

Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Кваліфікаційна наукова праця на правах рукопису

КРИСЕНКО ПАВЛО ІГОРОВИЧ

УДК 538.9, 539.2, 539.3, 620.3,

ДИСЕРТАЦІЯ
**ЗБАГАЧЕННЯ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ
МЕТАМАТЕРІАЛІВ**

Спеціальність 153 – Мікро- та наносистемна техніка
Галузь знань 15 – Автоматизація та приладобудування

Подається на здобуття наукового ступеня доктора філософії
Дисертація містить результати власних досліджень. Використання чужих ідей,
результатів і текстів мають посилання на відповідне джерело _____

Науковий керівник Корольок Дмитро Володимирович, доктор фізико-математичних
наук, професор

Київ – 2024

АНОТАЦІЯ

Крисенко П. І. Збагачення даних для прогнозування властивостей метаматеріалів. Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 153 – Мікро- та наносистемна техніка (галузь знань 15 – Автоматизація та приладобудування). – Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, 2024.

Робота присвячена розробці методики збагачення кількості даних для покращення прогнозування коефіцієнта пропускання метаматеріалів на основі топологічної структури та складу, параметрів опромінення поверхні метаматеріалу, а також для покращення можливостей розв'язку оберненої задачі із застосуванням штучних нейронних мереж.

У дисертаційній роботі досліджено науково-прикладні аспекти використання згорткових нейронних мереж для розв'язання задач у галузі проектування та прогнозування властивостей метаматеріалів та можливість задання параметрів опромінення поверхні метаматеріалу для збільшення кількості доступних даних

У сучасних дослідженнях у галузі науки та техніки глибоке навчання, зокрема застосування згорткових нейронних мереж, виявляється дієвим засобом розробки ефективних методів аналізу та прогнозування властивостей метаматеріалів у контексті їх дизайну. Цей підхід відзначається своєю легкістю в реалізації, доступністю відповідних даних, а також високою швидкістю обчислень порівняно з традиційними методами, що базуються на фізичних законах. Розуміється, що нейронні мережі є складними в реалізації та мають свої обмеження, такі як потреба у великій кількості даних, складність оптимізації та постановки завдань. Проте, з розвитком глибокого

навчання, ці обмеження поступово зменшуються, забезпечуючи більший доступ до цього підходу.

Використання нейронних мереж у дизайні метаматеріалів передбачає їх застосування для знаходження матеріалів та топології метаматеріалів за заданими характеристиками. Цей процес починається з навчання нейронних мереж на основі великої кількості вхідних даних, які включають параметри метаматеріалу, такі як розмір, форма та склад, і бажані властивості, такі як поглинання, відбивання та пропускання світла. Після завершення навчання нейронна мережа може використовуватися для прогнозування властивостей метаматеріалів на основі нових вхідних даних, тобто передбачати, як метаматеріал відреагує на нові електромагнітні хвилі.

Завдяки нейронним мережам можна автоматизувати та оптимізувати процес дизайну метаматеріалів, шукати складні залежності між параметрами матеріалу та функціональним відгуком, що є важко досяжним для традиційних методів. Важливо також відзначити можливість аналізувати дані про склад та структуру метаматеріалів та використовувати їх для передбачення електромагнітних властивостей.

У цьому дослідженні використовувалися сучасні знання та підходи до згорткових нейронних мереж для створення оптимальної архітектури нейронної мережі. Для отримання інформації про метаматеріали були використані дані з топологією та складом цих матеріалів. Застосовувалися програмні середовища для розробки цифрового коду та побудови 3D об'єктів метаматеріалів з визначеними властивостями.

Важливим аспектом була можливість прогнозувати коефіцієнт пропускання на основі інформації про структуру та склад метаматеріалів. Зокрема, була проведена конвертація даних з формату ".ply" у формат ".xyzrgb" за допомогою програмного пакета Open3D на базі Python. Цей формат містив важливі дані, які можна було

використовувати для навчання нейронних мереж. Показано важливі вимоги до даних, такі як однорідність та масштабованість.

Розроблено алгоритм для прогнозування властивостей метаматеріалів на основі їх структури та складу та умов експериментальних досліджень, використовуючи згорткову нейронну мережу, та методи кодування додаткової інформації в штучну нейронну мережу. Також був розроблений метод для ефективного зберігання інформації про склад метаматеріалів, що дозволяє передбачати їхні електромагнітні властивості. Обговорено різні цифрові формати, які можна використовувати для зберігання необхідних даних про метаматеріали.

Також досліджено способи представлення властивостей метаматеріалів у вигляді, зручному для навчання згорткової нейронної мережі. Проведено порівняльний аналіз ефективності різних методів, вказавши, що представлення характеристик у вигляді коефіцієнтів полінома, хоча швидше, не є найкращим для прогнозування характеристик метаматеріалів.

Проаналізовано формати збереження інформації про 3D структури. В цій роботі досліджено, що такі формати, як «.3ds», «.obj», «.fbx», «.stl», не підходять, через такі недоліки: відсутність швидкого та простого програмного забезпечення (для ефективного конвертування структур метаматеріалів у формат для навчання штучної нейронної мережі); неможливість зберігання додаткової інформації у канали про структурний склад метаматеріалів; неможливість видалення проприетарної інформації яка кодується та змішується з потрібними для навчання даними.

Розроблено конвеєр навчання тривимірної згорткової нейронної мережі для прогнозування частотних електромагнітних характеристик метаматеріалів на основі структури та складу метаматеріалів та можливістю встановлення інформації про умови експериментальних досліджень. Показані результати прогнозування для різних вибірок даних поділених по масштабу частоти опромінення, з урахуванням інформації о

умовах експериментальних досліджень. Був виконаний аналіз результату навчання та прогнозування. Було запропоновано шляхи покращення якості прогнозування шляхом збільшення вибірки даних для навчання та використання сучасних відходів у глибокому навчанні. Також було запропоновано шляхи вирішення зворотної задачі для генерації структури за заданими параметрами

Ключові слова: нейронна мережа, дизайн метаматеріалів, півперіодні випрямлячі, полярні діелектрик, глибоке навчання, електричні параметри, діелектрична проникність, діелектричні матеріали, метаматеріали, нелінійна модель, діелектрики, математичне моделювання, прогнозування, композитні структури, золото

ABSTRACT

Krysenko P. I. Data enrichment for predicting the properties of metamaterials. Qualifying scientific work on manuscript rights.

Dissertation for obtaining the scientific degree of Doctor of Philosophy in specialty 153 – Micro- and nanosystem engineering (field of knowledge 15 – Automation and instrument engineering). - National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, 2024.

The work is devoted to the development of a data enrichment technique for predicting the transmission coefficient of metamaterials based on topological structure and composition using deep learning techniques, namely, mathematical neural networks using three-dimensional convolution and auxiliary layers

The dissertation researched the scientific and applied aspects of using convolutional neural networks to solve problems in the field of designing and predicting the properties of metamaterials and the possibility of setting the parameters of metamaterial surface irradiation to increase the amount of available data.

In modern research in the field of science and technology, deep learning, in particular the use of convolutional neural networks, turns out to be an effective means of developing effective methods for analyzing and predicting the properties of metamaterials in the context of their design. This approach is noted for its ease of implementation, availability of relevant data, as well as high calculation speed compared to traditional methods based on physical laws. It is understood that neural networks are difficult to implement and have their limitations, such as the need for a large amount of data, the complexity of optimization and task formulation. However, with the development of deep learning, these limitations are gradually reduced, providing greater access to this approach.

The use of neural networks in the design of metamaterials involves their application to find materials and the topology of metamaterials according to given characteristics. This process begins with training neural networks based on a large amount of input data, which includes metamaterial parameters such as size, shape, and composition, and desired properties such as absorption, reflectance, and transmission of light. Once trained, the neural network can be used to predict the properties of metamaterials based on new input data, i.e. predict how the metamaterial will react to new electromagnetic waves.

Thanks to neural networks, it is possible to automate and optimize the design process of metamaterials, to search for complex relationships between material parameters and functional response, which is difficult to achieve with traditional methods. It is also important to note the possibility of analyzing data on the composition and structure of metamaterials and using them to predict electromagnetic properties.

This study used the state-of-the-art knowledge and approaches of convolutional neural networks to create an optimal neural network architecture. Data on the topology and composition of these materials were used to obtain information about metamaterials. Software environments were used to develop digital code and build 3D objects of metamaterials with defined properties.

An important aspect was the ability to predict the transmission coefficient based on information about the structure and composition of metamaterials. In particular, the conversion of data from the ".ply" format to the ".xyzrgb" format was carried out using the Python-based Open3D software package. This format contained important data that could be used to train neural networks. Important data requirements such as homogeneity and scalability are shown.

An algorithm has been developed for predicting the properties of metamaterials based on their structure and composition and the conditions of experimental studies, using a convolutional neural network, and methods of encoding additional information into an

artificial neural network. A method was also developed for efficient storage of information about the composition of metamaterials, which allows predicting their electromagnetic properties. Various digital formats that can be used to store the necessary metadata data are discussed.

Methods of presenting the properties of metamaterials in a form convenient for convolutional neural network training have also been investigated. A comparative analysis of the effectiveness of various methods was carried out, indicating that the representation of characteristics in the form of polynomial coefficients, although faster, is not the best for predicting the characteristics of metamaterials.

Formats for saving information about 3D structures have been analyzed. In this work, it was investigated that such formats as ".3ds", ".obj", ".fbx", ".stl" are not suitable due to the following disadvantages: lack of fast and simple software (for efficient conversion of metamaterial structures into format for training an artificial neural network); impossibility of storing additional information in channels about the structural composition of metamaterials; the impossibility of removing proprietary information that is coded and mixed with data necessary for training.

A three-dimensional convolutional neural network learning pipeline has been developed for predicting the frequency electromagnetic characteristics of metamaterials based on the structure and composition of metamaterials and the possibility of establishing information about the conditions of experimental research. Prediction results are shown for different data samples divided by the scale of exposure frequency, taking into account information about the conditions of experimental studies. An analysis of the learning and forecasting results was performed. Ways to improve the quality of prediction by increasing the training data sample and using modern waste in deep learning were suggested. Ways of solving the inverse problem for generating the structure according to the given parameters were also proposed.

Keywords: neural network, metamaterial design, half-wave rectifiers, polar dielectrics, deep learning, electrical parameters, dielectric permittivity, dielectric materials, metamaterials, nonlinear model, dielectrics, mathematical modeling, forecasting, composite structures, gold

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

Праці, в яких опубліковано основні результати дисертації:

[1] M. O. Zozyuk, D. V. Koroliouk, P. I. Krysenko, A. I. Yurikov and Y. I. Yakymenko, «Prediction of characteristics using a convolutional neural network based on experimental data on the structure and composition of metamaterials», *STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING*, vol. 11, no. 3, pp. 777–787, Jun. 2023. doi: 10.19139/soic-2310-5070-1707.

[2] P. I. Krysenko, M. O. Zoziuk, O. I. Yurikov, D. V. Koroliuk, Yu. I. Yakymenko «Chladni Figures Simulation on a Rectangular Plate», *Microsystems, Electronics and Acoustics*, vol. 26, no. 1, 2021, pp. 241698-1-241698-6, Dec. 2022. doi:10.20535/2523-4455.me.241698.

[3] P. I. Krysenko, M. O. Zoziuk “Using Information about Experimental Conditions to Predict Properties of Metamaterials,” *Microsystems, Electronics and Acoustics*, vol. 28, no. 3, 2023, pp. 287808.1-287808.7, Dec. 2023. doi: 10.20535/2523-4455.me.287808.

