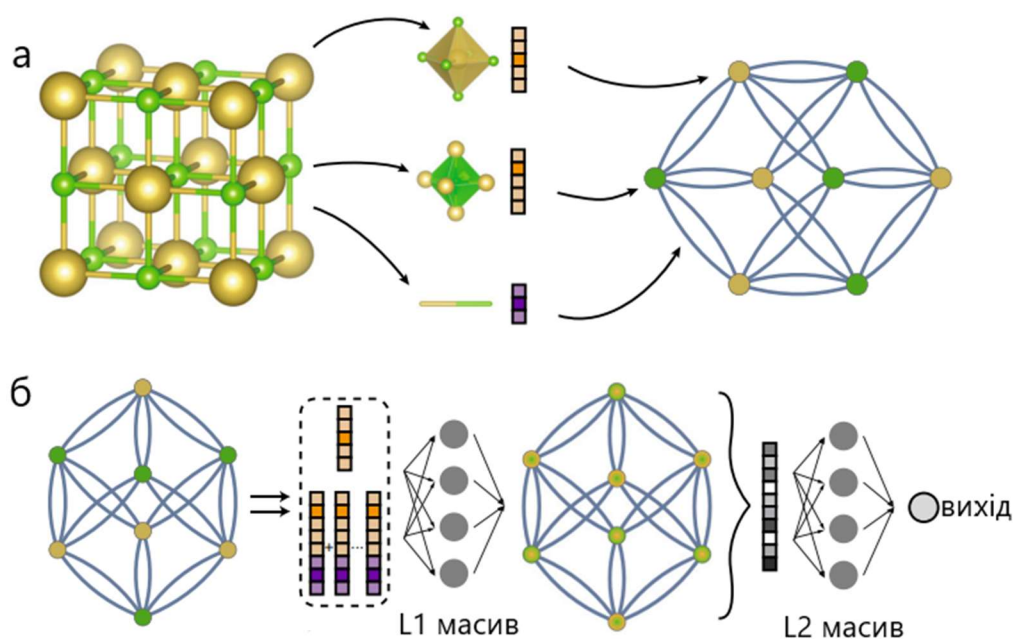


Рис 1.5 – Приклад нейронної мережі як допоміжної моделі обробки даних (перетворення кристалічної структури в граф-масив) а – перебудова кристалічної решітки як граф. б – структуризація графу та його переробка. Зображений як приклад застосування алгоритмів різного порядку та напрямку.[9]

Ридж-регресія (гребенева регресія) – один з методів метод зниження розмірності, який має своє застосування у тому, щоб виключати надлишкові дані з вибірки,

коли дані між собою корелюють (тобто має місце мільтиколінеарність). В наслідок застосування подібного типу регресії, ми можемо з цього отримати набагато менше значення середньої похибки, за умови, якщо дані мають статистичний характер [14].

Для багатовимірної ридж-регресії основною задачею для вирішення є задача лінійного розміру з методом найменших квадратів, включаючи також термін регуляції L_2 (рис 1.5) який має таке рішення:

$$\min_x \|Ax - b\|_2^2 + \lambda \|x\|_2^2 \quad (1.1)$$

Термін додаткової регуляризації включений для переваги конкретних рішень з набагато меншими коефіцієнтами.

Складні проблеми які виникають у регресійних методах зазвичай не можуть бути вирішені простою лінійною моделлю, тому у цей метод вводять нове поняття основи регресії, або як її називають по іншому, ядро регресії. Замість використання оригінального дескриптора x дані спочатку перетворюються у простір більш високих розмірів $\phi(x)$. У цьому просторі ядро $k(x, y)$ дорівнює внутрішньому добутку $\phi(x), \phi(y)$.

На практиці потрібно оцінювати лише основу (ядро), уникаючи неефективного або навіть неможливого явного обчислення функцій у новому просторі.

$$k_G(x, y) = e^{-\frac{|x-y|^2}{2\sigma^2}} \quad (1.2)$$

або поліном ядра

$$k_P(x, y) = (x^T y + c)^d \quad (1.3)$$

Для розв'язку задач мінімізації заданими рівняннями по відношенню до нових функцій, дії будуть призводити до утворення нелінійних регресій при вихідних результатах функцій. Подібні вихідні функції мають назву ядрова ридж-регресія (ЯРР).

Ядрова ридж-регресія, як правило, проста у використанні, оскільки для успішного застосування гребневої регресії потрібно вказати відносно невелику кількість існуючих гіперпараметрів. Отже, ядрова гребнева регресія зачасту використовується в матеріалознавстві для аналізу властивостей пов'язаних з міцністю, надпровідністю, та іншими складними параметрами.

Метод опорних векторів відноситься до групи граничних методів, які визначають класи за допомогою границь певних областей. У методі опорних векторів (МОВ), для певного вектору шукають гіперплани, які ділять набір даних на класи таким чином, щоб межа навколо гіперплана була максимальною (див.рис. 1.6). [36].

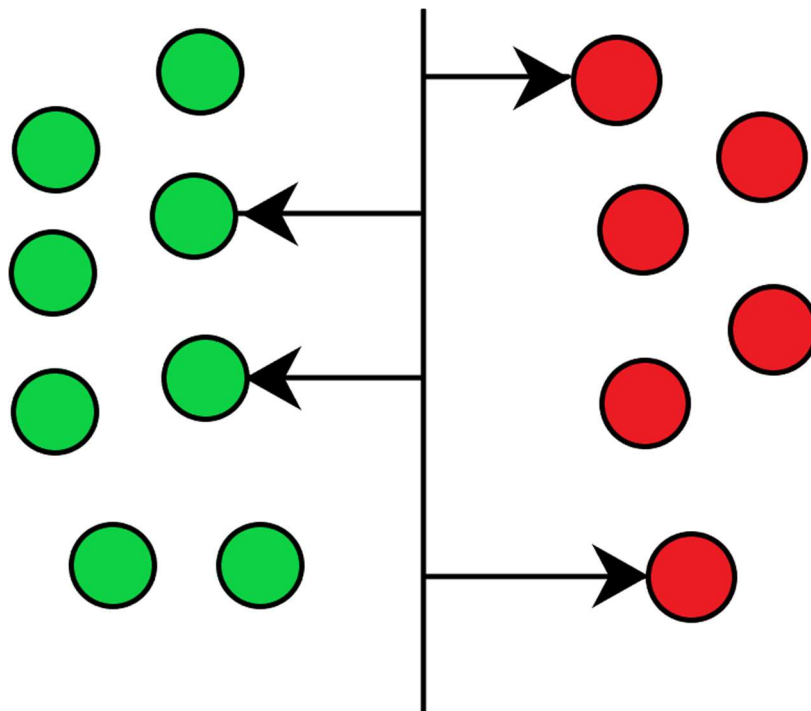


Рис 1.6 - Граніфікація межі підтримуючої векторної машини з «вектори підтримки», показані у вигляді стрілок. Чорною стрілкою вказана гіперплощина, яка є оптимальним розділенням дані [9]

Гіперплощина повністю визначається точками даних, що лежать найближче до площини, тобто векторами опори, з яких алгоритм походить свою назву. Аналогічно регресії хребта, фокус ядра може бути використаний для досягнення нелінійних МОВ. Регресори МОВ також створюють лінійну модель (нелінійну у випадку лінійної регресії з основою-ядром), але використовують так звану ϵ -нечутливу функцію втрат:

$$\text{Втрати} = \begin{cases} 0 & (\text{if } \varepsilon > |y - f(x)| \\ |y - f(x)| - \varepsilon & \text{для інших випадків} \end{cases} \quad (1.4)$$

де $f(x)$ - лінійна модель, а ε - гіперпараметр.

Якщо порівнювати МОВ та гребневу регресію ядра, великі відмінності в продуктивності не очікуються. Зазвичай МОВ отримують більш рідке представлення, що може бути перевагою; однак їх ефективність покладається на вдалу обстановку гіперпараметрів. У більшості випадків МОВ надають швидші прогнози та споживають менше пам'яті, тоді як методу гребневої регресії ядра знадобиться менше часу, щоб підходити до середніх наборів даних. Тим не менш, внаслідок загально низької обчислювальної вартості обох алгоритмів, ці відмінності рідко важливі для відносно невеликих наборів даних. На жаль, жоден метод неможливий для великих наборів даних, оскільки розмір матриці ядра квадратично масштабується з кількістю точок даних.

Регресія гауссовського процесу (РГП) спирається [9] на припущення, що дані тренінгу були сформовані за Гауссовим процесом і тому складаються з зразків багатовимірного гауссового розподілу. Єдине інше припущення, яке вводиться в регресію, - це форми коваріаційної функції $k(x, x')$ та середнє значення (яке часто вважається рівним нулю). На основі матриці коваріації, елементи якої представляють коваріацію між двома ознаками, можна передбачити середнє значення та дисперсію для кожного можливого значення функції. Здатність оцінювати дисперсію є головною перевагою РГП, оскільки невизначеність прогнозу може бути важливим компонентом процесу проектування матеріалів. РГП також використовує ядро для визначення функції коваріації.

На відміну від методу ридж-регресії, де гіперпараметри були оптимізовані за допомогою зовнішнього набору перевірки, гіперпараметри в РГП можуть бути оптимізовані за нахилом градієнта, якщо обчислені матриці коваріації та її оберненої обчислювальної можливості. Порівнюючи вказаний вище метод з звичайними методами ридж-регресії, у якому проводилася оптимізація гіперпараметрів для його даних,

можна вказати на непогані значення похибки, але цей метод досить важко використовувати та оптимізувати для машинного навчання

Хоча існують сучасні та швидкі впровадження гауссових процесів у матеріалознавстві (наприклад, COMBO [19]), їх притаманне масштабування є досить обмежувачим щодо розміру даних та розміру дескриптора, як наївне навчання вимагає інверсії коваріаційної матриці порядку $O(N^3)$ і навіть масштаби прогнозування з $O(N^2)$ щодо розміру набору даних. Виходячи з принципів РГП, можна також створити класифікатор. По-перше, РГП використовується для якісної оцінки ймовірності класифікації. Тоді до латентної функції застосовується сигмоїдна функція, що призводить до значень в інтервалі від нуля до одиниці.

У даному випадку слід зазначити той факт, що ми розглядаємо ті ситуації, коли вибір функцій попередньо відомий. Є також варіанти що вибір функції може бути досить складним. Через що виникає необхідність у виборі більш цей вибір може бути досить складним, методи вибору особливостей можуть мати суттєве значення.

У статистиці та машинному навчанні LASSO [13] (найменший абсолютний оператор усадки та відбору) - це метод регресійного аналізу, який виконує як вибір змін, так і регуляризацію з метою підвищення точності прогнозування та інтерпретації статистичної моделі, яку він виробляє. Спочатку він був введений в літературу з геофізики, а згодом незалежно відкрив і популяризував у 1996 році Робертом Тібширані, який придумав цей термін і надав подальші уявлення про спостережувані результати.

LASSO намагається покращити ефективність регресії шляхом створення розріджених моделей за допомогою змінного вибору. Він здебільшого використовується в поєднанні з лінійною регресією найменших квадратів, і в цьому випадку це призводить до наступної проблеми мінімізації. [5,13-16]

$$\min_{\beta, \beta_0} \sum_i (y_i - \beta_0 - \beta x_i)^2 \text{ за умови } \sum_i |\beta_i| < t \quad (1.5)$$

де y_i - результати, x_i - ознаки та β - коефіцієнти лінійної моделі, які необхідно визначити. На відміну від ридж-регресії, з використанням L_2 -норма терміну регуляризації,

LASSO алгоритм спрямований на перетворення більшості його коефіцієнтів до нульового значення. Для того, щоб знайти реальну модель з мінімальною кількістю ненульових компонентів, замість L_1 -норми, що використовується в прикладі з LASSO, слід використовувати так звану L_0 -норму коефіцієнта. (Норма L_0 вектора дорівнює його кількості ненульових елементів). Однак з точки зору обчислення для таких даних це здійснити неможливо. Крім того, норма L_1 є гарним наближенням у багатьох випадках. Здатність LASSO виробляти дуже рідкісні рішення робить його привабливим для випадків, коли потрібна проста, можливо, навіть імітаційна модель. Проблема мінімізації з рівняння (1.5.5), за обмеження L_0 -норми та теорії навколо неї, також відоме як стиснене зондування. [1,9,13]

1.6 Покращення методів машинного навчання

Кожний метод машинного навчання не є ідеальним у своєму стані. Часто, залежачи від певної поставленої задачі, потрібно оптимізувати у певному вигляді вже існуючі алгоритми або створювати нові на базисах існуючих.

Способи покращення зазвичай поєднують ряд простих алгоритмів для створення потужної та гнучкої моделі. Наприклад випадкові ліси, де декілька сильних алгоритмів навчаються незалежно та поєднуються шляхом простого усереднення, щоб зменшити дисперсію ансамблевої моделі.

Слабкі системи / алгоритми, що розвиваються, не навчаються самостійно і поєднуються для зменшення упередженості порівняно з одним слабким. Найпоширенішими методами, особливо в поєднанні з методами основаними на random forest, є адаптивне підвищення та зменшення похибок рішень (зменшення гілок та незручних шляхів). У матеріалознавстві їх почали застосовувати для прогнозування різних типів даних, починаючи від простих даних, до надпровідності та інших.

Для повного покращення штучного інтелекту слід застосовувати не тільки більш комплексно обширні алгоритми, а також більш адаптовані під них дані, які будуть аналізуватися за допомогою правильно обраного підходу. Вказуючи точніше, нам потрібно зібрати в одну систему сукупність тих методів та засобів, які забезпечують максимально можливу продуктивність, а у нашому випадку, найменшу похибку.

Саме тому в подальших пунктах і будуть розглядатися більш детально моделі random forest, яка найбільш близька до нашої задачі.

1.7 Приклади застосування машинного навчання та штучного інтелекту для покращення результатів прогнозування властивостей матеріалів

У матеріалознавстві, як вказувалося в пунктах вище, досить нещодавно розпочалось впровадження статистичних методів для аналізування матеріалів різної класифікації. Зазвичай подібні методи обмежуються не тільки апаратними можливостями, а також існуючими підходами, про які теж розповідалося пунктами вище. З більш нових розробок вже можливе використання нейронних мереж для аналізування механічних властивостей сплавів, які складаються з міді та алюмінію, до яких додається срібло та магній у певних пропорціях. [24]. Такі сплави утворюються завдяки порошковому напиленню і прогнозування властивостей до виготовлення буде набагато вигіднішим, бо є можливість спрогнозувати наперед існуючі властивості того чи іншого сплаву. Для цього використовувалася розробка «SMOreg/puk» сенс якої в використанні алгоритмів, які мінімально оптимізуються для регресії з використанням ядерних методів. При дослідженні матеріалу цей метод був найбільш точним та і мав найнижчу похибку серед порівняння з іншими методами. [24-25]

Побудову моделі машинного навчання при використанні нейронних мереж, можна розділити на 3 етапи [24]:

1. Збір даних
2. Моделювання
3. Валідація моделі (перевірка точності моделі та отриманих даних)

У зборі даних для застосування у подібному вигляді можна встановити три обмеження. Вони впливають з властивостей використаних матеріалів та їх призначення. Це можуть бути обмеження, наприклад масових часток легуючих елементів, які не повинні перевищувати певні норми. Також обмеження можуть бути представленими у вигляді використання лише одного методу синтезування матеріалу (наприклад порошкове напилення). Особливу увагу потрібно виділяти для фізико-хімічних властивостей матеріалу ще до початку аналізування. Деякі матеріали не можуть бути

поєднані один між одним для певної задачі (наприклад провідності та надпровідності).

Таблиця 1.1

Властивості литої та спеченої олов'яно-свинцевої бронзи як приклад дослідження в металургії [24-26]

Характеристика	Свинцева бронза	
	Лита	Спечена
Густина кг/м ³	9200	8500
Твердість по методу Брінелля, МПа	500	250,0
Міцність на розтяг, МПа	140	45
Подовження, %	6,0	3,5
Коефіцієнт тертя з мастилом	0,008	0,005

У порошковій металургії питання прогнозування найбільш вдачного матеріалу є досить важливим аспектом дослідження. Для матеріалів які утворюють за допомогою спікання є густина, твердість, міцність та інші професіональні приклади.

В нашому випадку, як і в прикладах хімічний склад матеріалів є важливим фактором, що регулює його властивості та мікроструктуру. Так як ми будемо брати лише певний набір даних в потрібній комбінації, подібні зауваження також вплинуть і на нашу роботу. У моїй роботі не розглядається саме синтез матеріалів, бо у роботі фактично ми опираємося на будову кристалічної решітки, яка в свою чергу вирішує які саме властивості будуть притаманні сплаву з узятих у наступних пунктах матеріалах

На табл. 1.1 на прикладі литої та спеченої олов'яно-свинцевої бронзи ми можемо побачити різницю між характеристиками технологій лиття та спікання.

З таблиці видно що литі матеріали є більш вигіднішими на відміну від матеріалів отриманих методами спікання, але враховуючи властивості експлуатаційного характеру [26], а також собівартість створення, зносостійкість є значно вищою ніж у ливарних матеріалах. З цього можна зробити висновок про релевантність методів спікання, а також порошкового напилення. Але також з цього виникає проблема в тому, які саме пропорції сплавів потрібно отримати. Ми б могли зробити декілька

практичних спроб та дослідити параметри матеріалу наприклад за методом Вікерса або Брінелля дослідити міцність матеріалу, але з цього випливає декілька проблем:

1. Економічна. Фактично ми витратимо велику кількість матеріалу.
2. Час плавлення матеріалу є досить затяжним навіть у промислових масштабах, що не вигідно включаючи два попередніх пункти.
3. Стандартизація виготовлення. Через дві попередні проблеми фактично неможливо утворювати нові пропорції або взагалі нові сплави матеріалів.

Саме подібні недоліки, а також більшу гнучкість у виготовленні матеріалів, незважаючи на методи вказані вище повинен вирішувати штучний інтелект, якщо розглядати проблему не тільки з научної точки зору, а й з промислової.

Аналізування матеріалів на основі Cu-Al дозволяє прискорити розробку багатоконпонентних сплавів, які мають еластокалоричні властивості та мають певний ефект пам'яті форми [29]. Такі матеріали при своєму виході крім зазначених властивостей є екологічно чистими (завдяки властивостям магнітокалоричного, електрокалоричного, барокалоричного та еластокалоричного охолодження). Це має досить важливий розвиток у подальшому.

Суть у цьому прикладі полягає у аналізі еластокалоричного ефекту та зміні ентропії ΔS , який вказує на зміну температури ΔT в адіабатичних умовах для певного заданого матеріалу. Для його знаходження використовують систему EGO – ефективну глобальну оптимізацію, яка використовує у своєму базисі глибинні Гауські процеси [30]. На рис 1.7 можна побачити приклад побудованої системи.

Розглядаючи у загальному випадку методи аналізування матеріалів не слід обмежуватися тільки сплавами та порівнянням їх властивостей. Використання методів аналізу матеріалів які базуються на методах штучного інтелекту також можна використовувати наприклад для синтезу термохімічного пального за допомогою оксидів металів перовскітів, які до цього мали використання у сонячній енергетиці. У 2013 році це був один з великих проривів для наукового дослідження [27-28]. Використання методів аналізу для цього матеріалу розглядалось більш детально у подальшому, що призвело до використання у багатьох сферах енергетики.

З цього прикладу можна зробити висновок, що в основі, відмінність кожної системи машинного навчання фактично відрізняється моделлю/методом, який використовується для обробки інформації та її відсіювання, а також від точності після проведення перевірок даних.

У своїй роботі ми будемо розглядати певні структуровані тернарні матеріали, які будуть відрізнятися один від одного кристалічною решіткою, а саме об'ємноцентрованою, та гранецентрованою, при цьому припустивши той факт, що вони можливі при певних заданих будовах [31].

Перед тим як розглядати прогнозування можливих матеріалів необхідно дізнатися які є особливості будови кристалів та види кристалічної решітки для того, щоб відокремити які саме типи генерованих матеріалів ми будемо брати у нашому випадку.

У кристалах частинки, з яких складається кристал, знаходяться досить близько, майже до стану зіткнення один з одним і розташовані по різному, але є закономірність в напрямках. Будову кристалу фактично можна представити у вигляді решітки з вузлами – атомами. Слід зазначити що у металах, у вузлах знаходяться позитивні іони, між якими рухаються електрони [31].

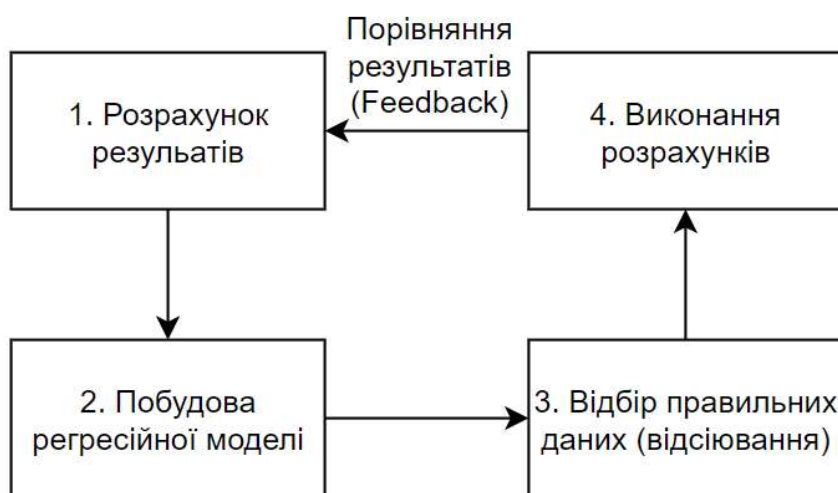


Рис.1.7 - Приклад побудованої системи для машинного навчання розрахунку зміни ентропійних властивостей матеріалів [29-30]. З рисунку можна побачити замкнену систему навчання та її основні аспекти.

1.8 Твердість матеріалу. Методи вимірювання

Для твердості матеріалу існує досить багато визначень. Твердістю можна назвати властивості матеріалу та його реакцію на проникнення іншого твердого тіла, або здатність матеріалу чинити опір до деформації під дією зовнішніх чинників чи властивостей матеріалів.

Твердість матеріалу є одним з головних аспектів для дослідження на сьогодні. Існує безліч напрямків та галузей, які займаються винаходженням та розробкою матеріалів та дослідженням цих самих матеріалів у галузі матеріалознавства. Конструкційні можливості завдяки дослідженням все набагато сильніше стрибнули вперед.

В чистих металах при взаємодії між сусідніми атомами виникає слабка залежність, яка пов'язана з конфігурацією групи, оскільки міжатомні сили в металах повністю залежать від локальної щільності електронів та мають велику дальність. В кристалічних структурах, ковалентно пов'язані сили необхідні для зсуву атомів, локалізовані і набагато більші за значеннями у порівнянні з металами, що обмежує їх дислокаційний рух.

В простих металах найбільш важливою є взаємодія між дислокаціями, а не взаємодія між атомами. Твердість металу залежить від деформаційного затвердіння, а не від індивідуальної рухливості частинок. При цьому еластична стійкість до сил зсуву відіграє ключову роль, так як бере безпосередню участь у дислокаційній рухливості, яка є основною для твердості матеріалу.

Так як твердість і модуль зсуву є пропорційними величинами. Модуль зсуву фактично є функцією локальної поляризованої і має залежність від електронної щільності і енергії, яка необхідна для просування валентного електрону до його першого збудженого стану. Це в свою чергу залежить між хімічною взаємодією двох атомів.

Існує декілька методів перевірки матеріалів на міцність, а саме метод Бринелля, метод Роквелла та метод Віккерса.

Для дослідження металів в основному використовується останній метод і його фізичну модель ми будемо використовувати як основу для перевірки твердості матеріалу. Сенс цього методу полягає у витисненні на матеріалі за допомогою алмазної

піраміди з кутом 136° відбитку на матеріалі. Температура при цьому варіюється в діапазоні $10-40^\circ\text{C}$ [42].

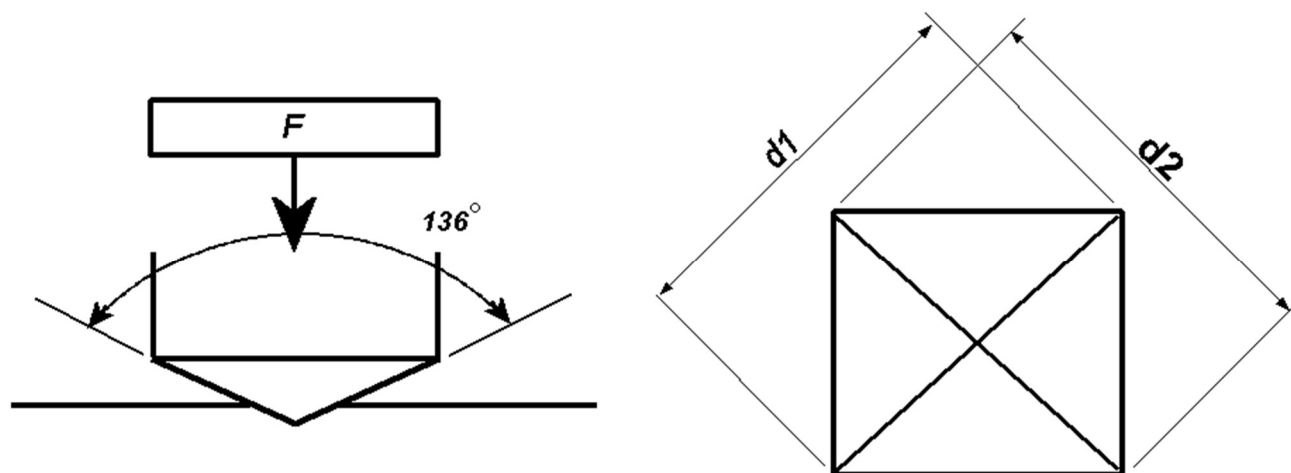


Рис 1.8 – Параметри та приклад індентора який використовується в методі Віккерса. [42].

Фізична модель, яка буде використовуватися для прогнозування буде описана у пункті 2.

1.9 Модуль Юнга

Важливо відрізнити жорсткість матеріалу, до якого прикладена певна сила, яка може викликати деформацію у матеріалі від міцності, яка відноситься до стійкості матеріалу до руйнування або деформації. Жорсткість зазвичай вимірюється шляхом накладення невеликих навантажень для виміру деформації матеріалу [35].

Так як при деформації у більшості матеріалів за таких умов малі показники, проблема є у тому, як саме вимірювати ці невеликі змінні при деформації та інших прикладених силах до матеріалу. При експериментальних дослідженнях матеріалів було виявлено, що навантаження відносно деформації має лінійне співвідношення до тих пір, поки навантаження не настає досить великим. Таке співвідношення знайоме нам як закон Гука і його можна описати у вигляді формули 1.6.

$$P = k\delta \quad (1.6)$$

У формулі k – константа жорсткості.

Деформацію яка вказана у формулі як δ можна нормалізувати, при цьому враховуючи, що навантаження буде розподілене по матеріалу рівномірно [35]:

$$\epsilon = \frac{\delta}{L_0} \quad (1.7)$$

L_0 - початкова довжина ϵ -безрозмірна міра розтягування (деформація). Виходячи з формули ми можемо представити закон Гука в іншому вигляді [35]:

$$\sigma = E\epsilon \quad (1.8)$$

У формулі 1.8 константа E -є модулем Юнга або модулем пружності. Робимо висновок, що закон Гука прямо залежить від модуля Юнга та параметрів і розмірів зразка. Тому жорсткість можна визначити як [35]:

$$k = \frac{AE}{L} \quad (1.9)$$

З рівняння 1.8 випливає що модуль Юнга характеризує здатність матеріалу чинити опір до розтягнення або стиснення при пружній деформації.

У роботі фактично ми будемо прогнозувати модуль пружності для сплавів різної кристалічної будови та різного набору металів використовуючи за модель фізичні аспекти модуля пружності.

1.10. Основні кристалічні структури, притаманні металам

У роботі для аналізу композитів та матеріалів буде використовуватися ОЦК та ГЦК структури. Для більшого розуміння їх потрібно пояснити, а у пункті 2 потрібно розкрити модель генерації структури та композитного складу.

Приклад решітки яка складена в ОЦК структуру можна побачити на рис 1.9.