Праці, які засвідчують апробацію матеріалів дисертації:

[1] М.О. Зозюк, Д.В. Королюк, П.І. Крисенко, О.І. Юріков, Ю.І. Якименко, «Використання нейронних мереж для прогнозування властивостей метаматеріалів,» *IX Міжнародної наукової конференції імені І. І. Ляшка «Обчислювальна та прикладна математика»*. м. Київ, Україна, 2022.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ.....	13
ВСТУП.....	14
Розділ 1. ОСНОВНІ ПОНЯТТЯ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ.....	20
1.1 Нейронна мережа та її різновиди.....	20
1.1.1 Штучні нейрони.....	21
1.1.2 Повнозв'язна нейронна мережа.....	23
1.1.3 Згорткова нейронна мережа.....	25
1.2 Типи електромагнітних матеріалів.....	29
1.3 Прогнозування властивостей метаматеріалів за допомогою згорткової нейронної мережі	33
Висновки до розділу 1.....	36
Розділ 2. ПІДГОТОВКА ДАНИХ ПРО ЕЛЕКТРОМАГНІТНІ МЕТАМАТЕРІАЛИ ТА УМОВИ ЕКСПЕРИМЕНТІВ.....	38
2.1 Постановка задачі.....	38
2.2 Способи збору даних про електромагнітні метаматеріали.....	41
2.3 Варіанти представлення інформації про метаматеріали для нейронної мережі ...	44
2.4 Підхід до кодування інформації про структуру, склад і властивості метаматеріалів та умови експерименту	51
Висновки до розділу 2.....	69
Розділ 3. АРХІТЕКТУРА ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ ЕЛЕКТРОМАГНІТНИХ МАТЕРІАЛІВ.....	71
3.1 Постановка задачі.....	71
3.2 Алгоритм програмного забезпечення з підготовки, кодування та представлення даних про метаматеріали та умови вимірювання їх властивостей	76

3.3 Метод закріплення інформації про умови експерименту до структури та складу метаматеріалів.....	90
3.4 Архітектура згорткової нейронної мережі для прогнозування коефіцієнту пропускання в залежності структури, складу та умов вимірювання властивостей метаматеріалів	98
Висновки до розділу 3.....	111
Розділ 4. ПРОГНОЗУВАННЯ КОЕФІЦІЄНТУ ПРОПУСКАННЯ ЕЛЕКТРОМАГНІТНИХ МЕТАМАТЕРІАЛІВ	114
4.1 Постановка задачі.....	114
4.2 Навчання нейронної мережі та тестування нейронної мережі	116
4.3 Аналіз та шляхи покращення отриманих результатів	137
4.4 Перспективи створення оберненого алгоритму для прогнозування структури, складу, в залежності від необхідних властивостей та умов вимірювання метаматеріалів	145
Висновки до розділу 4.....	149
ОСНОВНІ НАУКОВІ ТА ПРАКТИЧНІ РЕЗУЛЬТАТИ І ВИСНОВКИ	151
СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ.....	152
ДОДАТКИ.....	169
ДОДАТОК А. Список публікацій здобувача за темою дисертації	169

ПЕРЕЛІК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ

ШНМ – штучна нейронна мережа

CNN – згортова нейронна мережа (Convolutional neural network)

GNN – графова нейронна мережа (Graph neural network)

RNN – рекурентна нейронна мережа (Recurrent neural network)

FEM – метод скінченних елементів (Finite element method)

API – прикладний програмний інтерфейс (Application programming interface)

ASCII – Американський стандартний код для інформаційного обміну (American Standard Code for Information Interchange)

RGB – адитивна колірна модель (Red, Green, Blue)

MLP – багатошаровий нейрон (Multilayer perceptron)

PCA - метод головних компонент (Principal component analysis)

LSTM – довга короткочасна пам'ять (Long short-term memory)

GRU – вентильні рекурентні вузли (Gated recurrent units)

FCNN – повнозв'язна нейронна мережа (Fully-connected neural network)

ВСТУП

Актуальність

У сучасному світі існує потреба в матеріалах, які мають унікальні властивості, недосяжні для природних матеріалів. З розвитком науки кількість композиційних матеріалів та сплавів збільшилася доступних матеріалів з особливими характеристиками. Дослідники виявили закономірності у процесі створення різних композитів і сплавів, що дозволяє прогнозувати властивості нових матеріалів.

Метаматеріали представляють наступний етап у розвитку штучно створених матеріалів. Вони являють собою матеріали, властивості яких залежать від їх топологічної періодичної структури на різних масштабах. Математичний опис цих матеріалів за допомогою фізичних моделей обумовлений властивостями таких структур. Було запропоновано низку моделей для опису параметрів – електромагнітних, акустичних, і т.д.[1,2], в залежності від того яким параметром потрібно керувати. Розв'язок таких рівнянь не є тривіальною задачею, а тому, існує потреба в універсальному та спрощеному методі для прогнозування властивостей метаматеріалів, який буде брати до уваги окрім топології та матеріалів, ще умови експериментальних досліджень. Одним із таких методів може бути побудований на основі методів глибокого навчання [3, 4].

Одним із таких методів є застосування штучних нейронних мереж, які мають змогу знаходити залежності у вхідних даних та оброблювати інформацію представлено різними типами даних.

Одним з видів нейронних мереж є згорткова нейромережа, спроектована для створення моделей з можливістю визначення локальних залежностей у вхідних даних. У згортковій мережі використовується операція згортки ядра якої можуть мати різну просторову розмірність, що дозволяє вивчати топологічну залежність властивостей

вхідної інформації. З розвитком обчислювальної техніки та збільшенням інтересу дослідників до метаматеріалів були проведені дослідження по використанню штучного інтелекту для аналізу метаматеріалів [5-12]. Зазвичай використовуються дані про двовимірну топологію метаматеріалів та відповідні експериментальні характеристики. Так, деякі дослідження використали спрощену та неповну інформацію про топологію метаматеріалу [13,14].

Переваги використання штучного інтелекту для вирішення задачі прогнозування експериментально вимірних характеристик у порівнянні з класичним фізико-математичним моделюванням у тому, що нейронні мережі здатні знаходити більш глибокі зв'язки між параметрами та топологічною структурою, та використовують менший час для обчислень.

Для використання нейронних мереж існують певні вимоги до даних, на яких модель буде навчатися. Для стабільного та якісного прогнозування штучний інтелект за звичай потребує великої кількості даних та деякої однородності в зібраних та оброблених зразках для навчання.

Були опубліковані роботи, експериментальна інформація з яких може бути використана для навчання нейронної мережі для прогнозування властивостей метаматеріалу [15-18].

Актуальність і новизна дисертаційного дослідження обумовлені відсутністю підходів до прогнозування властивостей метаматеріалів з використанням інформації о умовах експериментальних досліджень та структури метаматеріалів. Необхідністю розширення умов збору даних для навчання штучної нейронної мережі. А також потребою вирішення оберненої задачі, а саме прогнозування структури метаматеріалу за заданими властивостями

Стан розробки даної тематики у вітчизняній та закордонній науці.

У зв'язку з активним використанням метаматеріалів у матеріально-технічній базі існує потреба у розробці інструменту для аналізу та прогнозування.

Використання глибокого навчання вимагає спеціального підходу до підготовки даних у відповідному форматі для використання нейронною мережею. У сучасний період ця область стрімко розвивається, особливо через виявлення нових можливостей глибокого навчання, зокрема згорткових нейронних мереж [19-30].

Зв'язок роботи з науковими програмами, планами, темами.

Дисертаційна робота виконана на кафедрі мікроелектроніки.

Мета та завдання дослідження.

Мета – розробити метод закріплення інформації про умови експериментальних досліджень для прогнозування властивостей метаматеріалів від їх структури та збільшення кількості даних для покращення прогнозування та вирішення оберненої задачі.

Для успішного досягнення мети роботи потрібно було розв'язати такі **завдання**:

- 1) Розробити метод кодування структурної інформації і інформації про склад метаматеріалу у єдиний тензор для використання як вхідної інформації для штучної нейронної мережі.
- 2) Розробити метод збільшення кількості даних шляхом кодування параметрів опромінення для використання їх у навчанні штучної нейронної мережі та уніфікації вибірки даних для експериментів
- 3) Показати можливість використання збільшеної кількості даних для вирішення оберненої задачі, а саме проектування метаматеріалів за заданими властивостями та параметрами опромінення.

- 4) Запропонувати вирішення оберненої задачі за допомогою згорткової штучної нейронної мережі, а також використання уніфікованих даних для навчання методом кодування параметрів опромінення у штучну нейронну мережу.

Об’єкт дослідження – метаматеріали та їх властивості.

Предмет дослідження – процедури збільшення кількості даних для навчання згорткової нейронної мережі за допомогою врахування умов експериментальних досліджень.

Методи дослідження:

Для підготовки і кодування структурної інформації і складу метаматеріалів, інформації о умовах експериментальних досліджень було використано програмне забезпечення програмне забезпечення для тривимірного моделювання.

Для навчання штучного інтелекту було використано методи глибокого навчання та математичної статистики.

Наукова та практична новизна отриманих результатів.

В роботі вперше:

- 1) Розроблено новий метод прогнозування властивостей метаматеріалів за допомогою штучного інтелекту, що відрізняється використанням архітектури, яка побудована на основі тривимірних згорткових операцій для обробки об’ємного представлення структурної інформації метаматеріалу та властивостей електромагнітного випромінювання, завдяки чому стало можливе тренування нейронної мережі на складних структурах матеріалів в нанометровому діапазоні та різних властивостях електромагнітного випромінювання, що знаходиться у терагерцовому діапазоні.
- 2) Розроблено новий метод збільшення кількості даних для навчання штучної нейронної мережі для прогнозування властивостей метаматеріалів, особливістю

якого є використання інформації о властивостях зовнішнього впливу на метаматеріал методом контекстного вектору, що дозволило використати усю доступну експериментальну інформацію з досліджень збільшивши кількість даних з 6-ти пар структура-властивість до 28 пар структура-властивість для навчання, з одних і тих самих джерел даних, а також зменшити середньоквадратичну похибку прогнозування з 0.051 до 0.025.

- 3) Показано, що розроблений метод збільшення кількості даних для навчання, особливістю якого є можливість використання усієї інформації з відкритих досліджень, відкриває шляхи до вирішення оберненої задачі по проектуванні структури метаматеріалу за потрібними властивостями
- 4) Запропоновано шляхи вирішення оберненої задачі, що відрізняються використанням розробленого методу збільшення кількості даних для навчання та адаптацією сучасних підходів до генерації інформації за потрібних умов.

Особистий внесок здобувача.

Був обраний напрямок дослідження, визначено об'єкт та предмет дослідження. Було сформульовану мету та визначено завдання роботи. В ході дослідження було проінтерпретовано та проаналізовано результати дослідження. Була запропонована можливість розроблення системи для прогнозування характеристик метаматеріалів з використанням їх структурної інформації та інформації о умовах експериментальних досліджень. У роботах [31*], [32*], які були опубліковані у співавторстві, було представлення ідей для вирішення задач прогнозування характеристик метаматеріалів. Було розроблено алгоритми кодування інформації о структурі метаматеріалів, інформації о умовах експериментальних досліджень, представлення цієї інформації для штучної нейронної мережі. Було проаналізовано якість прогнозування штучної нейронної мережі, було запропоновано ідеї вирішення зворотної задачі до прогнозування характеристик метаматеріалів за їх структурної інформації та

інформації у мовах експериментальних досліджень. Дисертантом було написано тексти статей та основні розділи у публікаціях [31*], [32*].

*джерела зі списку публікацій (Додаток А).

Апробація результатів дисертації.

Результати отримані в ході роботи над дисертацією були представлені на 1 наукових і науково-практичних конференціях і семінарах:

1) IX Міжнародна наукова конференція імені І. І. Ляшка «Обчислювальна та прикладна математика», присвячена 100-річчю академіка Івана Івановича Ляшка, 10.10.2022.

Публікації.

За результатами досліджень опубліковано 4 наукових праць, з них:

- 3 статті у яких висвітлено основні матеріали дисертації, одна з яких у виданні, віднесеному до другого квартилю (Q2) відповідно до класифікації SCImago Journal and Country Rank або Journal Citations Reports, чи одноосібних монографій, що відповідають зазначеним вимогам, інші у періодичних наукових фахових виданнях категорії Б,
- 1 доповідь та теза на науковій конференції

Структура та обсяг дисертації.

Дисертація складається із вступу, 4 розділів із висновками, загальних висновків, списку використаних літературних джерел зі 122 найменувань на 20 сторінках, 1 додатку. Загальний обсяг дисертації складає 170 сторінки, з яких основний зміст викладений на 129 сторінках, містить 52 рисунки та 4 таблиці.

РОЗДІЛ 1.

ОСНОВНІ ПОНЯТТЯ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ

1.1 Нейронна мережа та її різновиди

Штучна нейромережа (ШНМ) є революційною технологією, яка знаходить все більше застосувань у сучасному світі. Вона імітує роботу нервової системи людини та здатна до навчання на основі великого обсягу даних. ШНМ стає основним інструментом у галузі штучного інтелекту, машинного навчання та обробки даних.

Основна ідея штучної нейромережі полягає в наслідуванні біологічних нейронних мереж у мозку. ШНМ складається зі спеціальних обчислювальних одиниць, які називаються штучними нейронами. Кожен штучний нейрон приймає вхідні дані, обчислює їхню вагову суму, застосовує функцію активації та генерує вихідний сигнал.

Існує кілька типів штучних нейромереж, кожен з яких має свої особливості та застосування. Основні типи включають:

1. **Повнозв'язні нейромережі (перцептрони):** Усі нейрони одного шару пов'язані з усіма нейронами наступного шару. Вони застосовуються для класифікації, регресії та інших завдань[33].
2. **Згорткові нейромережі (CNN):** Використовуються для обробки зображень, розпізнавання об'єктів та візуального розуміння[34].
3. **Рекурентні нейромережі (RNN):** Підходять для аналізу послідовних даних, таких як текст або часові ряди[35].
4. **Сполучені нейромережі (GAN):** Використовують дві нейромережі для створення нового вмісту, такого як зображення чи музика[36].