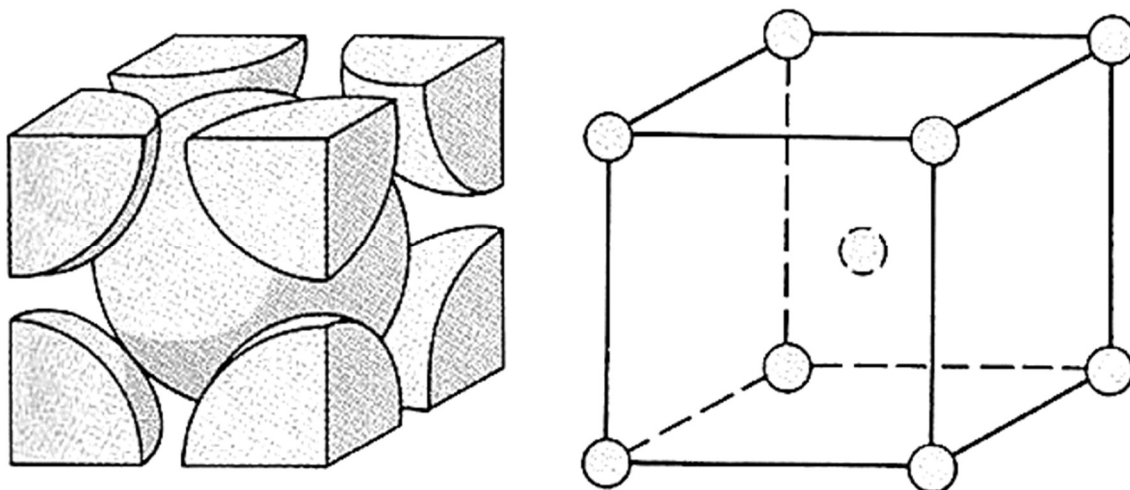


Рис 1.9 – Приклад будови ОЦК структури за щільністю пакування.

Подібна структура складається з 9 атомів, якщо брати до уваги повну структуру не враховуючи об'єм, який атоми займають у кристалічній решітці. За правилом жорстких сфер об'єм який будуть займати атоми всередині решітки буде дорівнювати 3. У кожному з кутів буде $1/8$ частинки, в центрі – повна сфера. Взаємозв'язок між кристалічними решітками і структуру матеріалу з подібних конструкцій можна побачити на рис 1.10.

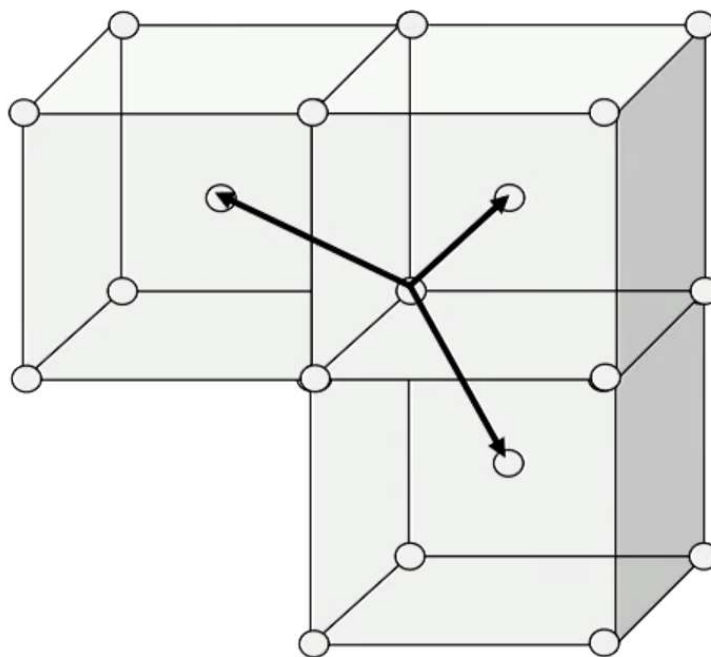


Рис 1.10 – Взаємодія частинок в кристалічних структурах.

Конструкція та зв'язок між частинками повністю визначає властивості матеріалу. Для цієї конструкції притаманні мала твердість, та непогані показники модулів деформації, так як вони мають невеликий атомний зв'язок один між одним. Досить важливим параметром для кристалічної решітки є коефіцієнт пакування (відсоток заповнення простору частинками). Його можна розрахувати за формулами 1.10-1.12 для ОЦК решітки. Коефіцієнт пакування представляє х себе об'єм сфер, який поділений на об'єм елементарної комірки (рис 1.9 для прикладу)

Для створення структури за теорією жорстких сфер потрібно щоб усі сфери доторкалися один до одної. Беручи до уваги діагональ з частинками, це дозволяє нам розписати взаємозв'язок, який дозволяє розрахувати щільність пакування кристалічної решітки (рис 1.11).

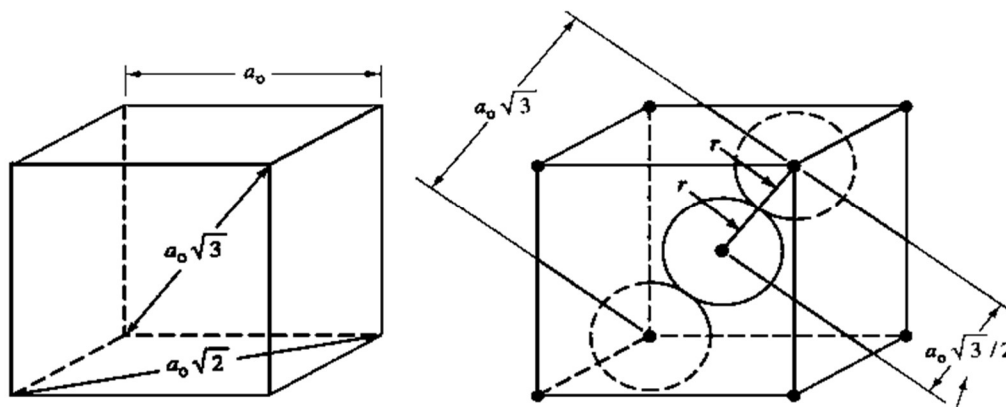


Рис 1.11 – Розрахунок щільності пакування на основі діагоналей.

$$pf = \left(\frac{2 \left(\frac{4}{3} \pi r^3 \right)}{a^3} \right) \quad (1.10)$$

$$a\sqrt{3} = 4r; a = \frac{4r}{\sqrt{3}} \quad (1.11)$$

$$pf = \left(\frac{2 \left(\frac{4}{3} \pi r^3 \right)}{\left(\frac{4r}{\sqrt{3}} \right)^3} \right) = 0.68 = 68\% \quad (1.12)$$

Приклад ГЦК структури можна побачити на рис 1.12. Для конструкції ГЦК характерні великі значення твердості через досить сильний зв'язок між частинками а також враховуючи що частинки мають більшу щільність пакування.

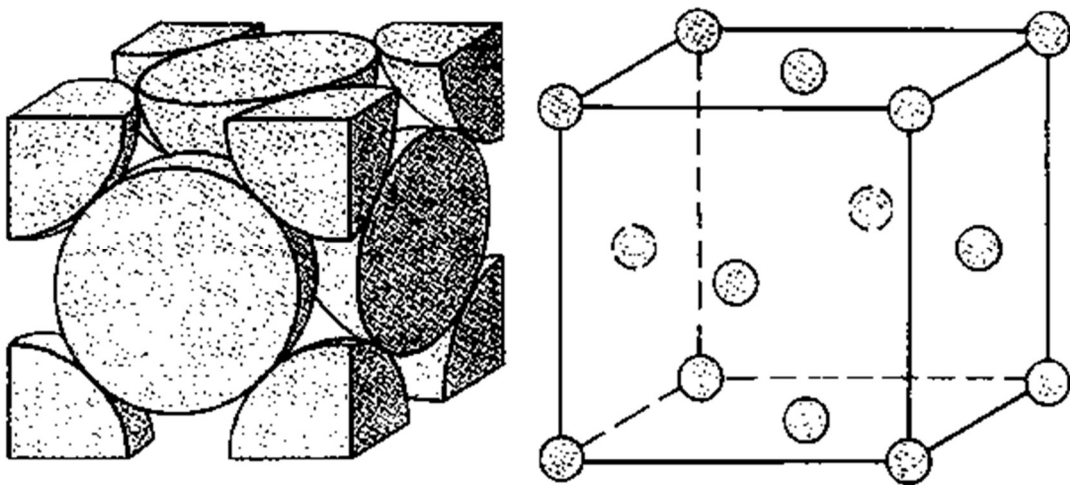


Рис.1.12 - Приклад будови ГЦК структури за щільністю пакування.

Подібна структура складається з 8 по $\frac{1}{8}$ сфер у кутах та 6 по $\frac{1}{2}$ сфер з боків, що в сумі дає займаний об'єм у 4 сфери. Коефіцієнт пакування можна представити як число сфер помножене на об'єм сфер та поділене на кубічну одиницю (формула 1.11; 1.13-14)

$$a\sqrt{2} = 4r; a = \frac{4r}{\sqrt{2}} \quad (1.13)$$

$$pf = \left(\frac{4 \left(\frac{4}{3} \pi r^3 \right)}{\left(\frac{4r}{\sqrt{2}} \right)^3} \right) = 0.74 = 74\% \quad (1.14)$$

Для побудови моделі у пункті 2 будуть використовуватися саме ці конструкції та на цих засадах і розрахунках при урахуванні жорстких сфер буде створена генерація матеріалів та отримання даних для подальшої обробки.

2 МЕТОДИ ДОСЛІДЖЕННЯ

2.1 Засоби розробки

Вибір мови програмування для створення алгоритму обирався по важливим критеріям розрахунків, а саме простоті кодування та можливостям розрахунку окремих існуючих параметрів та для інтеграції.

Головними аспектами/можливостями для засобів розробки повинні бути наступними:

- Можливість зчитування вхідних даних у вигляді файлу для подальшої обробки.
- Розв'язок задачі, обраним окремим алгоритмом
- Можливість кодування цього алгоритму та внесення правок
- Виведення результатів, отриманих обробленим алгоритмом у вигляді таблиці/файлу Excel
- Можливості виконання принципів машинного навчання
- Можливість перевірки та знаходження помилок при розрахунках.

У якості основної мови програмування, на якій буде створюватися програма / алгоритм (LinearForestRegressor) буде обрана мова програмування Python, так як подібна мова програмування вважається стандартною для обробки великих масивів даних, а також є досить легкою для навантаження на комп'ютер, що позитивно впливає на час отримання результатів.

У якості альтернативи можливо було використовувати такі засоби як MatLab або R. Так як MatLab фактично є комерційним продуктом та він не може підтримувати виконувати принципові функції та алгоритми машинного навчання, а тому й не може вважатися штучним інтелектом.

R був би гарною альтернативою, але так як для нашої задачі більш доцільніше використовувати мову програмування Python (простота кодування, швидкість та додаткові модулі).

Python вважається потужною мовою програмування з досить легким програмним кодом, яка має ефективну структуру складних даних на високому рівні та в

одночас є простою та з легкістю застосовується для об'єкто-орієнтованих типів програмування. В нашому випадку, простий та точний синтаксис, можливість динамічної обробки даних.

У цій роботі будуть використовуватися стандартні бібліотеки, які присутні у вільному доступі (як код для алгоритмів штучного інтелекту так і певні бінарні дистрибутиви для операційних систем). Скачаний з офіційного сайту Python має вільне розповсюдження на офіційному сайті та розповсюджує спеціальні рір-модулі, посилання, програми-утиліти та додаткову документація.

Одним з головних пунктів вибору також була інтеграція інших мов, так як мова програмування Python може бути доповнена та розширена за допомогою створених алгоритмів інших мов (наприклад C або C++ або інші мови програмування які пов'язані з мовою C)

З переваг вибору мови програмування можна відокремити такі пункти:

- Чистота написання синтаксису. Блоки можливо відокремити один від одного відступами
- Можливість перенесення програми як і більшості інтерпритованих мов.
- Можливість виведення отриманих даних у терміналі або діалоговому вікні (для перевірки правильності виконання окремих пунктів).
- Має досить великі можливості та розповсюджені математичні оператори, які дозволяють працювати з досить великими числовими масивами,

2.2 Особливості будови штучного інтелекту для дослідження модуля еластичності та твердості.

У даній роботі буде знайдений модуль об'ємної пружності та твердості на основі інших даних, які отримаємо з певної бази даних (material project) та за допомогою штучного інтелекту. Інакше кажучи, створений штучний інтелект, використовуючи при цьому трансферний тип навчання, буде за допомогою відкритих open-source алгоритмів та плагінів досліджувати певний структурований набір даних з певного багатокomпонентного сплаву. При цьому слід зазначити, що чим менше елементів, тим більш точніше буде визначатися об'ємний модуль пружності.

В цій роботі основним методом порівняння даних була лінійна лісова регресія (random forest з інтеграцією в лінійну регресію), метод моделювання залежності між скаляром у та векторною (у загальному випадку) величинами X , які будуть знаходитися у базі даних. При статистичному (ручному) аналізі доводиться проводити досить багато розрахунків, особливо якщо багатокomпонентний сплав є високоентропійним (матиме більше 5-ти окремих хімічних елементів/металів).

2.3 Принцип роботи алгоритмів штучного інтелекту

Для пояснення алгоритму потрібно пояснити спочатку які саме основні модулі використовуються для розрахунку потрібних параметрів. В роботі були використані модулі «rumatgen» та «matminer» в якості модулів з відкритим доступом та відкритим вихідним кодом. Також слід зазначити допоміжні модулі

Matminer – модуль для Python, який дозволяє вилучувати данні про властивості матеріалів. У своєму складі він містить підпрограми які дозволяють отримувати данні з різних «банків даних», як наприклад «material project», включаючи складні параметри матеріалів (кристалічна структура, склад матеріалу, та інш.). Сам модуль не має у собі реалізації машинного навчання, він дозволяє на його основі підготувати/перетворити набір даних, для його подальшого аналізу або обміну даними.[3]

Rumatgen – модуль/бібліотека, яка призначена для аналізу властивостей та параметрів матеріалу. Rumatgen дозволяє проводити складні розрахунки, так як містить потужні інструменти для аналізу. Дозволяє генерувати фазові діаграми, проводити дифузійний аналіз та інші складні параметри, як щільність, твердість, модуль еластичності та модуль пружності.

Numpy – розширення для мови Python, яка додає спеціальні розрахункові допоміжні властивості для роботи з багатовимірними масивами та матрицями, а також має у своєму складі високорівневі математичні функції для операцій над вказаними вище масивами. Створений з відкритим вихідним кодом і має багато прихильників, які користуються ним для розрахунку та має багато незалежних розробників.

Мова Python – інтерпретована мова. Для будь-якої інтерпретованої мови, математичні алгоритми працюють досить повільно. Для мови Python – цей час обробки

розрахунків набагато повільніший ніж в інших існуючих мовах програмування. NumPy фактично вирішує цю проблему, надаючи велику кількість алгоритмів, допоміжні функції які пов'язані з масивами та велику кількість математичних операторів.

Цей плагін ми використовуємо для заміни використання масивних формул, що дозволяє нам не використовувати інші готові рішення, як наприклад MatLab. Слід зазначити що вони обидві схожі один на одного вмістністю алгоритмів.

Pandas – плагін-бібліотека, яка була розроблена для мови програмування Python для того щоб полегшити загрузку та вивгрузку даних, допомагаючи інтерпретувати дані для Matminer

Фактично, об'єднуючи ці два модулі в єдину систему, ми зможемо завантажувати необхідні данні з різних джерел або «банків даних» та проводити їх повний аналіз, використовуючи ті самі данні.

Так як Rumatgen містить у собі необхідні алгоритми для теоретичних розрахунків, можна проводити ці розрахунки для певних нових стабільних елементів, враховуючи той факт, що існують відомі певні властивості елементів у сплаві/поєднанні матеріалів. Іншими словами, вилучаючи певні данні та проводячи розрахунок, ми отримуємо додатковий масив параметрів, які вподальшому також можна використовувати.

Фактично, виконується принцип машинного навчання, за яким створюється замкнений ланцюг аналізування отриманих даних.

2.3.1 Алгоритм лінійної регресії

Лінійна регресія - це лінійний підхід до моделювання зв'язку між скалярною відповіддю (або залежною змінною) та однією або кількома пояснювальними змінними (або незалежними змінними).

У лінійній регресії взаємозв'язки моделюються за допомогою лінійних функцій предиктора, параметри яких невідомі моделі оцінюються з даних. Найчастіше умовне середнє значення відгуку, що задається значеннями пояснювальних змінних (або предикторів), вважається афінною функцією цих значень; рідше використовується умовна медіана або якийсь інший квантил. Як і всі форми регресійного аналізу, лінійна

регресія зосереджується на умовному розподілі ймовірності відповіді з урахуванням значень предикторів, а не на спільному розподілі ймовірностей усіх цих змінних, що є областю багатовимірного аналізу. Приклад лінійної регресії можна побачити на рис 2.1

Цей метод буде використаний для порівняння з обраним нами методом випадкових лісів, який буде використовуватися як основна машинного навчання.

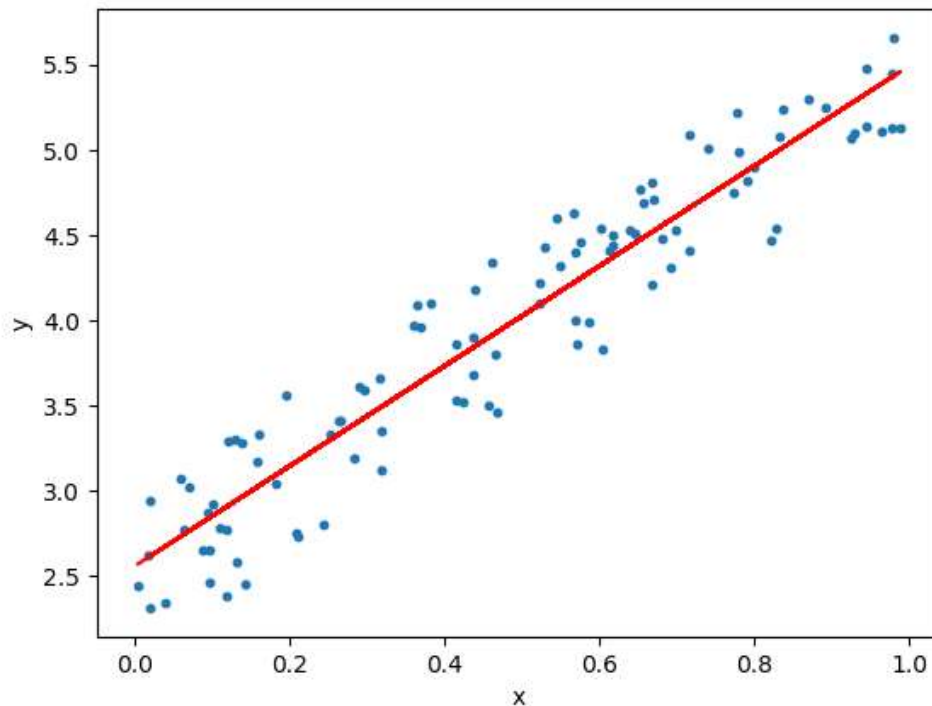


Рис - 2.1– Приклад графіку лінійної регресії для випадкових пов'язаних даних.

2.3.2 Випадкові ліси (Random forest)

Random forest (випадкові ліси) - це комбінація предикторів дерев, така що кожне дерево залежить від значень випадкового вектора, відібраного незалежно та з однаковим розподілом для всіх дерев у лісі. Похибка узагальнення для лісів сходиться до обмеження, оскільки кількість дерев у лісі стає великою. Похибка узагальнення лісу класифікаторів дерев залежить від сили окремих дерев у лісі та співвідношення між ними. Використовуючи випадковий вибір функцій для розділення кожного вузла, приводиться до помилок, які вигідніше порівняти з Adaboost [38], але є більш надійними щодо шуму (зайвих гілок). Внутрішні оцінки відстежують помилки, міцність та кореляцію, і вони використовуються для показу реакції на збільшення кількості

функцій, що використовуються при розщепленні. Внутрішні оцінки також використовуються для вимірювання змінної важливості. Ці ідеї також застосовні до регресії, яка є ключовим ядром у нашому розрахунку.[11]

На рис.2.2 можна побачити приклад «гілок», які представляють собою певні вибори / шляхи, за якими алгоритм обирається в залежності від операції

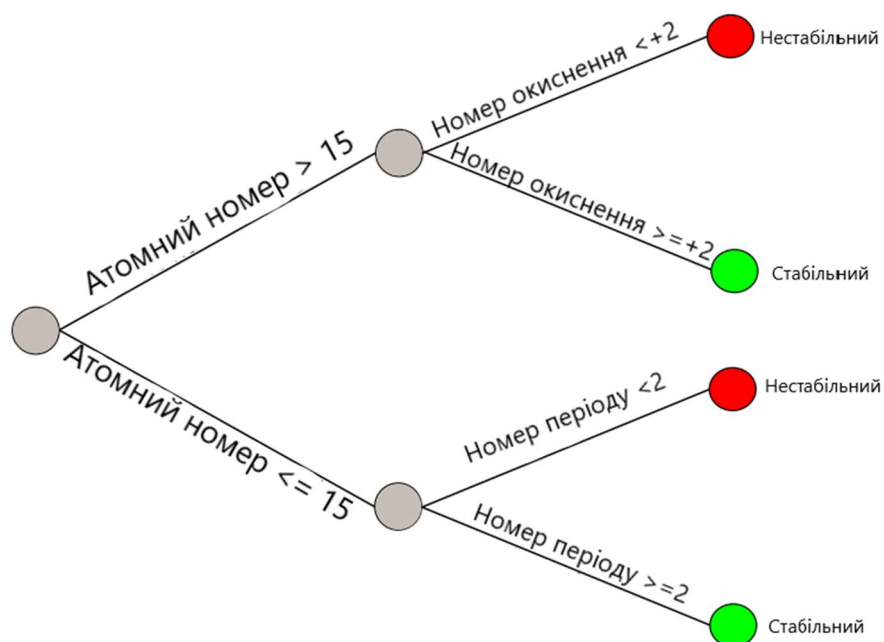


Рис 2.2 – Принцип роботи алгоритму випадкових лісів визначення стабільності

Значні покращення точності класифікації зумовили вирощування ансамблю дерев та надання їм вибору для найпопулярнішого класу.[1,4]

2.4 Особливості генерації структур та композитного складу.

Для генерації сплавів були використані матеріали на основі трьох складових, при цьому за основу. Крок кількості атомів при цьому буде взятий крок наявності атомів в залежності від порівнянного елементу для кожного елементу.

Генерація матеріалів буде відбуватися для ідеальних випадків, тобто будуть узяті ідеальні параметри елементів, які будуть використовуватися у моделюванні.

Також слід зазначити що генерація буде різною в залежності від того, яка саме кристалічна решітка використовується для певного поєднання матеріалів.

Генерація відбувається в залежності від даних, які ми отримуємо з бази даних за певними властивостями. Так як генерація відбуватиметься на основі ОЦК та ГЦК кристалічної решітки а також в залежності від того, скільки складових використовується при генерації матеріалу.

Головною задачею при генеруванні файлу з структурою та її даними потрібно розрахувати щільність кристалічної решітки. У пункті 1 були розглянуті основні формули для цього. Щільність – це маса на одиницю об'єму. В цьому випадку, для розрахунку потрібно зробити певні зміни у формулах 1.10-1.15

$$\rho = \frac{m}{V_{\text{ГЦК/ОЦК}}} \quad (2.1)$$

$$m = \sum \frac{m_i}{n_i} \quad (2.2)$$

Де m_i - маса кожної частинки, n_i - кількість елементарних комірок, у яких вона розташована

Для ОЦК структур:

$$V_{\text{ОЦК}} = a^3; a = \frac{4r}{\sqrt{3}} \quad (2.3)$$

Для ГЦК структур

$$V_{\text{ГЦК}} = a^3; a = \frac{4r}{\sqrt{2}} \quad (2.4)$$

Приклад генерації можна зобразити на прикладі з чистим залізом, який має ОЦК структуру. Фактично, знаючи параметри матеріалу можемо з легкістю визначити як саме буде знаходитися щільність. Радіус атома $Fe = 1.24 \times 10^{-8} \text{ см}$

$$\frac{2 \text{ атоми}}{\text{ОЦК решітку}} \times \frac{55.85 \text{ г/моль}}{\text{молярна маса}} \left(\frac{1 \text{ моль для } Fe}{6.023 \times 10^{23} \text{ (константа авогадро)}} \right) \quad (2.5)$$

$$V = 2.35 \times 10^{-23} \frac{\text{см}}{\text{ус}} ; \rho = 7.87 \frac{\text{г}}{\text{см}^3} \quad (2.6)$$

2.5 Фізична модель для знаходження твердості та модуля пружності

Так як у роботі використовується метод знаходження твердості за Віккерсом, є потреба у створенні моделі яка використовує подібний метод знаходження а також порівнює значення з базою експериментальних та теоретичних даних.

За Віккерсом твердіть є співвідношенням між силою навантаження індентора до площі відступу

$$H_V = \frac{2F \sin\left(\frac{\theta}{2}\right)}{d^2} \quad (2.7)$$

У формулі d – середня діагоналі відступу, θ – кут між протилежними гранями піраміди індентора (рис 1.8 та рис 2.3). Для правильного отримання моделі потрібно припустити два фактори:

1. Індентор пірамідальної форми можна розподілити на 4 сегменти у формі триглатних пірамідальних виступів.
2. Твердість Віккерса буде вимірюватися в межах еластичної шкали.

Припускаючи ці два аспекти можемо знайти для кожної трикутної частинки модуль зсуву G . [37]

$$G = \frac{F}{4A \tan(\alpha)} \quad (2.8)$$

Формула 2.8 фактично є співвідношення між напругою та деформацією зсувів. Ми можемо виразити точну область зсуву A через область деформації A^* та геометрію його відступу (формули 2.9-2.10)[37].

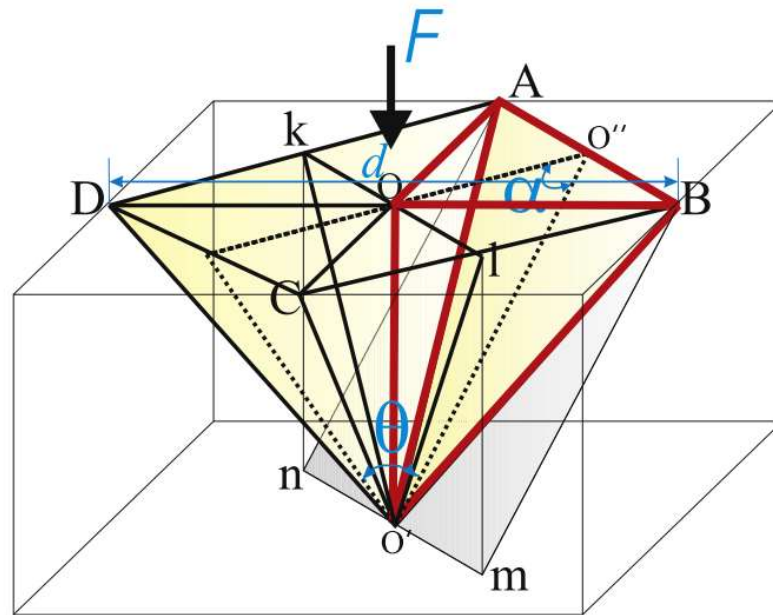


Рис 2.3 - Відбиток залишений алмазним індентором у квадратній площині. Червоним кольором вказано одну з чотирьох частин пірамідального виступу.

$$A^* = \frac{1}{8} d^2 \tan(\alpha) \quad (2.9)$$

$$A = cA^* = \frac{c}{8} d^2 \tan(\alpha) \quad (2.10)$$

В данному випадку значення c являє собою коефіцієнт пропорційності, що при пружній деформації області зсуву буде малою. При реальному вимірюванні нехтувати коефіцієнтом пропорційності ми не можемо через його невеликі значення, але якщо припустити що цей коефіцієнт достатньо великий, то формулу 2.7 можна зобразити у іншому вигляді та об'єднати з формулою 2.10

$$H_V = G \operatorname{tag}^2(\alpha) \sin\left(\frac{\theta}{2}\right) = 0.92G \operatorname{tag}^2(\alpha) \quad (2.11)$$

У формулі 2.11 $\sin\left(\frac{\theta}{2}\right)$ визначається параметрами індентора, і ми можемо припустити що цей параметр є постійним (що впливає з використаного методу та рівняння 1.8. Для алмазного індентора це значення дорівнює 0.92, для ідеального індентора:

$$\alpha = (\pi - \theta)/2.0 \quad (2.12)$$

Враховуючи формулу 2.6, можна спростити формулу 1.4.5 до вигляду:

$$H_V = 0.151G \quad (2.13)$$

Враховуючи запропоновані [39], автором Pugh співвідношення $\frac{G}{B}$ формулу 2.13 можна зобразити в наступному вигляді:

$$H_V = C k^m G^n; \left(k = \frac{G}{B}\right) \quad (2.14)$$

У формулі 2.14 H_V - твердість, G – модуль зсуву, B – модуль об'єму. Так як зазвичай ці параметри відомі і мають залежність один від одного експериментально [38], формулу 2.14 можна представити як остаточну модель

$$H_V = 2(k^2 G)^{0.585} - 3 \quad (2.15)$$

При аналізуванні матеріалу формула 2.15 буде грати роль фізичної моделі у нашій системі. Рівняння демонструє, що при збільшенні об'ємного модуля пружності твердість зменшиться до того випадку, коли він залишиться незмінним і навпаки, при

зменшенні об'ємного модуля пружності збільшиться твердість. Також з цього випливає висновок, що досліджувані та прогнозовані нами величини, а саме модуль пружності та твердість матеріалу будуть взаємопов'язані один з одним

2.6 Хід роботи

Спочатку ми отримуємо данні за допомогою «matminer», імпортувавши усі необхідні системи, такі як «pandas», для подальшої роботи з даними. Після цього потрібно підігнати данні під лінійну регресію, порівнюючи основні данні з розрахованими за допомогою штучного інтелекту (рис 2.4), через кілька ітерацій, а також порівняємо з моделлю випадкового дерева (рис 2.5), порівнявши розходження даних по діагоналі.