1.1.1 Штучні нейрони

Штучні нейрони є фундаментальними будівельними блоками штучних неймереж, які імітують роботу біологічних нейронів у мозку людини. Кожен штучний нейрон містить в собі декілька основних компонентів, і їхня взаємодія дозволяє штучним неймережам здійснювати різноманітні обчислення та вирішувати завдання навчання. Схематичний вигляд штучного нейрона зображено на рис. 1.1.

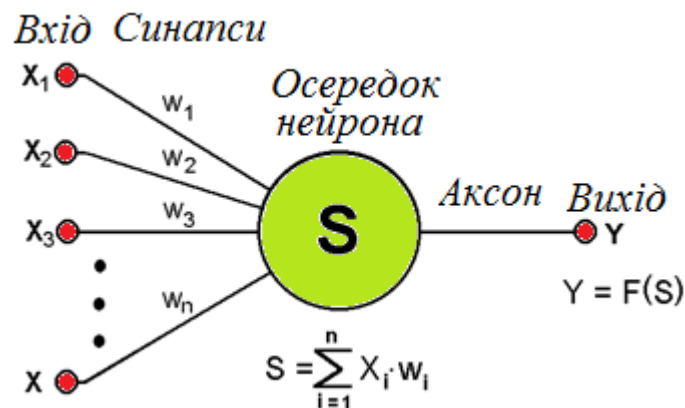


Рисунок 1.1 – Схематичне зображення штучного нейрону[33]

Кожен штучний нейрон має такі основні компоненти:

1. **Вхідні сигнали $x_1, x_2, \dots, x_n$:** Ці значення представляють вхідні дані, які подаються нашому штучному нейрону.
2. **Синапси $w_1, w_2, \dots, w_n$:** Кожному вхідному сигналу x_i призначена вага w_i , яка вказує на важливість цього сигналу для нейрона.

3. **Зважена сума z :** Обчислюється як лінійна комбінація вхідних сигналів та їхніх ваг з додаванням зсуву:

$$S = \sum_{i=1}^n w_i x_i \quad (1.1)$$

4. **Функція активації $F(S)$:** Це нелінійна функція, яка застосовується до зваженої суми S для введення не лінійності в роботу нейрона та генерації вихідного сигналу Y . Деякі з популярних функцій активації включають сигмоїду, гіперболічний тангенс та функцію ReLU (Rectified Linear Unit).
5. **Вихідний сигнал Y :** Після застосування функції активації до зваженої суми отримуємо вихідний сигнал Y :

$$Y = F(S) \quad (1.2)$$

Повний прохід вхідного сигналу через штучний нейрон буде описуватись формулою:

$$Y = F\left(\sum_{i=1}^n w_i x_i\right) \quad (1.3)$$

де, x_i – вхідний сигнал, w_i – вага зв'язку i -го сигналу, F – нелінійна функція активації

1.1.2 Повнозв'язна нейронна мережа

Повнозв'язна нейронна мережа є однією з найбільш поширених та важливих архітектур у глибокому навчанні. Вона здатна до моделювання складних залежностей у даних та вирішення різноманітних завдань, включаючи класифікацію, регресію та генерацію. Основний принцип роботи повнозв'язної нейронної мережі полягає у передачі інформації через шари нейронів, де кожен нейрон зв'язаний з усіма нейронами наступного шару[37].

У повнозв'язній нейронній мережі ми маємо декілька шарів: вхідний, приховані та вихідний, що продемонстровано на рис. 1.2. Кожен нейрону прихованих та вихідних шарів приймає вхідні сигнали, обчислює зважену суму цих сигналів та застосовує функцію активації для отримання вихідного сигналу.

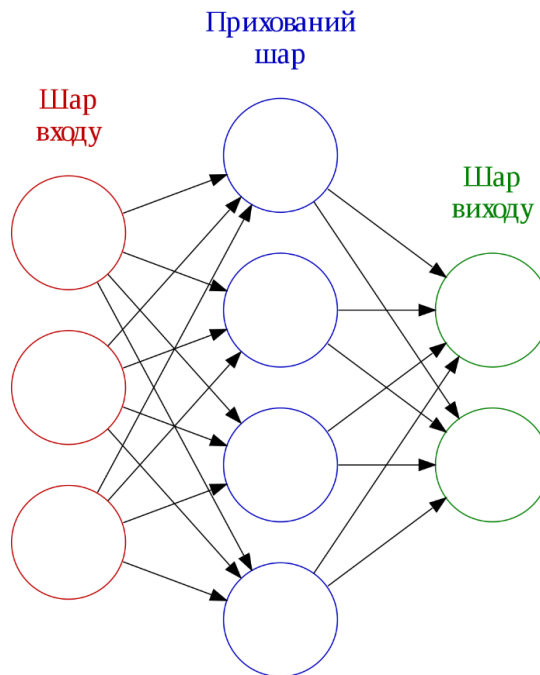


Рисунок 1.2 – Повнозв'язна нейронна мережа[33]

1. **Вхідний шар:** Кількість нейронів у вхідному шарі визначається розміром вхідних даних. Вхідні дані подаються як вектор x розмірністю n , де n - кількість вхідних сигналів.
2. **Приховані шари:** Повнозв'язна нейронна мережа може містити один чи кілька прихованих шарів, кількість нейронів у яких визначається архітектурою мережі. Кожен нейрон j у прихованому шарі має ваги w_{ji} для кожного вхідного нейрона i , а також зсув b_j .
3. **Функція активації:** Кожен нейрон у прихованому та вихідному шарі застосовує функцію активації $f(S)$ до зваженої суми z вхідних сигналів:

$$S_j = \sum_{i=1}^n w_{ji} x_i + b_j \quad (1.4)$$

4. **Вихідний шар:** Вихідний шар містить нейрони, які генерують фінальний результат нейронної мережі. Вихідні дані у генеруються шляхом застосування функції активації до зваженої суми z нейронів у вихідному шарі.

Повнозв'язна нейронна мережа може бути представлена матричними операціями, що спрощує розрахунки та прискорює обчислення. Позначимо матрицю ваг W розмірністю $m \times n$, де m - кількість нейронів у поточному шарі, n - кількість нейронів у попередньому шарі. Вектор зсувів b розмірністю m . Вхідний вектор x розмірністю n . Тоді зважена сума z та вихідний вектор a вираховуються як:

$$a = f(Wx + b) \quad (1.4)$$

де W – матриця вагів

x – вектор вхідного сигналу

b – вектор зсувів

a – вихідний вектор

1.1.3 Згорткова нейронна мережа

Згорткові нейронні мережі (ЗНМ) є потужним інструментом у глибокому навчанні, спеціально призначеним для обробки структурованих даних, таких як зображення та відео. Вони відіграють важливу роль у багатьох сферах, включаючи комп'ютерне зорове сприйняття, медичну діагностику, автономні та напівавтономні системи, а також аналіз даних.

ЗНМ складаються з декількох ключових компонентів, які дозволяють їм виконувати завдання обробки структурованих даних:

1. **Згорткові шари:** Основна ідея ЗНМ полягає у використанні згорткових шарів для виявлення локальних зразків у вхідних даних, таких як шаблони, краї об'єктів тощо. Кожен згортковий шар має набір фільтрів або ядер, які рухаються по вхідному зображенню та виконують операцію згортки.
2. **Операція згортки:** Операція згортки включає обчислення зваженої суми пікселів в околі певної позиції на зображенні з використанням фільтра. Це дозволяє визначити, чи є певний зразок чи ознака присутніми у певному регіоні.

3. **Підвибіркові шари (пулінг):** Після згортки виконується операція підвибіркового шару, або пулінгу, яка допомагає зменшити розмір зображення та підсилює видимість важливих ознак.
4. **Повнозв'язані шари:** Після послідовного виконання згорткових та підвибіркових шарів, зазвичай використовуються повнозв'язні шари, які здійснюють класифікацію або регресію на основі ознак, вивчених з попередніх шарів.

Ці операції зображені на рис.1.3.

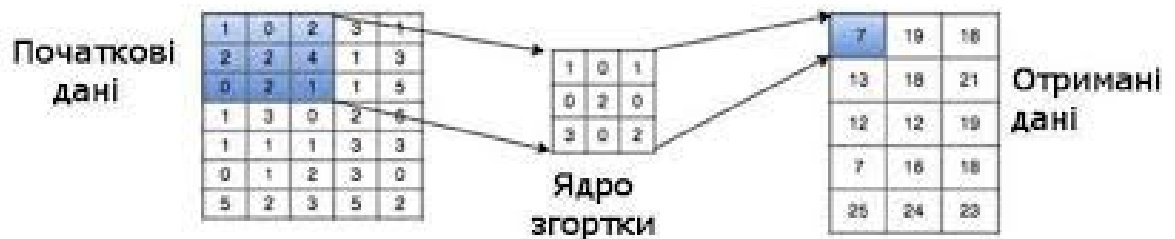


Рисунок 1.3 – Принцип роботи згорткових ядер[34]

Для випадку 2D зображення з кольоровими каналами D_{in} . Нехай I - це вхідне зображення розміром $W_{in} \times H_{in} \times D_{in}$, де W_{in} , H_{in} , D_{in} - відповідно ширина, висота та кількість каналів кольору. Розглянемо зворотну згортку з ядром K розміром $F \times F \times D_{in}$, де F - розмір ядра, D_{in} - глибина вхідного зображення.

Зворотна згортка між вхідним зображенням I та ядром K вираховує виходову карту ознак M розміром $W_{out} \times H_{out} \times D_{out}$, з кроком згортки S та величиною нульового доповнення P та, де

$$W_{out} = \frac{(W_{in} - F + 2P)}{S} + 1 \quad (1.5)$$

$$H_{out} = \frac{(H_{in} - F + 2P)}{S} + 1 \quad (1.6)$$

та D_{out} - кількість ядер.

Виходове значення $M(x,y,d)$ на (x,y) для глибини d вираховується за допомогою операції зворотної згортки та активаційної функції σ :

$$M(x,y,d) = \sigma \left(\sum_{i=0}^{F-1} \sum_{j=0}^{F-1} \sum_{k=0}^{D_{in}-1} I(x+i, y+j, k) \cdot K(i,j,k,d) + b_d \right) \quad (1.4)$$

де $K(i,j,k,d)$ - вага (ядро) на позиції (i,j) для вхідного каналу k та вихідного каналу d , b_d - зсув для каналу d .

Зазвичай ядра згортки використовуються у 3D форматі, а саме ширина, висота та кількість фільтрів. Але, сучасні нейронні мережі можуть використовувати також 2D формат, а саме час та кількість фільтрів. Або 4D формат для обробки просторових даних, які містять додаткову розмірність глибини. Саме такі згортки будуть використовуватися в цій роботі. Приклад архітектури згорткової нейронної мережі представлено на рис.1.4.

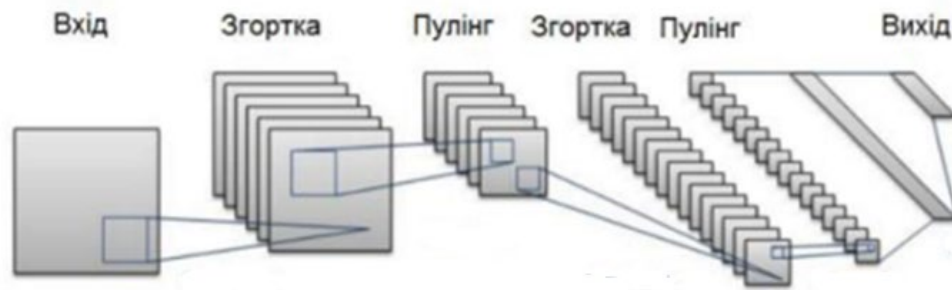


Рисунок 1.4 – Архітектура LeNet [34]

Пряме поширення (feed forward) є основним етапом обчислення в згорткових нейронних мережах (ЗНМ). Цей етап передбачає поширення вхідних даних через різні шари мережі для отримання вихідного результату. Процес прямого поширення яке містить кілька основних етапів, які ми розглянемо детальніше.

Перший шар згорткової нейронної мережі - це вхідний шар, який приймає вхідні дані. Це можуть бути зображення, аудіо сигнали, текстові дані тощо. Вхідні дані розглядаються як матриця або тензор, де кожен елемент відповідає певному значенню вхідного сигналу.

Після вхідного шару слідує згортковий та підвибірковий (пулінг) шари. Згорткові шари виявляють локальні ознаки та шаблони у вхідних даних за допомогою згорткових операцій. Після цього виконується операція підвибіркового шару, яка зменшує розмір отриманої інформації та виділяє найважливіші ознаки.

Після виконання згорткових та підвибіркових шарів можуть використовуватися повнозв'язані (fully connected) шари. Вони з'єднують отримані ознаки від попередніх шарів і виконують подальшу обробку для отримання вихідного результату. У цих шарах використовуються звичайні нейрони з активаційними функціями, які забезпечують нелінійність у моделі.