До того ми повинні створити масив даних, які ми згенеруємо також за допомогою matminer. Це буде виглядати фактично як великий масив з переліку різного поєднання металів, а також імпортування найбільш важливих даних. Для модуля об'єму цими даними будуть щільність та атомарний об'єм, за допомогою яких будуть розраховані подальші параметри.

Важливим параметром для правильного вибору методу є точність методів та кросс-валідація (перехресна перевірка).

Оцінка точності моделі є найважливішою частиною процесу створення моделей машинного навчання та є важливим критерієм у ефективності роботи моделі та її прогнозуванню. Показники оцінки змінюються залежно від типу проблеми.

Метод лінійної регресії може бути типовим прикладом такого типу проблем, і головна характеристика проблеми регресії полягає в тому, що цілі набору даних містять лише реальні числа. Помилки вказують, наскільки модель допускає помилок у своєму передбаченні. Основна концепція оцінки точності полягає в порівнянні вихідної цілі з передбачуваною за певними показниками. Знаходження точності методів дає нам такі висновки. Перший, ми одразу можемо знайти похибку кожного методу, так як похибка повністю залежить від вихідних даних, яка б розмірність у них не була. По друге, це дозволяє оцінити точність самого методу та одразу дізнатися наскільки він є точним, а з цього випливає, чи можливо його використовувати взагалі.

Точність кожного методу можна розрахувати за формулами 2.10-2.13

MAE – середня абсолютна похибка, різниця між початковими даними та прогнозованими значеннями, які узяті з абсолютної різниці по набору усіх даних

MSE – середньоквадратична помилка, різниця між початковими та передбачуваними значеннями, взяті в квадрат середньої різниці по набору усіх даних.

R^2 - коефіцієнт визначення, який являє собою відповідність значень порівняно з вихідними значеннями. Значення отримуються в відсотках и чим більший відсоток – тим кращий метод моделювання

$$MSE = \frac{1}{N} \sum_{I=1}^N (y_i - \hat{y})^2 \quad (2.10)$$

$$RMSE = \frac{1}{N} \sqrt{\sum_{I=1}^N (y_i - \hat{y})^2} \quad (2.11)$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})}{\sum(y_i - \bar{y})} \quad (2.12)$$

$$MAE = \frac{1}{N} \sum_{I=1}^N |y_i - \hat{y}| \quad (2.13)$$

Де \hat{y} -спрогнозоване значення; \bar{y} - середнє значення.

Фактично ми розраховуємо загальну середньоквадратичну помилку масиву даних, що і вказує на точність кожного з методів

Розраховуючи ці данні отримуємо для обох систем значення R:

Для лінійної регресії у випадку модуля пружності

$$R^2 = 0.607 = 60.7\%$$

$$RMSE = 59.02$$

Для регресії випадкового лісу для випадку модуля пружності

$$R^2 = 0.98 = 98\%$$

$$RMSE = 13.386$$

Одразу можемо зробити висновок про те, що точність другого методу набагато більша ніж у лінійної регресії. За допомогою кросс-валідації можна перевірити достовірність точності та уточнити наскільки точними є значення

Кросс-валідація - це техніка валідації моделі для перевірки того, наскільки успішно застосовується в моделі статистичний аналіз здатний працювати на незалежному наборі даних. Зазвичай крос-валідація використовується в ситуаціях, де метою є передбачення, і хотілося б оцінити, наскільки пророкує модель здатна працювати на практиці. Один цикл крос-валідації включає розбиття набору даних на частини, потім побудова моделі на одній частині (званої тренувальним набором), і валідація моделі на іншій частині (званої тестовим набором). Щоб зменшити розкид результатів, різні цикли крос-валідації проводяться на різних розбитті, а результати валідації усереднюються по всіх циклах (формула 2.14-2.15). [5]

$$T^l = F_1 \cup \dots \cup F_k \quad (2.14)$$

$$CV_k = \frac{1}{k} \sum_{i=1}^k Q\left(\mu\left(\frac{T^j}{F_i}\right) F_i\right) \rightarrow \min \quad (2.15)$$

Після проведення кросс-валідації в нашому випадку

$$RMSE: 35.809$$

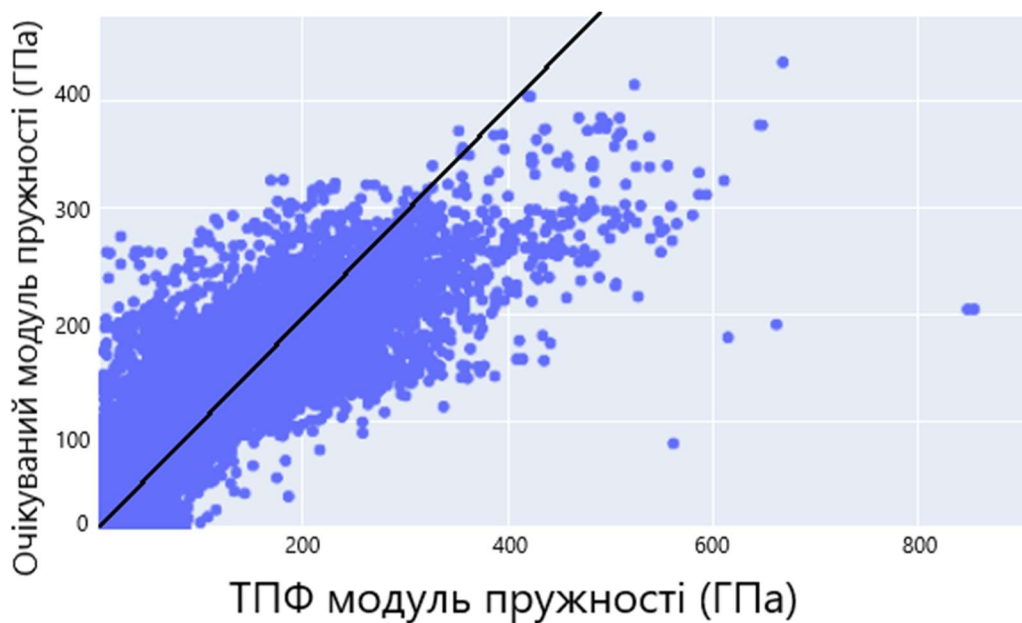


Рис 2.4 – Порівняння очікуваного модуля пружності для лінійної регресії

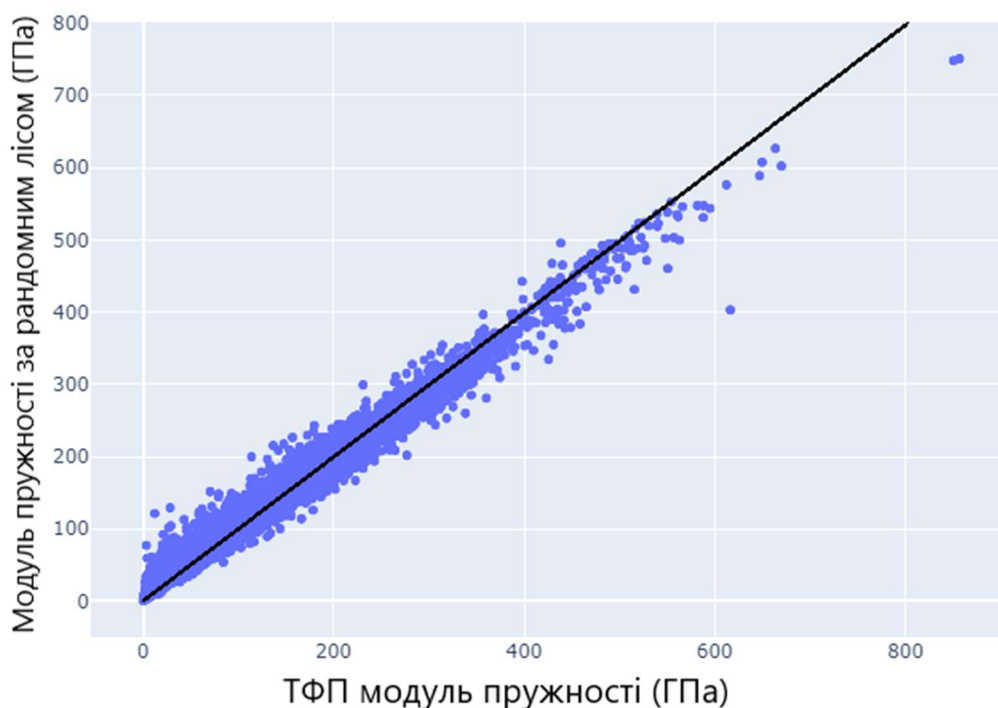


Рис 2.5 – Порівняння отриманих за допомогою алгоритму та відомих

З графіків одразу можемо зробити висновок про точність та мінімальні розходження двох підходів та який саме з них буде ключовим на даному етапі. Також важливим пунктом буде розрахунок «важливості кожного з параметрів для необхідного нам розрахунку об'ємного модуля пружності, а у подальшому для твердості.

Також слід зазначити що на рис 2.4-2.5 модель використовує для прогонки 90% від загальної бази даних яку імпортує, і 10% для тестування, змінюючи сектор с кожним циклом, яких усього 10.

Головним показником при прогнозуванні виходячи з даних будуть такі параметри:

Для модуля Юнга це об'єм елементарної комірки, яку займає кожний атом (46-47%), температура кипіння (19%), температура плавлення (9%) та атомний радіус (2%). Слід зазначити що ці дані можуть варіювати свій відсоток в залежності від параметрів, які потрібно прогнозувати.

Для твердості найбільш важливими параметрами для прогнозування будуть радіус атому (18%) об'єм атому та маса атому (10%), а також параметри частинок такі як група та період елементів та щільність комірки (в залежності від її типу)

Лише після цього ми можемо отримати розрахунок модуля пружності та твердість для попередньо згенерованих нами матеріалів, враховуючи попередньо обраний метод машинного навчання, та відокремивши необхідні нам параметри для розрахунку.

3 РЕЗУЛЬТАТИ

Основними результатами у нашій роботі – прогнозування параметрів різних типів сплавів, а саме бінарних та тернарних. Так як наш алгоритм працює фактично ґрунтуючись на попередніх розрахунках схожих елементів та/або нових елементів, слід зазначити, що деякі з розрахунків в силу недосконалості системи будуть досить сильно відрізнятися від істини, що залежить від складності самого матеріалу.

В даній роботі був проведений розрахунок / прогнозування модуля пружності та твердість промодельованих матеріалів.

Підгружаючи потрібні нам дані за допомогою `matminer` та використовуючи розрахунковий модуль `rumatgen`, ми можемо розпочати теоретичний аналіз, враховуючи дочірні параметри, які у пункті 2 ми виявили як найбільш важливі.

`Rumatgen` може розраховувати досить масивну кількість матеріалів вне залежності від кількості їх у сплаві, тому попередньо згенерований набір сплавів можна зручно та швидко обробити за допомогою цього модуля.

Також `Rumatgen` може генерувати ці дані в єдину таблицю, отримуючи набір з досліджуваних даних та їх теоретичною обробкою, за тієї умови, що це можливий стабільний елемент.

3.1 Твердість та модуль Юнга матеріалу для Ni-Nb с ГЦК та ОЦК структурою.

В методі дослідження була повністю описана модель, за якої виконується дослідження нашого матеріалу. Слід зазначити що генерований матеріал не буде мати дефекти та фактично являє собою монокристал. Тобто ми розглядаємо матеріал за ідеальних параметрів, при відсутності різних зовнішніх чинників. Це також буде впливати на різницю між отриманими результатами та експериментальними значеннями які будуть. Також слід зазначити що експериментальні значення будуть отримані в інших розбіжностях, тобто нам потрібно одразу конвертувати коефіцієнт Віккерса в потрібну нам розмірність (ГПа). Це можна зробити, помноживши значення коефіцієнту Віккерса на 0.009807.

В експериментальних значеннях беруться не ідеальний приклад сполуки ніобій-нікелю, але для прикладу та для порівняння прогнозованих значень і отриманих експериментально цей приклад підходить і дозволить наглядно побачити особливості будови матеріалів.