Останній шар є вихідним, де отримується кінцевий результат прямого поширення. Це може бути класифікаційний результат, числове значення або інша форма вихідних даних, в залежності від завдання мережі. Ці блоки зображені на рис. 1.4.

Процес прямого поширення в ЗНМ можна математично описати наступним чином:

1. Вхідний шар: $\mathbf{X}$ - вхідний тензор (зображення або дані).
2. Згорткові шари: $\mathbf{Z}_i = \text{Conv}(\mathbf{X}, \mathbf{W}_i)$, де $\mathbf{W}_i$ - ядра (фільтри) згорткового шару i .
3. Підвибіркові шари: $\mathbf{P}_i = \text{Pool}(\mathbf{Z}_i)$, де Pool - операція підвибіркового шару.
4. Повнозв'язані шари: $\mathbf{H}_j = \text{ReLU}(\mathbf{W}_j \cdot \mathbf{P}) + \mathbf{b}_j$, де $\mathbf{W}_j$ - ваги повнозв'язаного шару j , $\mathbf{b}_j$ - зсув, ReLU - активаційна функція.
5. Вихідний шар: $\mathbf{O} = \text{Softmax}(\mathbf{W}_{\text{out}} \cdot \mathbf{H}_{\text{last}} + \mathbf{b}_{\text{out}})$, де Softmax - активаційна функція для класифікації.

Процес прямого поширення в згорткових нейронних мережах допомагає здійснити подальшу обробку та аналіз вхідних даних для отримання корисної інформації або вихідного результату.

1.2 Типи електромагнітних математеріалів

Електромагнітні метаматеріали є унікальними структурами, які дозволяють маніпулювати електромагнітними хвилями на мікро та нанорівнях. Вони відрізняються від природних матеріалів і здатні досягати надзвичайних електромагнітних властивостей, що відкриває двері до різноманітних застосувань.

1. Діелектричні метаматеріали:

Діелектричні метаматеріали - це клас метаматеріалів, які складаються зі штучно створених структур, зазвичай на мікро- та нанорівні, і відрізняються від природних діелектриків своїми електромагнітними властивостями. Діелектричні метаматеріали використовуються для маніпулювання розповсюдженням електромагнітних хвиль у широкому спектрі частот, від радіохвиль до світла.

Основні принципи та властивості діелектричних метаматеріалів:

Структура: Діелектричні метаматеріали складаються з мікроструктур, які мають періодичну або неперіодичну геометрію. Ці мікроструктури можуть бути виготовлені з діелектричних матеріалів, таких як пластик або кераміка. Діелектрична проникність і магнітна проникність:

- від'ємна діелектрична проникність: Діелектричні метаматеріали можуть мати від'ємну діелектричну проникність, що означає, що вони розсіюють світло навпаки до того, як це робить природний матеріал;
- придушення втрат: метаматеріали можуть бути спроектовані так, щоб подавити втрати електромагнітної енергії, зокрема, втрати в радіо- та мікрохвильовому діапазоні;
- надзвичайна дисперсія: діелектричні метаматеріали можуть відзначатися надзвичайною дисперсією, коли їх властивості змінюються значно в залежності від частоти[38].

Застосування:

- лінзи та підсилювачі для мікрохвильового та оптичного випромінювання: Діелектричні метаматеріали використовуються для створення плоских лінз та антен, які можуть покращувати роздільну здатність джерела світла. [39, 40]. Метаматеріали можуть бути використані для створення мета-лінз, які

можуть усувати аберації та покращувати роздільну здатність в оптичних системах[40];

- інфрачервоні камуфляжні покриття: діелектричні метаматеріали використовуються для створення камуфляжних покриттів, які можуть робити об'єкти невидимими для інфрачервоних камер із засобами нічного бачення[41] [42];
- антени та комунікації: діелектричні метаматеріали можуть бути використані для створення ультра тонких антен, які знаходять застосування в багатьох сучасних комунікаційних системах[43];
- терагерцова технологія: діелектричні метаматеріали допомагають в розробці терагерцової технології, яка використовується для зображення і діагностики матеріалів та біологічних тканин у терагерцовому діапазоні[44].

Діелектричні метаматеріали мають великий потенціал для різних застосувань, і їх дослідження триває в багатьох наукових лабораторіях та промислових дослідницьких центрах. Вони відкривають нові можливості для розвитку електроніки, оптики, комунікаційних технологій та інших галузей, де важливе керування електромагнітними хвилями. [38]

1. Плазмонні метаматеріали:

Плазмонні метаматеріали - це клас матеріалів, які використовують властивості плазмонів для керування електромагнітними хвильовими процесами. Плазмони - це колективні коливання електронів у металевих структурах, які виникають при взаємодії електромагнітного випромінювання з металевою поверхнею. Ці коливання можуть бути використані для маніпулювання світлом на нанометровому рівні, що робить плазмонні метаматеріали потужними інструментами для створення нових оптичних та фотонних пристроїв.

Основні характеристики плазмонних метаматеріалів:

- від'ємний індекс рефракції: плазмонні метаматеріали можуть мати від'ємний індекс рефракції. Це відкриває можливості для створення невидимих об'єктів або лінз, які можуть фокусувати світло на нанометровому рівні[45];
- велика поглинаюча здатність: плазмонні метаматеріали можуть мати високу поглинаючу здатність в обраному діапазоні частот. Це робить їх ідеальними для розробки сенсорів, які можуть виявляти навіть дуже низькі концентрації молекул[46] [47];
- керована дисперсія: плазмонні метаматеріали дозволяють контролювати швидкість поширення світла в матеріалах. Це може бути корисним для створення оптичних комунікаційних пристроїв та оптоволокон з покращеною ефективністю[48];
- плоскі лінзи та імплантація в офтальмології: плазмонні метаматеріали також досліджуються для створення плоских лінз та інших оптичних пристроїв, які можуть замінити традиційні скляні лінзи у медицині, зокрема в офтальмології[49];
- терагерцова та метаматеріалів в мікрохвильовому діапазоні: плазмонні метаматеріали також мають застосування в терагерцовому та мікрохвильовому діапазоні, де вони можуть бути використані для створення нових пристроїв для обробки сигналів та зображень[50].

Загалом, плазмонні метаматеріали є дуже цікавою галуззю досліджень, яка має потенціал для революції в різних галузях науки і технології, зокрема в оптиці, фотоніці, сенсоріці та медицині.

2. Магнітні метаматеріали:

Магнітні метаматеріали мають властивості, які залежать від магнітних полів. Вони використовуються для створення ефективних антен та пристроїв для магнітного резонансу у медицині.

Приклади використання електромагнітних метаматеріалів:

- інфрачервоний діапазон: метаматеріали в цьому діапазоні використовуються для розробки теплових камуфляжних покриттів, які роблять об'єкти невидимими для інфрачервоних камер із засобами нічного бачення[51] [52];
- мікрохвильовий діапазон: електромагнітні метаматеріали в мікрохвильовому діапазоні використовуються для розробки ультра тонких антен та відбиваючих поверхонь, що знаходять застосування в сучасних комунікаційних системах[53] [54].

Електромагнітні метаматеріали також знаходять застосування в медицині для розробки нових методів діагностики та лікування, таких як образи та терапія з використанням плазмонних метаматеріалів. [55]. [56]. [57]

Застосування електромагнітних метаматеріалів є широким і постійно розширюється завдяки постійному розвитку технологій та досліджень в цій області. Вони відкривають нові можливості у різних галузях, включаючи електроніку, оптику, акустику, комунікації та медицину.

1.3 Прогнозування властивостей метаматеріалів за допомогою згорткової нейронної мережі

Останніми роками активно вивчається сфера застосування глибокого навчання для дизайну метаматеріалів. Вчені запропонували різноманітні методи, включаючи як класичні нейронні мережі, так і згорткові моделі[58-60]. Більшість досліджень зосереджена на створенні метаматеріалів для метаповерхонь, де вхідні дані зображені у двовимірному форматі. У таких моделях третя координата може бути закодована через яскравість пікселів або додатковий канал у вхідних даних для нейронних мереж.

Прогнозування електромагнітних характеристик метаматеріалів за допомогою нейронних мереж має свої особливості. Оскільки цифрова симуляція характеристик метаматеріалів може бути великою та ресурсномісткою задачею, для навчання нейронних мереж необхідно мати прості дані в базі даних. Саме тому важливо обирати елементи бази даних, які легко підготувати та які мають відповідні характеристики.

Для навчання нейронної мережі ідеально підходить симуляція характеристик двовимірної метаповерхні. Використання тривимірної симуляції для задач глибокого навчання є більш складним завданням, як підтверджено дослідженнями [61], [62]. Однак існує спосіб використання інформації про об'єм як додатковий канал кольору або яскравості пікселя на двовимірному зображенні [63].

Для розв'язку проблеми з частковою відсутністю даних і спрощення представлення інформації запропонована ідея використання сферичних наночасток, які утворюють метаатоми. Їхні властивості стають вхідними даними для нейронної мережі [64]. Симулювання характеристик таких сфер (метаатомів) є простішою задачею. Подібний підхід застосовувався також в інших роботах [65], [66], де метаматеріали змодельовані як шари матеріалів з властивостями та товщинами.

Обидва підходи (з використанням сфер або площин) дозволяють генерувати метаматеріали з потрібними характеристиками, але обмежені складністю поверхневих структур. Розв'язок задачі генерації структур за допомогою нейронних мереж вимагає додаткових досліджень.

В контексті прогнозування електромагнітних властивостей метаматеріалів з двовимірними зображеннями, симуляція характеристик метаповерхонь менш складна порівняно з об'ємними структурами. Ця категорія досліджень поділяється на два напрямки. Перший напрямок використовує операції згортки (CNN)[67][68], інший - нейронні мережі без цих операцій [69-72]. Застосування згорткової нейромережі доцільне, коли розмір двовимірного зображення великий, а процес растеризації не є ефективним через втрату інформації про структуру. Згорткова нейромережа дозволяє зменшити кількість пікселів, при цьому зберігаючи інформацію про структуру. Схематичний вигляд такої операції продемонстровано на рис. 1.5 Цей підхід також може бути використаний для дизайну молекул та кристалічних структур, де структура відображається через відображення 3D-структури[73].

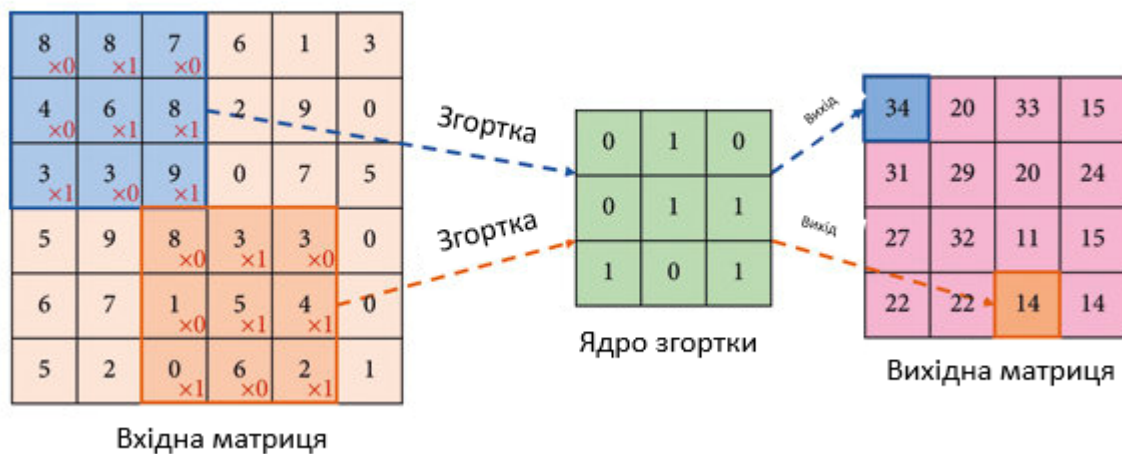


Рисунок 1.5 – Схематичне зображення зменшення розмірності за допомогою згорткового ядра[34]

Для подолання цих труднощів важливо використовувати спрощені структури, які можуть бути симульовані в невеликому масштабі. Такі структури можуть включати метаатоми, метаповерхні з обмеженим розміром або інші елементи. Інший підхід

полягає в створенні структур, які складаються з «немасивних» елементів, таких як шари матеріалів або невеликі структури. Також можна представити властивості великих структур, розділивши їх на менші підструктури, властивості яких вже вивчені. З таким підходом можна симулювати характеристики великих структур, використовуючи знання про характеристики їх складових частин. Більшість робіт [69-73] в різній мірі використовують ці підходи.

Стосовно представлення графіків характеристик або спектрів, це відносно проста задача. Її можна вирішити за допомогою різних методів, таких як вибір ключових точок на графіку, конвертування графіка в поліноміальні коефіцієнти або вираз графіка у вигляді функції.