Для розуміння фазового складу матеріалу в залежності від концентрації ніобію слід звернути увагу на фазову діаграму та за яких параметрів матеріал змінює свої властивості

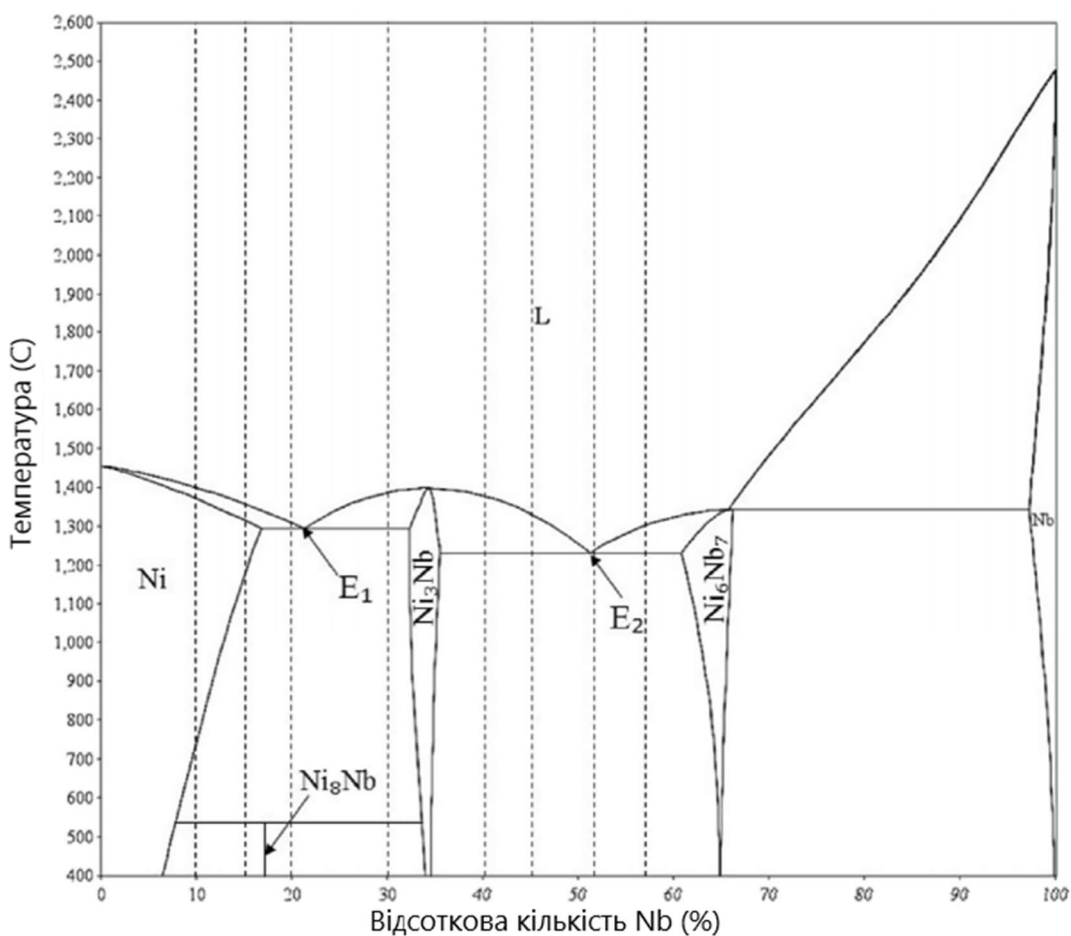


Рис 3.1 – Фазова діаграма ніобій-нікелю. Приклад узятий з експериментальних значень для порівняння [43]

Для дослідження матеріалу, крім порівняння прогнозованих параметрів та експериментальних значень, одразу отримаємо необхідні значення, а також знайдемо одразу найбільш тверде поєднання нікель-ніобію. Прогнозовані дані отримані за допомогою фізичної моделі яка вказана у пункті 2 роботи та фактично є прогнозованими

даними для ідеального матеріалу без дефектів (монокристал ГЦК структури). Генерація буде відбуватися з кроком наявності елементів в 5, від 20 до 80.

Таблиця 3.1

Значення прогнозованої твердості для ГЦК структури сплаву Ni-Nb

№	Твердість (ГПа)	Твердість (Hv)	Матеріал (сплав)
1	2	3	4
1	6,50905	663,7147151	Nb20Ni80
2	6,133804	625,4516654	Nb22Ni78
3	7,118425	725,8513878	Nb24Ni76
4	7,259851	740,2723158	Nb26Ni74
5	7,586479	773,5779205	Nb28Ni72
6	7,301816	744,5514351	Nb30Ni70
7	6,504838	663,2852034	Nb32Ni68
8	7,110914	725,0855593	Nb34Ni66
9	7,875828	803,0822797	Nb36Ni64
10	7,650828	780,1395229	Nb38Ni62
...
30	5,739565	585,2519094	Nb78Ni22
31	5,82743	594,2113181	Nb80Ni20

Одразу можемо відокремити найкращі показники твердості матеріалів узявши найбільші значення твердості матеріалів. Потім ми зможемо їх порівняти з отриманим модулем пружності та зробити певні висновки [36].

Таблиця 3.2

Найкращі значення для згенерованих матеріалів ГЦК структури сплаву Ni-Nb

№	Твердість (ГПа)	Твердість (Hv)	Матеріал (сплав)
1	2	3	4
11	8,813927764	898,738428	Nb42Ni58
12	8,266914695	842,9606093	Nb44Ni56
13	8,309929909	847,3467839	Nb46Ni54
14	8,121012848	828,0832924	Nb48Ni52
15	8,120908665	828,072669	Nb50Ni50

За отриманими даними одразу можемо зробити висновок про те, що найкращі пропорції для твердості, враховуючи дану структуру, буде поєднання Nb₄₂Ni₅₈. Для перевірки правильності прогнозування нашою моделлю потрібно перевірити декілька пунктів, а саме:

1. Згенерувати для такого-ж самого матеріалу / сплаву структуру, яка побудована за іншим алгоритмом (тобто згенерувати ОЦК структуру для перевірки).
2. Порівняти ці дані з експериментальними для цього матеріалу, для порівняння точності моделювання, враховуючи похибки.

Для ОЦК структури генерація матеріалу буде відрізнятися і враховуючи розглянуті дані у пункті 2, значення повинні бути меншими.

Таблиця 3.3

Значення прогнозованої твердості для ОЦК структури сплаву Ni-Nb

№	Твердість (ГПа)	Твердість (Hv)	Матеріал (сплав)
1	2	3	4
1	5,242848656	534,6026976	Ni20Nb80
2	4,933318923	503,0405754	Ni22Nb78
3	4,654624287	474,6226458	Ni24Nb76
4	5,518118825	562,6714413	Ni26Nb74
5	5,605446084	571,5760257	Ni28Nb72
...
30	6,977385857	711,4699559	Ni78Nb22
31	6,960737301	709,7723362	Ni80Nb20

Таблиця 3.4

Найкращі значення для згенерованих матеріалів ОЦК структури сплаву Ni-Nb

№	Твердість (ГПа)	Твердість (Hv)	Матеріал (сплав)
1	2	3	4
19	8,675927831	884,6668533	Ni58Nb42
17	8,122044015	828,1884383	Ni54Nb46
18	8,093970958	825,3258854	Ni56Nb44
15	8,014370509	817,2091882	Ni50Nb50
20	7,984341367	814,1471772	Ni60Nb40

За отриманими даними одразу можемо зробити висновок про те, що найкращі пропороції для твердості, враховуючи дану структуру, буде поєднання Nb₅₈Ni₄₂ [35-36].



Рис 3.2 – Порівняння значень твердості для ОЦК та ГЦК структур Ni-Nb

З отриманих даних, порівнюючи їх з експериментальними даними [35-36] можемо зробити декілька висновків:

1. Для генерації з кристалічною решіткою ГЦК значення міцності будуть більшими за значеннями ніж у ОЦК структури.
2. Проходячи певну межу, для ОЦК структур міцність стає набагато більшою, в залежності від відсоткового вмісту ніобію, на що впливає атомний радіус частинок
3. Враховуючи досліджені матеріали зі схожими структурами, є розходження які пояснюються похибками прогнозування та експерименту а

також в прогнозуванні не враховуються зовнішні чинники та дефекти матеріалу, та йде перевірка реальної моделі.

Для знаходження модуля Юнга використовуються ті ж самі генерації структур з ОЦК та ГЦК решітками. Потім можливо знайти найкращі матеріали по відношенню до твердості та модуля Юнга водночас [36].

Таблиця 3.5

№	ОЦК модуль Юнга (ГПа)	ГЦК модуль Юнга (ГПа)	Сплав
1	2	3	4
1	224,1384	232,7796	Nb20Ni80
2	219,75	236,4696	Nb22Ni78
3	209,3196	264,5568	Nb24Ni76
4	206,5536	311,1228	Nb26Ni74
5	205,9608	313,0788	Nb28Ni72
6	221,13	312,072	Nb30Ni70
7	219,372	315,3432	Nb32Ni68
8	211,8852	258,1212	Nb34Ni66
9	240,6036	262,2072	Nb36Ni64
10	240,4668	258,1404	Nb38Ni62
...
30	224,7696	230,2404	Nb78Ni22
31	228,4776	229,362	Nb80Ni20

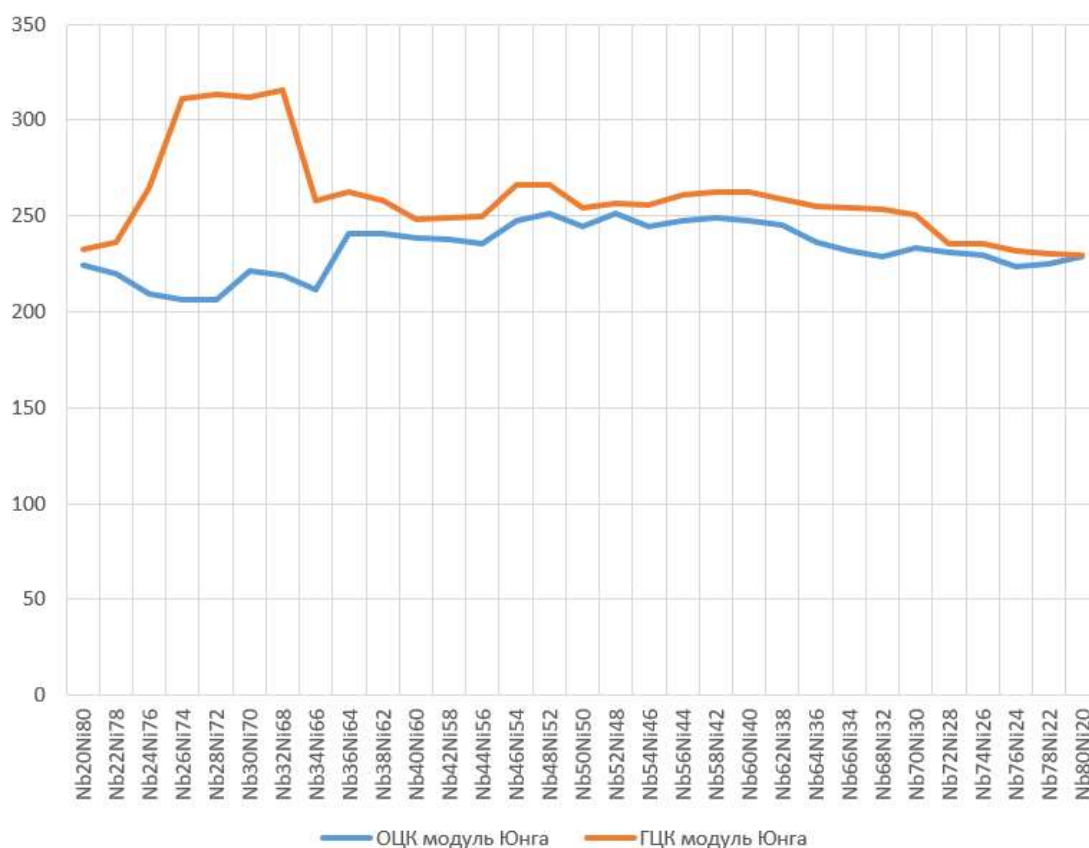


Рис 3.3 – Залежність модуля Юнга ОЦК та ГЦК структури

Так як попередньо ми вказали на те, що у нашій роботі самим важливим елементом для перевірки твердості та пружності є атомний радіус. Це можна пояснити тим, що при збільшенні концентрації атомів іншої речовини у сплаві вона починає деформувати кристалічну решітку, через що виникають розходження даних на певному пороговому значенні.

Також ми можемо виявити найкращий матеріал по відношенню твердості до модуля Юнга. Для цього використаємо формулу 3.1. Чим менше буде між ними розходження, тим кращі показники твердості та модуля Юнга для сплаву.

$$K = \frac{\text{Твердість}}{\text{Модуль Юнга}} \quad (2.13)$$

Проаналізувавши масив даних можна виявити, що найкращими значеннями для ОЦК структури буде поєднання Nb₂₄Ni₇₆, а для ГЦК структури це Nb₃₂Ni₆₈

3.2 Твердість та модуль Юнга матеріалу для матеріалів с ОЦК структури для Ti, V, Nb, Mo, Ag

Так як алгоритм для знаходження бінарних та тернарних сполук досить схожий і відрізняється лише генерацією потрібних даних, ми можемо одразу спрогнозувати твердість та модуль Юнга [20]. Для перевірки візьмемо схожі за властивостями матеріали, які можуть складатися в чисту ОЦК структуру.

Таблиця 3.6

Розрахований модуль пружності для згенерованого набору сплавів

№	Елементи	Щільність	Температура кипіння	Температура плавлення	Електронегативність	Модуль пружності (Гпа)
1	2	3	4	5	6	7
1	Ti20V40Nb40	6,71	4190,8	2361,4	1,6	142,5972

Продовж.табл 3.6

1	2	3	4	5	6	7
2	Ti20V40Mo40	7,07	4148,8	2419,8	1,824	193,9272
3	Ti20V40W40	11,09	4515,2	2739,4	1,904	258,5448
4	Ti20V40Ag40	7,25	3158	1755,372	1,732	138,3288
5	Ti20Nb40W40	11,92	5050	2966,2	1,892	264,6276
6	Ti20Nb40Ag40	8,25	3692,8	1982,172	1,72	109,1964
7	Ti20Mo40W40	12,47	5008	3024,6	2,116	285,2832
8	Ti20Mo40Ag40	8,65	3650,8	2040,572	1,944	154,9944
9	Ti20W40Ag40	12,41	4017,2	2360,172	2,024	214,5852
10	V20Nb40Mo40	8,5	4707,6	2695	1,83	223,3344
11	V20Nb40W40	12,4	5074	3014,6	1,91	275,796
...
295	V40Mo40Ag20	8,24	3923,8	2278,586	1,902	185,9196
296	V40W40Ag20	12,27	4290,2	2598,186	1,982	247,7712
297	Nb40Mo40W20	11,08	5137,2	2997,4	1,976	277,764
298	Nb40Mo40Ag20	9,22	4458,6	2505,386	1,89	174,5016
299	Nb40W40Ag20	13	4825	2824,986	1,97	260,6712
300	Mo40W40Ag20	13,58	4783	2883,386	2,194	275,9472

З зроблених розрахунків можна зробити висновок, що алгоритм отримує приблизно подібні результати до тих, які утворюються за цим методом. Враховуючи складність матеріалу, це спрощує його дослідження. У табл. 3.2 можна побачити те, які саме данні утворюються та порівняти модуль пружності з значеннями твердості [22]. В таблиці також присутні коефіцієнти електронегативності. Вони потрібні для перевірки можливості отримання сплаву за правилами Юма-Розери. Електронегативність не повинна мати велику різницю між чистими елементами для сплаву. Так як ми взяли елементи які можуть складатися в ОЦК структуру та мають близькі значення один до одного, а також різниця між їх електронегативністю мінімально відрізняється, то можна вказати що можливе утворення твердого розчину-сплаву.