Однорідність даних також грає важливу роль у якості прогнозування. Дані повинні бути зібрані при однакових умовах та параметрах отримання. Параметри, які впливають на нейронну мережу, повинні бути враховані в задачі для прогнозування. Якщо є параметри, що не мають впливу на мережу, вони повинні залишатися постійними для всіх даних. Хоча це може бути важко досяжною метою, важливо зменшити вплив параметрів, які впливають на характеристики, до мінімуму.

Висновки до розділу 1

Використання того чи іншого основного типу штучної нейронної мережі залежить від задачі котру потрібно розв'язати методом глибокого навчання. Для задач, в яких у вхідних даних існує просторова та локальна залежність, використовуються згорткові нейронні мережі. Їх основний базовий блок має можливість знаходити та вивчати локальні залежності в даних за рахунок операції згортки. Існує декілька реалізацій такої штучної нейронної мережі в залежності від розмірності вхідних даних. Для

двовимірного типу даних, таких як картинки або двовимірні структури, розроблено архітектури з двовимірним ядром згортки. Для задач з тривимірним типом даних, таких як об'ємні зображення або тривимірні структури, є можливість використання вже розроблених та досліджених архітектур для двовимірних структур зі зміною ядра згортки на тривимірне.

Для проектування архітектури існує потреба в виборі гіперпараметрів штучної нейронної мережі, таких як функція активації, розмірність вагів

Електромагнітні метаматеріали мають значний потенціал у використанні в сучасних приладах та пристроях. Можуть використовуватися для керованого розсіювання та для створення ефекту невидимості. Їх революційний потенціал може застосовуватися у різних галузях, від оптики до телекомунікацій.

Існує потреба у створенні інструментів на основі глибокого навчання для проектування метаматеріалів. В існуючих дослідженнях було використано різні типи штучних нейронних мереж для прогнозування властивостей метаматеріалів. В цих дослідженнях не використовувалися умови експериментальних досліджень, що свідчить про те, що дослідники були обмежені у кількості даних, або ці характеристики не бралися до уваги, що може призводити до погіршення якості прогнозування.

У підсумку, використання штучних нейронних мереж для аналізу та дизайну електромагнітних метаматеріалів відкриває нові горизонти для дослідження та інновацій. Проте, для досягнення більш точних та надійних результатів необхідно глибше дослідження та вдосконалення алгоритмів, що використовуються в цих роботах.

РОЗДІЛ 2.

ПІДГОТОВКА ДАНИХ ПРО ЕЛЕКТРОМАГНІТНІ МЕТАМАТЕРІАЛИ ТА УМОВИ ЕКСПЕРИМЕНТІВ

2.1 Постановка задачі

Збір даних про структуру метаматеріалів є ключовим етапом дослідження, оскільки від цих даних залежить точність та ефективність подальших аналізів та моделювання властивостей матеріалів. Збір даних повинен бути систематичним, добре структурованим та дослідженим з урахуванням різних властивостей метаматеріалів. Нижче будуть описані основні вимоги до збору даних:

Перш за все, необхідно чітко визначити мету збору даних. Це може бути прогнозування оптичних, електромагнітних, акустичних або інших властивостей метаматеріалів. Визначення мети допоможе уточнити, які конкретні параметри структури слід виміряти або вивчити.

Вибір параметрів структури метаматеріалів є ключовим кроком. Це можуть бути розміри, форма юніт-клітини, орієнтація, відстані між елементами та інші параметри, які впливають на їх властивості. Важливо вибрати ті параметри, які є релевантними для конкретних характеристик метаматеріалів.

Збір даних може включати експериментальне вимірювання та моделювання структур. Експерименти можуть включати вимірювання мікроскопічних структур за допомогою рентгенівської дифракції, електронної мікроскопії тощо. Моделювання

може проводитися з використанням методів обмежених елементів, методів скінченних різниць тощо.

Метаматеріали можуть мати різні структурні конфігурації, що впливають на їх властивості. Важливо врахувати різні можливі варіанти структурної організації та забезпечити достатню вибірковість при зборі даних.

Метаматеріали можуть виявляти свої унікальні властивості на різних масштабах, від нанометрів до міліметрів. Важливо враховувати масштабні особливості структур при зборі даних.

Зібрані дані потрібно обробити та проаналізувати. Це може включати використання статистичних методів, машинного навчання або фізичних моделей для встановлення зв'язку між структурою та властивостями.

Для підтвердження достовірності зібраних даних і моделей важливо виконати перевірку на незалежних наборах даних або порівняти прогнози з експериментальними результатами.

Усі кроки збору даних, методи вимірювання, моделювання та аналізу повинні бути детально задокументовані. Це допоможе зрозуміти методологію та перевірити достовірність результатів.

Дані, отримані в результаті дослідження, повинні бути доступні для інших науковців і дослідників у відкритому форматі. Це сприятиме спільному розвитку досліджень у галузі метаматеріалів.

Зі зміною технологій та методів збору даних важливо постійно оновлювати підходи до збору даних та моделювання, щоб використовувати найсучасніші методи.

Врахування цих вимог при зборі даних про структуру метаматеріалів допоможе забезпечити якісні та надійні дані для подальшого дослідження та розробки нових матеріалів з унікальними властивостями.

Молекулярну структуру матеріалів можна відтворити у цифровому форматі, який може бути підданий аналізу штучного інтелекту, і встановити відповідність між цим цифровим відображенням та ключовими параметрами метаматеріалів, які підлягають вивченню. Головна мета полягає в установленні чіткої відповідності між інформацією про структуру матеріалу і фізичними характеристиками цієї структури. Звичайно, існує безліч факторів, що впливають на експериментальні фізичні характеристики. Однак важливо розуміти, що більша кількість доступної інформації про структуру та інші параметри, що впливають, призводить до зменшення необхідного обсягу вибірки для навчання нейронної мережі.

В якості вхідної інформації до нейронної мережі був обраний структурний рендер метаматеріалу який зберігає просторову та матеріальну інформацію, що збільшить точність прогнозування, але й збільшить вимоги до архітектури такої штучної нейронної мережі.

Враховуючи складність електромагнітних явищ у метаматеріалах, збір даних повинен включати наступні аспекти:

Електромагнітні властивості метаматеріалів можуть виявлятися на різних діапазонах частот, включаючи радіочастоти, мікрохвилі, інфрачервоне та видиме випромінювання. Збір даних повинен включати спектральний аналіз у відповідних діапазонах для повного розуміння властивостей матеріалу.

Врахування орієнтації та поляризації електромагнітних хвиль відносно структури метаматеріалу допомагає зрозуміти анізотропію та напрямленість його властивостей.

Збір даних повинен передбачати вимірювання в різних напрямках та для різних поляризацій.

Електромагнітні властивості метаматеріалів, такі як комплексний імпеданс, коефіцієнт відбивання, пропускання та поглинання, є важливими характеристиками для визначення їх поведінки в реальних умовах. Збір даних повинен включати точні вимірювання цих параметрів.

Товщина шару метаматеріалу та мікроструктура юніт-клітин можуть суттєво впливати на його електромагнітні властивості. Збір даних повинен включати вимірювання для різних товщин та структурних конфігурацій.

Електромагнітні властивості метаматеріалів можуть змінюватися під впливом зовнішніх факторів, таких як температура, вологість, механічне навантаження тощо. Збір даних повинен включати вимірювання під різними умовами.

Для підтвердження отриманих даних важливо використовувати обчислювальні моделі та порівнювати їхні результати з експериментальними даними.

Вибір відповідних джерел діапазону (вимірювальних приладів) для проведення експериментів важливо для отримання точних і достовірних даних.

2.2 Способи збору даних про електромагнітні метаматеріали

Збір даних про метаматеріали може бути здійснений за допомогою різноманітних джерел, які охоплюють експериментальні, обчислювальні та літературні джерела.

Збір структурної інформації та даних щодо електромагнітних властивостей метаматеріалів відіграє важливу роль у розвитку цих унікальних матеріалів та їх застосуваннях. Джерела збору даних про метаматеріали, такі як наукові статті, є

важливим інформаційним ресурсом для дослідників, які вивчають їхні властивості та застосування. Ці статті надають доступ до результатів проведених досліджень, експериментів та моделювання, що допомагає розширити знання про метаматеріали

Збір структурної інформації вимагає від підрозділу дослідників ретельно розробленого підходу. У статті[74], автори розглядають використання мікроскопічних методів, таких як скануюча електронна мікроскопія (SEM) та атомно-силова мікроскопія (AFM), для вимірювання деталей мікроструктури метаматеріалів. Ці методи дозволяють отримувати високороздільну інформацію про розміри та форму структурних елементів. Моделювання рентгенівської дифракції, як розглядається в статті[75], може допомогти реконструювати кристалічну структуру метаматеріалів, що є важливим для розуміння їхнього поведіння.

Щодо даних про електромагнітні властивості, стаття[76] пропонує ретельний аналіз різних методів вимірювання, включаючи вимірювання коефіцієнтів відбиття та пропускання. Вона наголошує на важливості збору даних для різних напрямків поляризації та різних діапазонів частот. В роботі[77] оглядається вимірювання комплексних параметрів, таких як комплексний імпеданс, що вказують на ефективність матеріалу у конкретних умовах.

З іншого боку, обчислювальне моделювання також відіграє важливу роль у зборі даних. Моделі методу скінченних елементів (FEM) та методу дискретних дірок (DDM), описані в статті "Finite Element Simulation of Electromagnetic Metamaterials: Modeling the Effective Material Properties" (Chen et al., 2017), дозволяють оцінити ефективні електромагнітні характеристики метаматеріалів.

Згідно з обговоренням у статтях, важливим аспектом є те, що поєднання експериментальних даних та результатів моделювання дозволяє отримати більш повне розуміння електромагнітних властивостей метаматеріалів. Такий підхід сприяє

розвитку точних моделей, які можуть використовуватися для передбачення поведінки матеріалів у різних умовах та діапазонах частот.

Самостійне розв'язання рівнянь Максвелла є потужним інструментом для дослідження властивостей метаматеріалів. Воно дозволяє розглядати різні аспекти їхнього поведінки та взаємодії з електромагнітним полем. Такий підхід сприяє розвитку нових знань, що можуть впливати на розробку передових технологій, включаючи антени, датчики, оптичні пристрої та багато інших застосувань.

Моделі, створені на основі рівнянь Максвелла, можуть бути використані для порівняння з експериментальними даними, що допомагає підтвердити точність моделі та розуміння процесів, які відбуваються в матеріалі[78].

Цей підхід дозволяє аналізувати та моделювати поведінку електромагнітних полів у структурах зі складною геометрією та взаємодією між компонентами метаматеріалу. Розглянемо конкретні переваги та можливості використання самостійного вирішення рівнянь Максвелла для збору даних про властивості метаматеріалів.

Розв'язання рівнянь Максвелла дозволяє прогнозувати ефективні електромагнітні властивості метаматеріалів при заданому спектрі частот та формі структури. Це важливо для дизайну матеріалів з певними електромагнітними характеристиками, такими як від'ємна дисперсія, антирезонансі явища тощо.

Самостійне вирішення рівнянь Максвелла дозволяє дослідникам проводити детальний аналіз структури метаматеріалу та впливу геометричних параметрів на його електромагнітні властивості. Завдяки цьому підходу можна вивчати різні аспекти, такі як резонансні явища, анізотропію, вплив зміни розмірів та орієнтацій структур на спектральні характеристики[78].

2.3 Варіанти представлення інформації про метаматеріали для нейронної мережі

Нейронні мережі можуть використовувати різноманітні типи даних, як наприклад числові дані, які є найпоширенішим типом даних для нейронних мереж, що можуть бути використані у вигляді векторів або матриць, як наприклад, числа з датасету, розміри об'єктів, вимірювання сенсорів, фізичні параметри та багато іншого. тексту до зображень та числових значень.

Також використовуються зображення, які широко використовуються для завдань комп'ютерного зору. Вони можуть бути у форматі RGB (червоний, зелений, синій) або в відтінках сірого. Зображення можуть бути подані як тривимірні матриці, де кожний піксель має відповідні значення кольору або інтенсивності.

З розвитком LLM(великі лінгвістичні моделі) популярні стають нейромережі для обробки природньої мови, для їх використання потребується представлення тексту у цифровому вигляді. Для цього використовується послідовність слів, речень, або навіть векторів слів (word embeddings)[79], які представляють слова у векторному просторі. Вибір конкретного типу даних залежить від задачі, яку потрібно вирішити.

Одним з найпростіших типів для аналізу є числові послідовності. Такі типи даних використовуються найчастіше для прогнозування наступних значень у числовому ряді та для класифікації. Існують декілька способів їх представлення для нейронної мережі:

1. Векторний підхід (Vector Representation): У цьому підході кожен часовий ряд може бути представлений у вигляді вектора, де кожен компонент відповідає значенню

ряду в конкретний момент часу[80]. Для багатовимірних часових рядів (наприклад, багатовимірних сигналів) цей підхід особливо ефективний.

$$X_t = [x_{t,1}, x_{t,2}, \dots, x_{t,n}] \quad (2.1)$$

де X_t - вектор значень часового ряду у момент часу

$x_{t,i}$ - значення i -го компонента ряду у момент часу t

n - кількість компонент.

2. Спектральний аналіз (Spectral Analysis): Спектральний аналіз перетворює часовий ряд у частотний домен за допомогою дискретного перетворення Фур'є (ДПФ) чи інших спектральних методів[81]. Це дозволяє виявити панівні частоти та шаблони в ряді.

$$X(f) = \sum_{t=0}^{N-1} x(t) e^{-i2\pi f t} \quad (2.2)$$

де $X(f)$ - спектр часового ряду

$x(t)$ - значення ряду у момент часу t

f - частота

N - кількість відліків.

3. Сплайн-інтерполяція (Spline Interpolation): Сплайн-інтерполяція дозволяє інтерполювати складний часовий ряд гладкою функцією, яка може бути використана для аналізу та прогнозу[82]. Наприклад, кубічні сплайни інтерполюють ряд за допомогою кубічних функцій у кожному інтервалі.

$$S(x) = a + bx + cx^2 + dx^3 \quad (2.3)$$

де $S(x)$ - апроксимація ряду кубічним сплайном

a, b, c, d - коефіцієнти сплайну

x - значення ряду.

4. Представлення у вигляді послідовності (Sequence Representation): У глибокому навчанні можна використовувати послідовність підходу, де вхідні дані подаються в мережу покроково. Наприклад, варіант "Sequence-to-Sequence" може бути використаний для прогнозування майбутніх значень ряду.

$$X = (x_1, x_2, \dots, x_T) \quad (2.4)$$

де X - послідовність значень часового ряду

x_t - значення ряду в момент часу t

T - кількість кроків часу.

Представлення зображень є ключовим аспектом для успішного застосування нейронних мереж у завданнях комп'ютерного зору. Зображення можуть бути великими і складатися з багатьох пікселів як на рис. 2.1, тому їхнє оптимальне представлення допомагає нейронній мережі здійснювати ефективну обробку. Є декілька способів представлення структурної інформації зображення:

Найпростіший спосіб представлення зображення полягає в поданні значень пікселів у вигляді матриці.

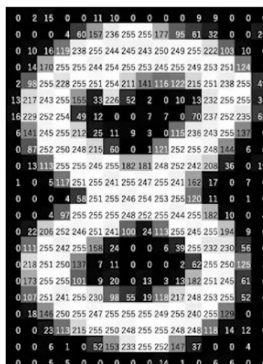


Рисунок 2.1. Піксельне представлення зображення[36]

Якщо зображення має розмір $M \times N$ пікселів та є чорно-білим, то воно може бути представлене як матриця X розміром $M \times N$, де кожний елемент x_{ij} відображає інтенсивність пікселя у i -му рядку та j -му стовпці.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1} & x_{M2} & \dots & x_{MN} \end{bmatrix} \quad (2.5)$$

2. Кольорове зображення: Кольорові зображення, зазвичай в форматі RGB (Red, Green, Blue), можуть бути представлені у вигляді тривимірних матриць, що представлено на рис. 2.2.

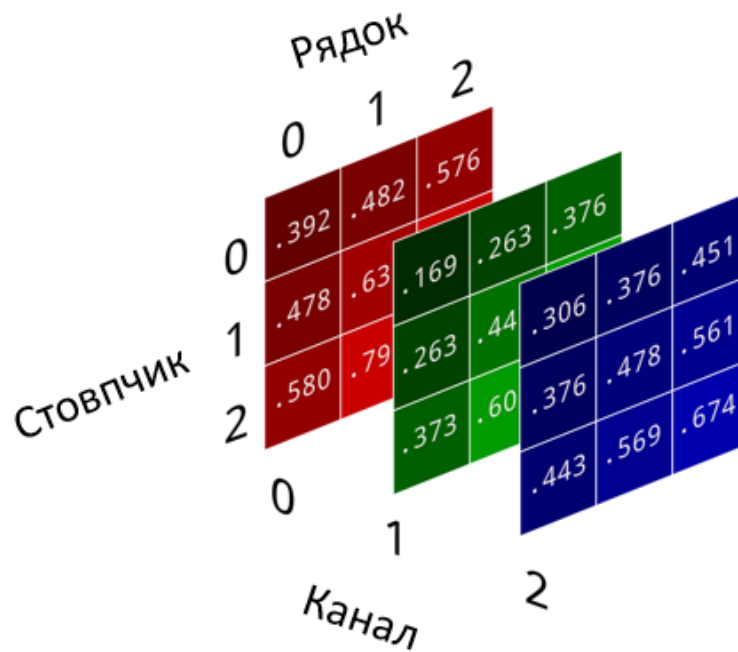


Рисунок 2.2 – Піксельне представлення з RGB каналами[36]

Якщо зображення має розмір $M \times N$ пікселів та містить кольорові компоненти, то воно може бути представлене як тривимірна матриця X розміром $M \times N \times 3$, де третій розмір відповідає каналам R, G та B.

$$X = \begin{bmatrix} x_{11}^R & x_{11}^G & x_{11}^B & \cdots & x_{1N}^R & x_{1N}^G & x_{1N}^B \\ x_{21}^R & x_{21}^G & x_{21}^B & \cdots & x_{2N}^R & x_{2N}^G & x_{2N}^B \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{M1}^R & x_{M1}^G & x_{M1}^B & \cdots & x_{MN}^R & x_{MN}^G & x_{MN}^B \end{bmatrix} \quad (2.6)$$

3. Векторне представлення (Flattened Representation): Інколи зображення може бути "сплющено" у вектор, в якому всі пікселі послідовно розташовані. Це може бути корисним для використання звичайних нейронних мереж, які приймають на вхід вектори.

$$X = [x_{11}, x_{12}, \dots, x_{1N}, x_{21}, x_{22}, \dots, x_{MN}] \quad (2.7)$$

4. Використання згорткових шарів: Для ефективної обробки зображень нейронні мережі можуть використовувати згорткові шари, які автоматично виявляють різноманітні функції та ознаки в зображенні. У цьому випадку зображення буде оброблено з використанням згорткових ядер, які пропускаються по всьому зображенню та вилучають функції.

$$Y = \text{Convolutional}(X, \text{kernel}) \quad (2.8)$$

де Y - оброблене зображення, X - початкове зображення, kernel - згорткове ядро.

Представлення зображень грає важливу роль у роботі нейронних мереж для завдань комп'ютерного зору, і вибір оптимального підходу залежить від конкретної задачі та архітектури мережі.

Представлення тривимірних (3D) структур для нейронних мереж є важливим завданням у розв'язку задач, пов'язаних з молекулярною біологією, хімією, матеріалознавством та іншими областями. Такі структури можуть бути дуже складними та багатовимірними, тому їх оптимальне представлення допомагає нейронним мережам аналізувати та моделювати їх властивості. Розглянемо декілька способів представлення 3D структур:

1. Використання графічних вкладень (Graph Embeddings): 3D структури, такі як молекули, можуть бути представлені у вигляді графів, де вузли представляють атоми, а ребра - хімічні зв'язки між ними. Графічні вкладення перетворюють цей граф на числовий вектор, який враховує семантику атомів та зв'язків[87].

$$\begin{aligned} G_{\text{molecule}} &= \{V, E\} \\ v_i &= \text{AtomEmbedding}(a_i), \quad \forall v_i \in V \end{aligned} \quad (2.11)$$

$$e_{ij} = \text{BondEmbedding}(b_{ij}), \quad \forall e_{ij} \in E$$

де G_{molecule} - граф структури молекули

V - множина вузлів (атоми), E - множина ребр (зв'язки)

v_i - векторне вкладення атома a_i

e_{ij} - векторне вкладення зв'язку b_{ij} .

2. Представлення геометрії молекул: для молекул можна використовувати координати атомів в 3D-просторі для їх представлення. Кожна молекула може бути описана набором координат атомів x , y та z .

$$M = \{(x_i, y_i, z_i)\}, \quad \forall (x_i, y_i, z_i) \quad (2.12)$$

3. Використання віконних підходів: 3D структури можуть бути розбиті на невеликі вікна або фрагменти, які потім представляються у вигляді числових векторів. Це дозволяє обробляти великі структури покроково та об'єднувати інформацію з окремих фрагментів.

$$F_i = \text{FeatureExtraction}(W_i) \quad (2.13)$$

$$V = \text{Aggregation}(F_1, F_2, \dots, F_n)$$

де F_i - векторне представлення фрагмента i

W_i - фрагмент i у вікні

$V_{\text{structure}}$ - векторне представлення всієї структури.

4. Використання ґрид-представлення: ґрид-представлення - це метод, де простір навколо структури розділяється на рівномірну сітку, а значення властивостей об'єктів визначаються в кожному вузлі цієї сітки. Це дозволяє представити 3D структури у вигляді тривимірної матриці.

$$G_{grid}(x,y,z) = Property(x,y,z) \quad (2.13)$$

де $G_{grid}(x,y,z)$ - значення властивості в точці (x,y,z) .

Цей метод був обраний в якості основного в цій роботі. Із його недоліків можна відзначити те, що він не є компактним. Тензор, який зберігатиме інформацію про властивості об'ємної структури, буде 4-го порядку, а саме $[Width, Length, Height, N]$, де $Width$ – ширина структури, $Length$ – довжина структури, $Height$ – висота структури та N – розмірність вектору представлення властивостей. Якщо кількість властивостей буде 3 у точці (x,y,z) , то й N буде дорівнювати 3.

Так як різні метаматеріали можуть мати різні розміри, то й ґрид-представлення цих метаматеріалів буде мати різний розмір. Деякі види штучних нейронних мереж дозволяють мати довільний розмір вхідних даних, або часткову довільність. Наприклад, рекурентні нейронні мережі дозволяють мати довільну часову розмірність, але матимуть фіксовану feature розмірність. Довільну feature розмірність мають FCNN(Fully convolutional neural networks), які зазвичай використовуються у задачах image2image, semantic segmentation мають можливість подавати на вхід картинки довільного розміру, так як згорткові фільтри оперують не повним зображенням, а лише сегментним. Такий фільтр може бути використаний для зображення будь-якого розміру. В випадку класифікації, передбачення будуть використовуватися шари повнозв'язної нейронної мережі, котрі потребують фіксованого розміру. Тому, є потреба в додатковій обробці ґрид-представлення для зведеного розміру.

2.4 Підхід до кодування інформації про структуру, склад і властивості

метаматеріалів та умови експерименту

Дані про структуру, склад і властивості метаматеріалів були отримані з опублікованих робіт [17][18][88-90]. Ці дослідження підпадають під такі критерії:

- були визначені всі характеристики топології метаструктур;
- були визначені всі матеріали з яких створені метаструктури;
- експериментальні характеристики були однакового типу;
- були вказані умови проведення експериментальних досліджень, які будуть використані для навчання штучної нейронної мережі;
- складність математичного моделювання структур.

Така мала кількість робіт обумовлена тим, що, такі складні структури потребують значного часу для рендеру та обробки алгоритмами для представлення інформації для глибокого навчання

Зібрані дані про метаматеріали графічно представленні на рис. 2.7-2.11. На них відображені всі топологічні розміри потрібні для рендеру 3D структур в програмному забезпеченні для створення тривимірної графіки.

Всі структури були включені у вибірку даних для навчання штучної нейронної мережі. Було сформовано декілька вибірок даних для навчання, опираючись на їх частотний діапазон.

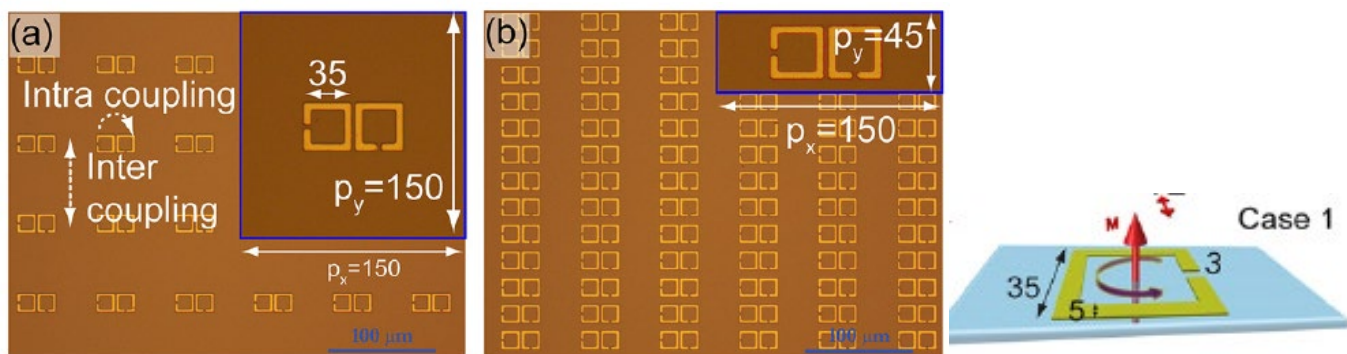


Рисунок 2.3 – Схематичний вигляд метаматеріалу - 1. Товщина антен 1 мкм. Товщина підкладки 5 мкм. $p_y = 150$ мкм(а), $p_y = 45$ мкм(б)[17].