Тепер можливо спрогнозувати значення твердості для таких самих сполук.

Таблиця 3.7

Отримані значення коефіцієнту твердості

№	Твердість (ГПа)	Твердість (Hv)	Сплав
1	2	3	4
1	7,927414157	808,342425	Ti20V40Nb40

Продовж.табл 3.7

1	2	3	4
2	7,592937705	774,236536	Ti20V40Mo40
3	8,381073261	854,601128	Ti20V40W40
4	5,071572108	517,137974	Ti20V40Ag40
5	5,3506992	517,137974	Ti20V40Ag40
6	13,082538	537,949322	Ti20Nb40Mo40
7	20,535858	879,360841	Ti20Nb40W40
8	8,0652768	412,360022	Ti20Nb40Ag40
9	21,359646	971,334735	Ti20Mo40W40
10	8,8890648	584,062033	Ti20Mo40Ag40
11	16,3423848	846,340293	Ti20W40Ag40
12	12,4117392	607,805539	V20Nb40Mo40
13	19,8650592	886,186603	V20Nb40W40
14	7,394478	493,941966	V20Nb40Ag40
15	20,6888472	1061,97681	V20Mo40W40
16	8,218266	619,495512	V20Mo40Ag40
...
299	7,042592512	718,118947	Nb40W40Ag20
300	7,026854918	716,514216	Mo40W40Ag20

З даних видно одразу що найбільш міцними матеріалами є ті, які мають у собі вміст вольфраму. В чистому вигляді він також має досить великі значення міцності та непогані показники гнучкості.

Для визначення найкращих матеріалів з коефіцієнтом твердості, можна взяти матеріали з максимальними коефіцієнтами (таблиця 3.8), а також проаналізувати залежність між модулем пружності та твердістю (рис 3.4).

Таблиця. 3.8

Матеріали з найкращим коефіцієнтом значення твердості

Твердість (ГПа)	Сплав
1	2
10,48118109	V40W40Ag20
10,41480659	V20Mo40W40
10,20995813	V25Mo35W40
10,07415956	V30Mo30W40
9,878348213	V25Mo40W35
9,768881612	V35Mo30W35
9,74730259	V30Mo35W35
9,69317189	V35Mo35W30

Продовж. Табл 3.8

9,650394741	V35Mo25W40
9,580998634	Ti25Mo35W40
9,565911361	V30Mo40W30

Видно що найкращій матеріал по значенням твердості це V40W40Ag20. За відношенням кращого матеріалу за формулою 3.1 можна знайти найкращій матеріал за обома параметрами ($Ti_{35}Nb_{30}Mo_{35}$)

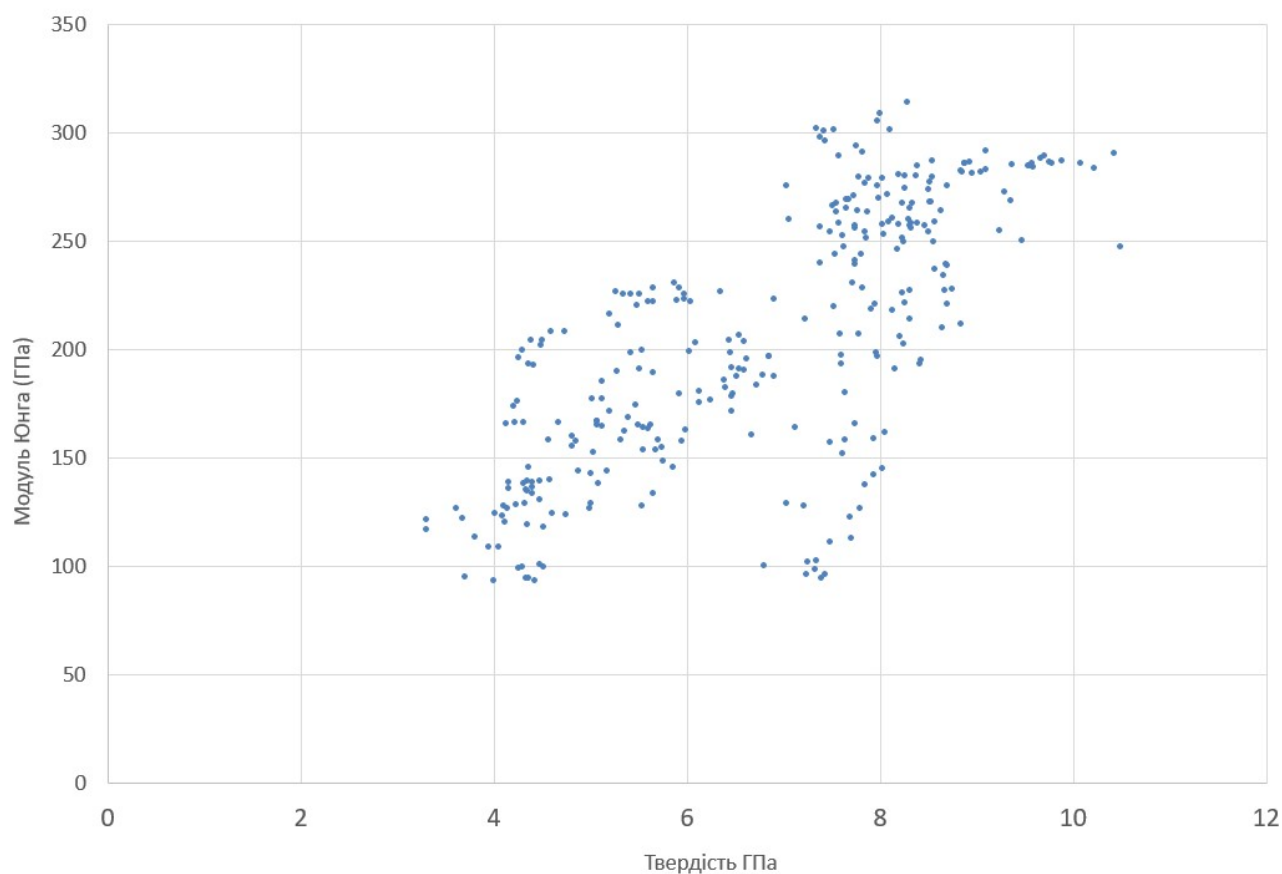


Рис 3.4 – Залежність твердості від модуля Юнга для тернарних сплавів

4 ТЕХНІКА БЕЗПЕКИ ТА ОХОРОНА ЗДОРОВ'Я

При роботі за комп'ютером одним з головних критеріїв є безпека та здоров'я працюючого на робочому місці. До важливих критеріїв праці можна віднести: Зручність робочого місця, правильне освітлення приміщення, вентиляція, температура оточуючого середовища, вологість повітря, виробничий «шум» та інші чинники.

Перед початком роботи була проведена бесіда з техніки безпеки на підприємствах та на робочому місці, певні нормативні норми, а також була поставлена задача преддипломної практики, та були визначені дати її виконання. [37]

1. Загальні вимоги

1.1. Вимоги інструкції поширюються на всіх працівників, які у своїй діяльності здійснюють роботу, пов'язану з персональними електронно-обчислювальними машинами (ПЕОМ) з візуальними дисплейними терміналами (ВДТ) і периферійними пристроями (ПП) (клавіатура, маніпулятор "миша", сканери, друкуючі пристрої тощо).

1.2. Дана інструкція призначена для запобігання або зменшення несприятливої дії на працівників шкідливих факторів, які супроводжують роботу з ПЕОМ:

- зорове та нервово-емоційне напруження при перебуванні у вимушеній робочій позі, яка супроводжується локальним напруженням верхніх кінцівок на фоні обмеженої загальної м'язової активності (гіподинамії);
- наявність шуму та вібрації (при роботі друкуючих пристроїв);
- м'яке рентгенівське випромінювання при роботі з ВДТ, що мають електронно-променеву трубку;
- електромагнітне випромінювання;
- ультрафіолетове та інфрачервоне випромінювання;
- електростатичне поле між екраном та оператором;
- наявність пилу, озону, оксидів азоту та аероіонізації.

1.3. До роботи на ПЕОМ допускаються особи, що пройшли попередній та періодичний медогляд, вступний та первинний інструктаж з питань охорони праці та пожежної безпеки.

1.4. Оператор ПЕОМ один раз на 6 місяців повинен проходити повторний інструктаж з питань охорони праці та один раз на рік інструктаж з питань пожежної безпеки.

1.5. Забороняється допускати осіб, молодше 18 років, до виконання самостійних робіт з профілактичного обслуговування, налагодження та ремонту ПЕОМ.

1.6. Кожен працівник, під час виконання своїх обов'язків, повинен:

- виконувати тільки ту роботу, яка доручена йому керівником або передбачена посадовими чи робочими інструкціями;
- дотримуватися правил внутрішнього трудового розпорядку;
- дбати про особисту безпеку і здоров'я, а також про безпеку і здоров'я оточуючих людей в процесі виконання будь-яких робіт чи під час перебування на території університету;
- знати і виконувати вимоги правил пожежної безпеки, знати місця знаходження первинних засобів пожежогасіння та уміти ними користуватися;
- знати і виконувати вимоги нормативно-правових актів з охорони праці, вимоги санітарно-гігієнічних норм і правил, вміти надавати домедичну допомогу;
- користуватися передбаченими засобами колективного та індивідуального захисту.

1.7. ПЕОМ належать до електрообладнання з напругою до 1000 В, тому, при виконанні робіт працівник повинен пройти інструктаж для отримання І групи з електробезпеки та один раз на рік перевірку знань відповідних нормативних актів з охорони праці та пожежної безпеки.

1.8. Обслуговування, ремонт та налагодження ПЕОМ слід виконувати в окремому приміщенні.

1.9. Приміщення, де розміщуються ПЕОМ, повинні відповідати вимогам «Державних санітарних правил і норм роботи з візуальними дисплейними терміналами електроннообчислювальних машин» (ДСанПІН 3.3.2.007-98).

1.10. Штепсельні з'єднання та електророзетки крім контактів фазового та нульового робочого провідників повинні мати спеціальні контакти для підключення нульового захисного провідника. Конструкція їх має бути такою, щоб приєднання нульового захисного провідника відбувалося раніше ніж приєднання фазового та нульового робочого провідників. Порядок роз'єднання при відключенні має бути зворотним.

1.11. Неприпустимим є підключення ПЕОМ, ВДТ та ПП, а також устаткування для їх обслуговування, ремонту та налагодження звичайної двохрановідної електромережі, в тому числі з використанням перехідних пристроїв.

1.12. Забороняється застосування саморобних подовжувачів, які не відповідають вимогам «Правил улаштування електроустановок» до переносних електропроводок.

1.13. Приміщення з ЕОМ повинні бути забезпечені переносними вуглекислотними вогнегасниками з розрахунку 2 шт. на кожні 20 кв. м площі приміщення.

1.14. Працівник має право відмовитися від роботи, якщо умови її виконання створюють загрозу його здоров'ю або життю чи суперечать вимогам нормативних актів з охорони праці.

1.15. За порушення вимог цієї інструкції працівник несе адміністративну, дисциплінарну та карну відповідальність згідно діючого законодавства України.

2. Вимоги безпеки перед початком роботи

2.1. Оглянути робоче місце і привести його в порядок; переконатися, що на ньому відсутні сторонні предмети, все обладнання і блоки ПЕОМ з'єднані між собою за допомогою з'єднувальних шнурів.

2.2. Перевірити зовнішнім оглядом загальний стан апаратури, перевірити справність електропроводки, з'єднувальних шнурів, штепсельних вилок, розеток.

2.3. Перевірити надійність установки апаратури на робочому столі. Не допускається встановлювати ВДТ на краю стола. Він має бути встановлений так, щоб було зручно дивитися на екран – під прямим кутом (а не збоку) і трохи зверху вниз (при цьому екран має бути трохи нахилений таким чином, щоб нижній його край знаходився ближче до оператора (користувача)).

2.4. Відрегулювати освітленість робочого місця. За необхідності ввімкнути загальне та/або місцеве освітлення.

2.5. Відрегулювати та зафіксувати висоту крісла, зручний для оператора (користувача) нахил його спинки.

2.6. У разі потреби приєднати до системного блоку необхідну апаратуру (принтер, сканер тощо) – всі кабелі, що з'єднують системний блок з іншими пристроями, слід приєднувати та роз'єднувати тільки при вимкненому та відключеному від електричної мережі комп'ютері.

2.7. Включити апаратуру в наступній послідовності: стабілізатор напруги (блок безперебійного живлення), монітор, системний блок, принтер (якщо передбачається друкування).

2.8. Щодня перед початком роботи необхідно очищати екран ВДТ від пилу та інших забруднень.

2.9. Забороняється самостійно усувати помічені під час огляду обладнання несправності, пошкодження обладнання, порушення вимог безпеки тощо. Про таке необхідно терміново доповісти своєму керівнику або технічному персоналу, що обслуговує ПЕОМ, та до роботи не приступати. Приступати до роботи дозволяється тільки після усунення виявлених недоліків.