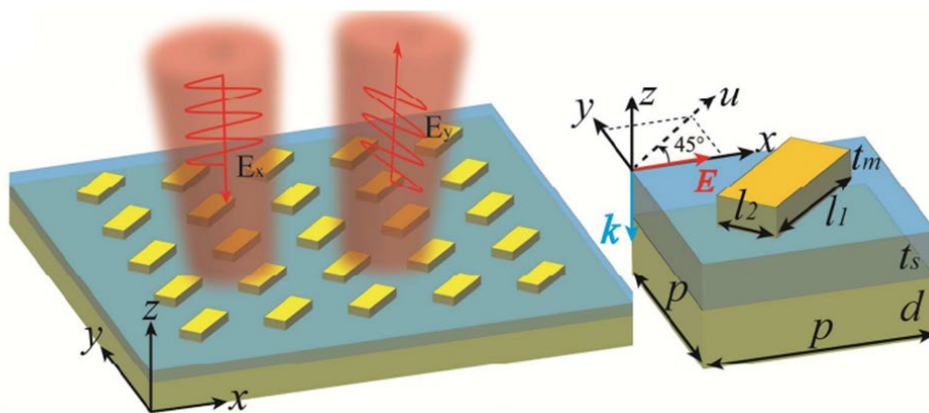


Рисунок 2.4 – Схематичний вигляд метаматеріалу - 2: $p = 600$ нм, $l_1 = 390$ нм, $l_2 = 165$ нм, $t_m = 80$ нм, $t_s = 110$ нм, $d = 150$ нм [18].

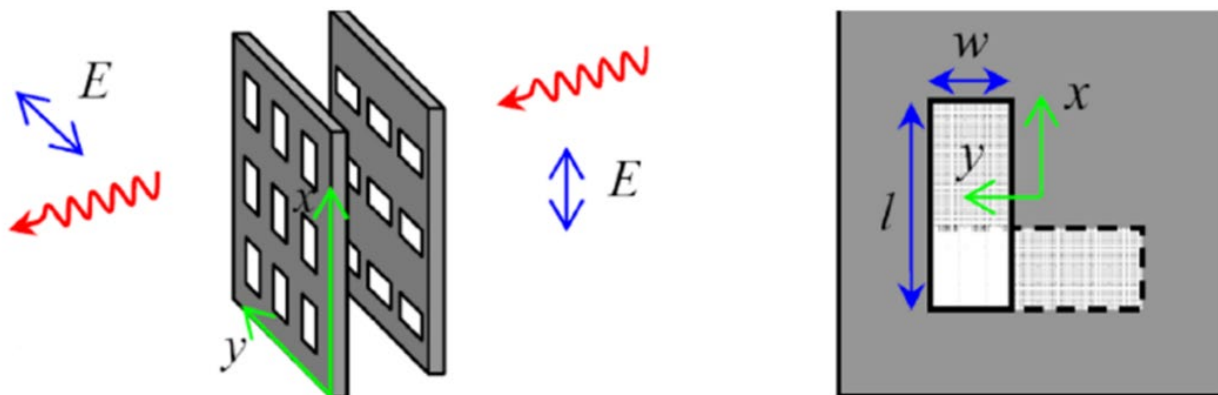


Рисунок 2.5 - Схематичний вигляд метаматеріалу - 3. $l = 750$ нм, $w = 280$ нм, $t = 150$ нм (товщина металічного шару), $d = 1000$ нм (відстань між вирізами) h - товщина діелектричного шару [88].

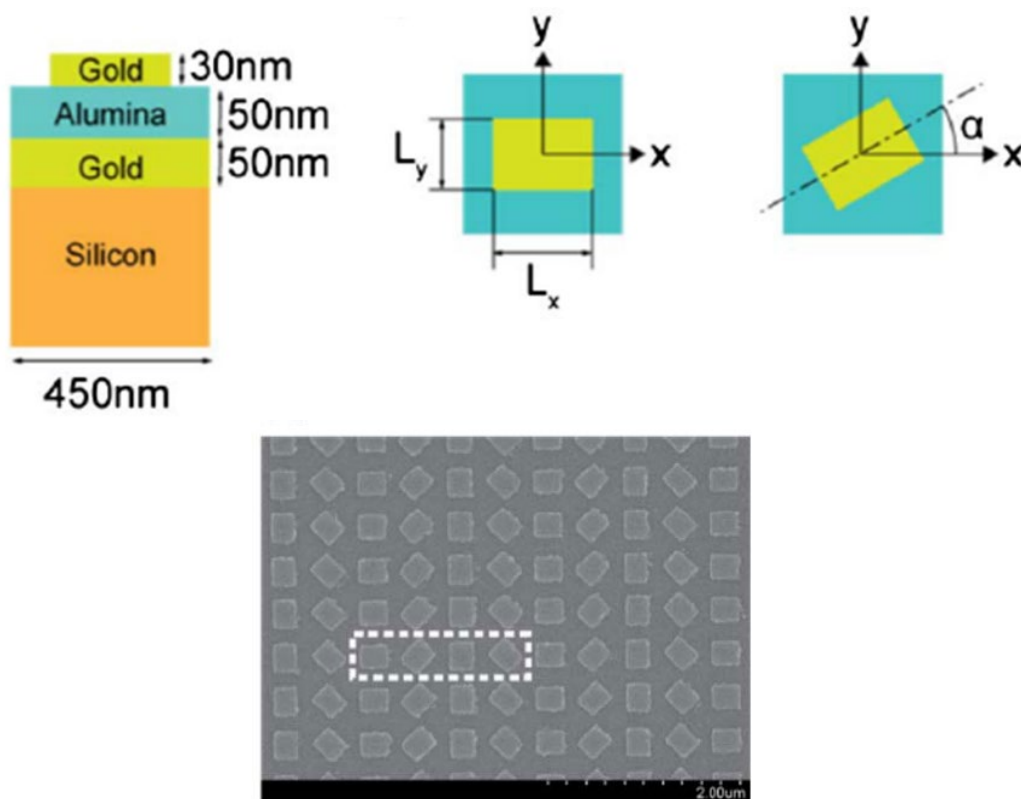


Рисунок 2.6 - Схематичний вигляд метаматеріалу - 4. $\alpha = 0^\circ; 45^\circ; 90^\circ; 135^\circ$. $L_x = 280$ нм, $L_y = 230$ нм [89].

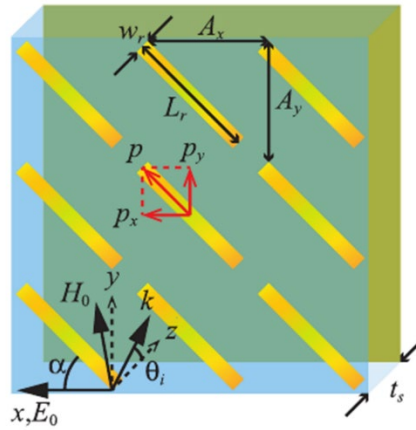


Рисунок 2.7 - Схематичний вигляд метаматеріалу - 5. $t_s = 33$ мкм, $A_x = A_y = 68$ мкм, $L_r = 82$ мкм, $w_r = 10$ мкм; товщина антен і шару золота 200 нм; товщина проміжку між смугами 10 мкм [90].

Так як більшість електромагнітних параметрів має залежність від частоти опромінення, для спрощення було обрано середні значення за частотою для всіх зразків. Це, безумовно, може викликати неточності в описі елементів метаструктур, та призвести до втрати інформації про фізичні властивості матеріалів. Але, з іншого боку, необхідно встановити фізичну характеристику, яка описує всі складові для створення можливості навчання моделі. Далі в дослідженні, будуть продемонстровані результати навчання штучної нейронної мережі, з яких буде видно, що модель все ще матиме можливість прогнозування та знаходження структурних залежностей метаматеріалів. Експерименти проводились на даних з різними діапазонами опромінення, тому було сформовано 3 вибірки даних для навчання та тестування штучної нейронної мережі:

- 1) діапазон від 0.3 до 0.6 ТГц. Такі метаструктури, зображені на рисунках 2.3 та 2.7;
- 2) діапазон від 131 до 374 ТГц. Такі метаструктури зображені на рисунках 2.4, 2.5, 2.6;

3) об'єднаний діапазон від 0.3 ТГц до 374 ТГц.

Для структури, що можна спостерігати на рисунку 2.3, матеріал підкладки представлений кремнієм, тоді як матеріал антен виготовлений зі золота. Щодо кремнію, діелектрична проникність дорівнює 12, а магнітна проникність позначена як $\mu = 1,25 \cdot 10^{-6}$ (загалом це відповідає діаманетичному характеру кремнію). Щодо золота: діелектрична проникність $\epsilon = -74$, магнітна проникність $\mu = -36 \cdot 10^{-6}$ Гн/м. Відносна провідність золота виражена $\rho = 0,4 \cdot 10^8$ См/м. Додаткові подробиці про структуру можна знайти за посиланням [17]. Ці параметри актуальні для діапазону від 0.3 ТГц до 0.6 ТГц.

У випадку структури, що відображена на рисунку 2.4, золото позначено жовтим кольором, діоксид кремнію - голубим, а підкладку складає кремній. Для діоксиду кремнію властива діелектрична проникність 2.1, а магнітна проникність - 1 (це загальні характеристики для діаманетичного діоксиду кремнію). Діелектрична проникність золота приймається як середнє значення в залежності від частоти (зважаючи на те, що для досліджуваних частот - ближній інфрачервоний діапазон - діелектрична проникність має змінюватись). Таким чином, $\epsilon = -74$ для золота, а магнітна проникність $\mu = -36 \cdot 10^{-6}$. Відносна провідність золота залишається $\rho = 0,4 \cdot 10^8$ См/м. Докладніші відомості про цю структуру наведені за посиланням [18]. Ці параметри актуальні для діапазону від 131 ТГц до 374 ТГц

Структура, яку бачимо на рисунку 2.5, відрізняється тим, що жовтим позначено золото, голубим - діелектричний заповнювач (Al_2O_3), а оранжевим - кремній. Характеристики золота щодо діелектричної, магнітної проникності та відносної провідності співпадають з характеристиками структури на рисунку 3.3. Щодо діелектричного заповнювача властиві значення $\epsilon = 10$, $\mu = 1,25 \cdot 10^{-6}$ Гн/м [89]. Ці параметри актуальні для діапазону від 131 ТГц до 374 ТГц

У випадку структури, що зображена на рисунку 2.16, золото позначено жовтим, діелектричний поліамід позначено синім, і його діелектрична проникність $\varepsilon = 4$, а магнітна $\mu = 1,25 \cdot 10^{-6}$ Гн/м. Щодо золота значення діелектричної проникності $\varepsilon = 7$; магнітна проникність $\mu = -36 \cdot 10^{-6}$ Гн/м. Відносна провідність золота залишається $\rho = 0,4 \cdot 10^8$ См/м. Докладніше інформацію про дослідження цієї структури можна знайти за посиланням [90]. Ці параметри актуальні для діапазону від 131 ТГц до 374 ТГц

Важливо відзначити, що всі наведені значення були встановлені, враховуючи інформацію з вказаних джерел, і у випадках, коли інформація про компоненти була недоступна, вона була визначена відповідно до частотних діапазонів експериментів. Під час навчання нейронної мережі використовувалися тільки цілі значення, виключаючи 10^{-6} .

На рис 2.12-2.16 представлені скріншоти структури метаматеріалів з програмного забезпечення для створення 3D об'єктів. Їх розміри повністю відповідають реальним розмірам з рис. 2.7-2.11

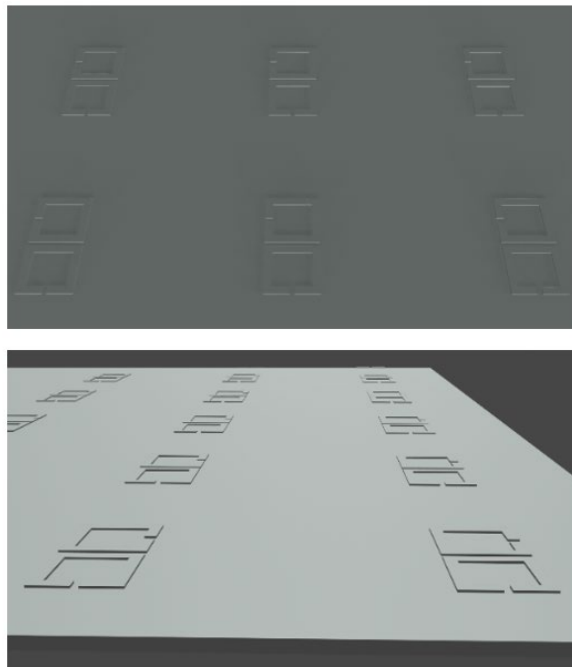


Рисунок 2.8 – Знімки 3D структур, розміри представлені на рис. 2.3 (а).