3. Вимоги безпеки під час роботи

3.1. Робота з персональним комп'ютером:

- необхідно стійко розташувати клавіатуру на робочому столі, не допускаючи її хитання. Разом з тим має бути передбачена можливість її поворотів та переміщень. Положення клавіатури та кут її нахилу мають відповідати

- побажанням оператора (користувача). Якщо в конструкції клавіатури не передбачено простір для опори долонь, то їх слід розташовувати на відстані не менше 100 мм від краю столу в зручній для роботи зоні. Під час роботи з клавіатурою необхідно сидіти прямо, не напружуватися;
- для зменшення несприятливого впливу на оператора (користувача) маніпулятора типу «миша» (вимушена поза, потреба постійного контролю за якістю дій) необхідно забезпечити достатню для переміщення «миші» і зручного упору ліктьового суглоба поверхню столу;
 - не дозволяються сторонні розмови, подразнюючі шуми;
 - періодично при вимкненому комп'ютері прибирати бавовняною ганчіркою, ледь змоченою мильним розчином, пил з поверхонь апаратури. Екран ПЕОМ протирають ватою, змоченою в спирті, або спеціальними серветками. Не дозволяється використовувати рідину або аерозольні засоби для очищення поверхонь комп'ютера;
 - забороняється самостійно виконувати ремонт апаратури, а також класти будь-які предмети на апаратуру ПЕОМ, оскільки це може вивести її з ладу;
 - для зняття статичної електрики рекомендується час від часу, торкатися рукою металевих поверхонь (батареї центрального опалення тощо);
 - для зниження напруженості праці при роботі з ПЕОМ необхідно рівномірно розподіляти і чергувати характер робіт відповідно до їх складності. З метою зменшення негативного впливу монотонії доцільно застосувати чергування операцій введення тексту та числових даних (зміна змісту робіт), чергування редагування текстів та введення даних (зміна змісту та темпу роботи) тощо. и) з метою збереження здоров'я працюючих, зменшення негативного впливу на них виробничих факторів, запобігання професійним захворюванням, а також підтримки працездатності повинні бути передбачені для відпочинку наступні регламентовані перерви
 - під час регламентованих перерв з метою зниження нервово-емоційного напруження і втоми зорового аналізатора, усунення негативного впливу гіподинамії

- та гіпокінезії, а також запобігання розвитку позотонічної втоми рекомендується виконувати комплекс вправ, зазначених у додатку А до цієї інструкції;
- у разі виникнення у працюючих за ПЕОМ зорового дискомфорту та інших несприятливих суб'єктивних відчуттів, що настають, незважаючи на дотримання санітарно-гігієнічних і ергономічних вимог режимів праці та відпочинку, слід застосовувати індивідуальний підхід в обмеженні часу робіт за ПЕОМ і корекцію тривалості перерв для відпочинку або проводити заміну іншими видами робіт (не пов'язаних з використанням ПЕОМ).

4. Вимоги безпеки після закінчення роботи

4.1. Після закінчення роботи необхідно відключити від електричної мережі обладнання, підключене до ПЕОМ, в наступній послідовності:

- відключити додаткове обладнання (принтер, модем, або інше);
- відключити монітор;
- відключити системний блок.

4.2. Привести в належний стан робоче місце, прибрати носії інформації, робочі папери, інструменти, пристосування, захисні засоби та інше.

4.3. Повідомити безпосереднього керівника про всі виявлені під час роботи недоліки.

4.4. Вимкнути в робочому приміщенні освітлення та інше обладнання.

4.5. Працівник, який останнім залишає приміщення, обладнане протипожежною та охоронною сигналізацією, повинен здати його під охорону.

5. Вимоги безпеки під час аварійної ситуації

5.1. В процесі роботи можуть виникнути наступні аварійні ситуації:

- загоряння обладнання або матеріалів, виникнення пожежі;
- пошкодження ізоляції електрообладнання, виникнення напруги дотику та крокової напруги;

- дія природних факторів або явищ (підвищена або знижена температура, снігопад, гроза тощо).

5.2. У разі виникнення аварійної ситуації необхідно:

- обгородити небезпечну зону і не допускати в неї сторонніх осіб; якщо обстановка не загрожує життю і здоров'ю працівників і не призведе до більш тяжких наслідків, необхідно її зберегти такою, якою вона була на момент настання події, для проведення розслідування;
- при пожежі приступити до її ліквідації; якщо погасити пожежу своїми силами неможливо, викликати пожежну команду;
- повідомити про те, що сталося, керівника робіт.

5.3. У разі настання нещасного випадку:

- негайно звільнити потерпілого від дії уражаючого фактору;
- перенести його в безпечне місце;
- визначити стан потерпілого, при необхідності надати йому домедичну допомогу, а у тяжких випадках викликати "швидку медичну допомогу".

ВИСНОВКИ

Робота присвячена досить важливому дослідженню актуальної теми у матеріалознавстві, а саме прогнозуванню властивостей бінарних і тернарних сполук. Це досить важливе дослідження, так як зазвичай у нових сплавах досить складно передбачити властивості, особливо при його синтезуванні.

У процесі роботи були проведені глибокі дослідження даної теми. В літературному огляді була розглянута актуальність цього напрямку, важливі аспекти будови машинного навчання та штучного інтелекту, для подальшого застосування у роботі. Завдяки розглянутим алгоритмам у роботі став можливим вибір основного методу, у якості якого був вибраний `random forest`. Також були розглянуті інші застосування подібних структур по відношенню до аналізу та передбаченню властивостей матеріалів. У літературному огляді також були використані основні фізичні поняття, розглянута модель твердості та модуля пружності (модуль Юнга) та яким чином вони перевіряються у експериментальному полі.

У ході роботи були порівняні два алгоритми, які базуються на лінійній регресії та на `random forest`. З цього ми виявили що останній дає досить малі похибки для подальших значень після машинного навчання. Також були більш детально розглянуті моделі, які використовувались для отримання значень. На основі цих моделей було проведено прогнозування властивостей бінарних та тернарних матеріалів, які можуть складатися у обрані нами структури ГЦК та ОЦК.

Для дослідження були взяті сполуки які складаються з титану, ванадію, молібдену, ніобію, аргентуму (срібла) та вольфраму для тернарних матеріалів, та бінарних (Ni-Nb). В якості основного алгоритму для машинного навчання був використаний алгоритм `random forest`, як алгоритм за найбільшою точністю навіть після додаткової перевірки. Це дозволило якісно розрахувати тернарні сполуки. За допомогою `rumatgen` та `matminer` у роботі вийшло розрахувати такі параметри як твердість матеріалу та модуль пружності.

У процесі роботи було проаналізовано алгоритми штучного інтелекту, які базуються на лінійній регресії та обрали один з найбільш правильних методів для нашої

задачі, проаналізувавши його точність та перевіривши її, а також змогли отримати результати, у процесі яких виявили та перевірили залежності модуля пружності та твердості, а також порівняли дані з отриманими даними у процесі експериментальних досліджень.

В результаті були отримані значення та виявлений найкращий коефіцієнт для значення твердості та модуля пружності, також зроблені висновки, а саме що в процесі прогнозування основним чинником для розрахунку є щільність комірки та атомний радіус, по якому він розраховується. Найкращими матеріалами по твердості для бінарних сплавів були знайдені, а саме для бінарних це ОЦК $Nb_{24}Ni_{76}$, та ГЦК $Nb_{32}Ni_{68}$ для бінарних, та $V_{40}W_{40}Ag_{20}$ для тернарних сплавів.

Розвиток штучного інтелекту в цій галузі набагато спростить дослідження більш складніших високоентропійних матеріалів, як як таких матеріалів синтезовано досить мало, а при вдалому прогнозуванні великих масивів подібних матеріалів можна раніше знайти найбільш підходящий матеріал для створення нових сплавів. Також при достатньому розвитку, це дозволяє знаходити раніше невідомі залежності параметрів різних матеріалів. Дослідження в цій темі також досить сильно здешевлюють подальше дослідження матеріалів, так як не доводиться приходити до методів фізичної взаємодії та не має потреби псувати / деформувати сплави.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. From DFT to machine learning: recent approaches to materials science—a review / [R. S. Gabriel, C. P. Antonio, M. A. Carlos та ін.]. // *Journal of Physics: Materials*. – 16. – С. 47.
2. Gregory A. L. Application of machine-learning methods to solid-state chemistry: ferromagnetism in transition metal alloys / A. L. Gregory, G. Hagh. // *Academic press*. – 21. – С. 587–593.
3. Ward L. Matminer: An open source toolkit for materials data mining / L. Ward, A. Dunn, A. Faghaninia. // *Elsevier*. – 16. – С. 10.
4. Trevor H. The Elements of Statistical Learning Data Mining, Inference, and Prediction / H. Trevor, T. Robert, F. Jerome., 2009. – 764 с.
5. Уэс Маккинли Python и анализ данных / Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2015. – С. 482.
6. Data mining techniques for materials informatics: datasets preparing and applications / Y. Gang, C. Jingzhong, L. Zu. – China. Wuhan: School of Material Science and Engineering, China University of Geosciences, 2009. – 189 - 192 с. – (Second International Symposium on Knowledge Acquisition and Modeling).
7. Louridas P. Embedded Analytics and Statistics for Big Data / P. Louridas, E. Christof. // *Software technology*. – 2013. – С. 33–39.
8. [Xie T. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties / T. Xie, J. Grossman. // *Phys. Rev. Lett.* – 2018. – Vol 120. – С. 6.
9. Jonathan S. Recent advances and applications of machine learning in solidstate materials science / S. Jonathan, M. Marques, S. Botti. // *Shanghai Institute of Ceramics of the Chinese Academy of Sciences*. – 2019. – №5. – С. 36.
10. Marsland S. Machine Learning / Stephen Marsland. // *An Algorithmic Perspective*. – 2014. – С. 205.
11. Breiman, Leo. Random Forests // *Machine Learning (journal)* // *Machine Learning : journal*. — 2001. — Vol. 45, no. 1. — С. 5—32.

12. Altmann A. Permutation importance: a corrected feature importance measure / A. Altmann, L. Tolosi, O. Sande. // Department of Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, Saarbrücken, Germany. – 2010. – №26. – С. 1340–1347.
13. Tibshirani, R. Regression shrinkage and selection via the LASSO J. R. // Stat. Soc. Ser.. – 1996. – Vol 58, C. 267–288
14. Noura A. CrystalGAN: Learning to Discover Crystallographic Structures with Generative Adversarial Networks / A. Noura, N. Sokolovska, J. Crivello. // University Paris Est, ICMPE (UMR 7182). – 2009. – С. 9.
15. Sutton R. Reinforcement Learning: An Introduction / R. Sutton, A. Barto. // A Bradford Book The MIT Press Cambridge, Massachusetts London, England. – 2014. – №2. – С. 352.
16. COMBO: An efficient Bayesian optimization library for materials science / [T. Ueno, T. Rhone, H. Zhufeng та ін.]. // Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Toky. – 2016. – P. 4.
17. Machine-learning-assisted prediction of the mechanical properties of Cu–Al alloy / Zheng-hua Deng, Hai-qing Yin, Xue Jiang та ін.] – China: International Journal of Minerals, Metallurgy and Materials, 2020. – (3). – P. 362–373.
18. Machine leaning aided study of sintered density in Cu-Al alloy / Zhenghua Deng, Haiqing Yin, Xue Jiang та ін.] // Computational Materials Science – China: a Collaborative Innovation Center of Steel Technology, University of Science and Technology Beijing, 2018. – P. 48–54.
19. Новітні технології виробництва стандартизованих виробів (підшипників) / Олег Анатолійович Гавриш, Юлія Юріївна Віцюк, Тетяна Анатоліївна Роїк та ін.]. – Україна: НТУУ КПІ, 2012. – С. 204
20. Experimental, computational and thermodynamic studies in perovskites metal oxides for thermochemical fuel production: A review / Alicia Bayon, Alberto de la Calle, Krishna Kamol Ghose // International Journal of Hydrogen Energy / Alicia Bayon, Alberto de la Calle, Krishna Kamol Ghose. – Australia: c Priority Research Centre for

- Frontier Energy Technologies and Utilisation, The University of Newcastle, Callaghan, 2020. – (45; 23). – С. 12629–13128.
21. Robert Coontz. Science's Top 10 Breakthroughs of 2013 [Электронный ресурс] / Robert Coontz // sciencemag. – 2013. – Режим доступа до ресурсу: <https://www.sciencemag.org/news/2013/12/sciences-top-10-breakthroughs-2013>
 22. [Accelerating the development of multi-component Cu-Al-based shape memory alloys with high elastocaloric property by machine learning / Xin-Peng Zhao, Hai-You Huang, Cheng Wen та ін.]. – China, Beijing: Institute for Advanced Materials and Technology, University of Science and Technology Beijing, 2020. – 7 с.
 23. Efficient Global Optimization using Deep Gaussian Processes / Ali Hebbal, Loic Brevault, Mathieu Balesdent та ін.]. – France: ONERA, DTIS, Université Paris Saclay, F-91123 Palaiseau Cedex, 2018. – 12 с.
 24. Ryan Jacobs. The Materials Simulation Toolkit for Machine Learning (MAST-ML): an automated open source toolkit to accelerate data-driven materials research / Ryan Jacobs, Tam Mayeshiba, Ben Afflerbach. – USA: Department of Materials Science and Engineering, University of Wisconsin-Madison, Madison, WI, 53706,, 2019. – P.36 . – (1).
 25. Афанасьєва, О. В. Матеріалознавство та конструкційні матеріали / Афанасьєва, О. В. – Харків: ХНУРЕ, 2016. – С. 190. – (ХНУРЕ).
 26. Leo Breiman. Random forests / Leo Breiman. – California Berkeley: Statistics Department University of California, 2001. – С. 33.
 27. Comparison of incipient plasticity in bcc and fcc metals studied using nanoindentation / Suman Vadalakonda, Rajarshi Banerjee, Anantha Puthcode, Reza Mirshams // Materials Science and Engineering – USA: Department of Materials Science and Engineering, University of North Texas, Denton, 2006. – С. 208–213.
 28. Mariette Awad. Efficient Learning Machines / Mariette Awad, Rahul Khanna. – Germany 88 schwarz-weiß Abbildungen: apress open, 2015. – 241 с.
 29. В.В.Вьюгин. Математические основы теории машинного обучения и прогнозирования / В.В.Вьюгин. – Москва, 2013. – 387 с.

30. Modeling Hardness of Polycrystalline Materials and Bulk Metallic Glasses / Xing-Qui Chen, Naiyang niu, Dianzhong Li, Yiyi Li. – China: Shenyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, 2011. – 1275 с. – (72).
31. Yoav Freund. A Short Introduction to Boostin / Yoav Freund, Robert E. Schapire – USA: AT&T Labs Researc, 1999. – (Journal of Japanese Society for Artificial Intelligence). – С. 771–780.
32. John J. Gilman. Chemistry and Physics of Mechanical Hardness / John J. Gilman. – USA: John Wiley & Sons Inc, 2009. – 232 с.
33. Peter Flach. Machine Learning The Art and Science of Algorithms that Make Sense of Data / Peter Flach. – USA: Cambridge University Press. – 409 с.
34. <https://www.gordonengland.co.uk/hardness/vickers.htm>
35. David Roylance. Mechanical properties of materials / David Roylance. – USA: MIT 2008, 2008. – 128 с.
36. Characterization, corrosion resistance and hardness of rapidly solidified Ni-Nb alloys / Conrado R.M. Afonso, Katherine Martinez-Orozco, Vicente Amigo та ін.] // Journal of Alloys and compounds. – 2020. – №829. – С. 12
37. ІНСТРУКЦІЯ № 2 з охорони праці під час експлуатації персональних електронно - обчислювальних машин – Суми. Україна: Міністерство освіти і науки України Сумський Державний Університет, 2012. – 7 с.