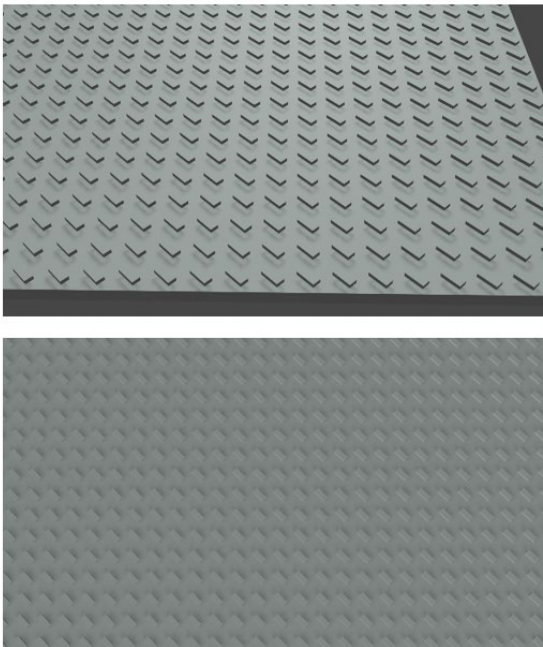


Рисунок 2.9 – Знімки 3D структур, розміри представлені на рис. 2.4.

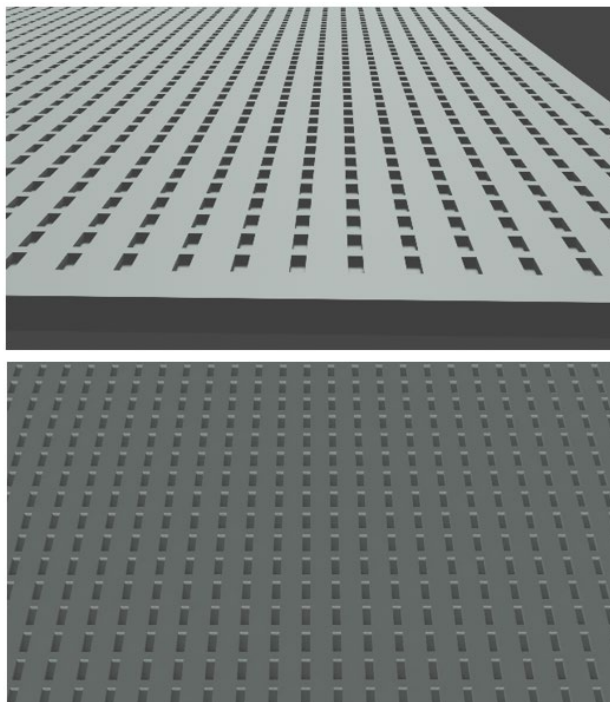


Рисунок 2.10 – Знімки 3D структур, розміри представлені на рис. 2.5.

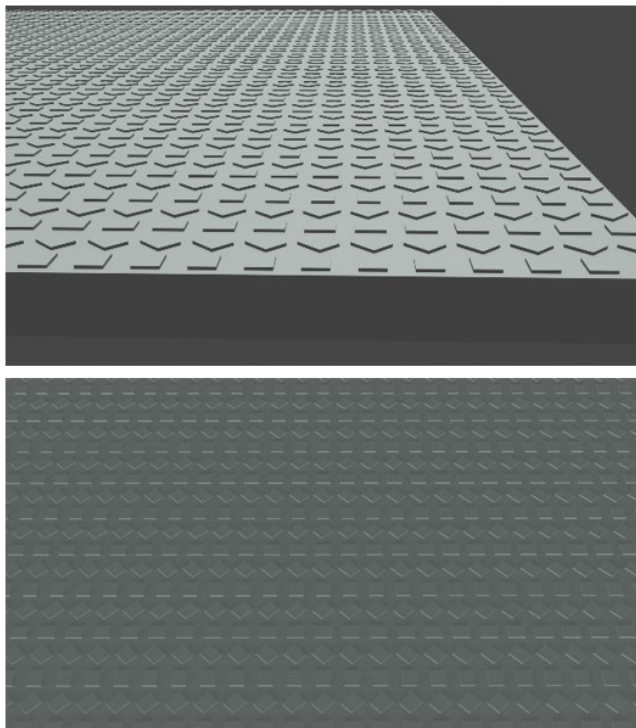


Рисунок 2.11 – Знімки 3D структур, розміри представлені на рис. 2.6.



Рисунок 2.12 – Знімки 3D структур, розміри представлені на рис. 2.7.

На рис 2.12 можна спостерігати лише верхні ділянки 3D структури, оскільки значна частина цієї структури не може бути адекватно показана через значну різницю у розмірах між антенами та підкладкою. Всі зображення були зроблені з погляду зверху та під деяким невеликим кутом нахилу.

Задача, яку треба вирішити, це обрати оптимальний спосіб представлення даних. Так як обрана архітектура базується на згортковій нейронній мережі, мною була запропонована ідея кодування інформації о матеріалі в додаткові “кольорові” канали.

Ідея підкріплена тим, що обробка багатоканальних зображень вже дуже добре досліджена іншим науковцями. Згорткові нейронні мережі мають здатність збирати та обробляти інформацію у єдиний вбудований простір представлення.

Інформація о матеріалі кодується у саме 3D зображення матеріалу таким чином, що в кожній трійці координат (x,y,z) співвідноситься трійка параметрів матеріалу, а саме діелектрична проникність, магнітна проникність та питома провідність.

Для реалізації такого кодування у сирий матеріал було використано програмне забезпечення *Blender*. Так як читання даних відбувалося із середовища мови програмування *Python*, потрібно було визначитися з форматом зберігання даних.

Існує декілька різних форматів, серед яких можна виділити наступні: *.dae*, *.abc*, *.obj*, *.ply*, *.stl*, *.fbx*, *.x3d* та інші. Кожен з них має свій унікальний спосіб представлення 3D-моделей, сцен та цифрової інформації, яку можна створювати у відповідному програмному середовищі.

Вибір найбільш слушного формату залежить від кількох чинників. Один з них - можливість ефективної роботи з даними в даному форматі. Крім того, важливо мати

можливість виконати конвертацію до формату, який оптимально підходить для використання у нейронних мережах.

З огляду на зручність обробки даних, а також простоти конвертування, був обраний формат, де в кожний рядок файлу записувався шестивимірний вектор формату:

$$X = [x, y, z, R, G, B], \quad (2.14)$$

де x, y, z - координати відповідних осей

R, G, B - значення інтенсивності кольорів червоного, зеленого та синього відповідно, в які закодований матеріал з якого було створено метаматеріал.

Існує виклик, пов'язаний з отриманням такого тензора напряму за допомогою вбудованих форматів програмного середовища. Тому був обраний підхід на основі програмного коду. За використання бібліотеки Open3D у мові програмування Python, всі моделі були збережені у форматі ".ply" та далі перетворені у дані у форматі ".xyzrgb".

Формат ".xyzrgb" включає шестивимірний тензор, що представляється у вигляді масиву даних, де кожен стовпець містить інформацію про просторове розташування та властивість матеріалу, а кожен рядок відповідає елементу метаструктури.

Оскільки представлення матеріалів за допомогою RGB значень не містить жодної прямої інформації про реальні фізичні властивості, необхідно виконати конвертацію значень RGB у характеристики матеріалів. Враховуючи, що в даній роботі передбачалось прогнозування електромагнітних характеристик метаструктур, зокрема залежності коефіцієнта пропускання від частоти випромінювання, було вибрано властивості матеріалів, що складають метаструктури, для подальшої конвертації.

Іншими словами, проекція RGB була перетворена на $\epsilon\chi\sigma$, і кожен окремий елемент тепер представляється так:

$$X = [x, y, z, \epsilon, \chi, \sigma], \quad (2.14)$$

де ϵ - діелектрична проникність

χ - магнітна проникність

σ – питома провідність

x, y, z – координати відповідних осей.

До вигляду даних, котрі будуть використовуватися для навчання накладається низька вимог, а саме:

- при будь-якій конвертації даних повинна зберігатися залежність фізичних властивостей матеріалів;
- коректно проведене масштабування;
- розмірність кожного вхідного масиву повинна бути однаковою;
- координатна розмірність не повинна перевищувати 1024x1024x20, із-за проблем пов'язаних з великими шарами згортки (уповільнення швидкості навчання, збільшення ємності нейронної мережі, що призводить до ефекту “перенавчання”).

Звісно, можливість включення параметрів у властивості матеріалів не обмежується лише трьома. Очевидно, що зі зростанням кількості параметрів збільшується обчислювальне навантаження. З іншого боку, важливо розуміти, що в параметри матеріалів варто включати лише ті параметри, які відображають фізичну природу експерименту. Розуміючи, що матеріали, такі як золото чи полімери, мають додаткові фізичні характеристики, які включають жорсткість, хімічний потенціал,

теплоту плавлення тощо, їх варто враховувати в параметрах матеріалів лише тоді, коли прогножуються властивості, які взаємодіють з цими параметрами.

Наприклад, якщо необхідно передбачити частоти акустичного руйнування певного матеріалу, то до вхідних параметрів слід включити коефіцієнт жорсткості або модуль пружності. Це пояснюється тим, що всі параметри, що служать еталонами для кожного матеріалу, були вивчені теоретично, з використанням відповідних моделей.

Розуміло, що введення залежностей параметрів, які не мають прямого відношення до властивостей, наприклад хімічного потенціалу, від залежності сили струму від напруги, може вплинути на прогноз відповідних властивостей при достатньому обсязі вибірки. Однак результат може бути менш точним, оскільки існують інші параметри, які мають аналогічно важливий внесок. Висновок полягає в тому, що при збільшенні кількості включених параметрів, що впливають на вихідні властивості матеріалу, можна покращити якість передбачень, зберігаючи при цьому розмір вибірки незмінним.

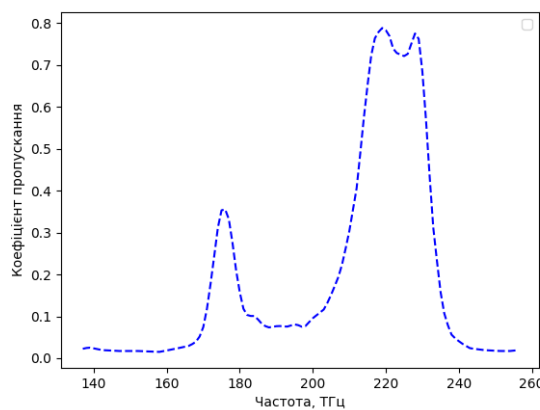
Але при цьому не слід забувати, що збільшення кількості вхідної інформації потребує і покращення ємності нейронної мережі шляхом збільшення кількості згорткових фільтрів, збільшення кількості згорткових шарів, збільшення кількості повнозв'язних шарів, що в більшості випадків призводить до перенавчання, якщо не збільшується навчальна вибірка. Тому, може скластися така ситуація, що при збільшенні кількості вхідної інформації та збільшення самої нейронної мережі, ми отримуватимемо гіршу якість передбачення.

Подання експериментальних характеристик не є настільки складним процесом, як подання структури та складу матеріалів у цифровому форматі. Варіанти подання експериментальних даних різноманітні. До класичних методів наочного представлення відносяться таблиці, ряди, розподіли, графіки, гістограми, схеми, малюнки. Дослідники у галузі глибокого навчання, як правило, розуміють, як відображати інформацію щодо

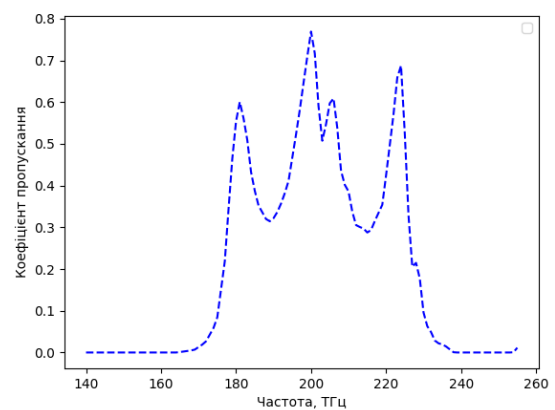
експериментальних досліджень, коли вони безпосередньо беруть участь у дослідницьких групах. У випадку, якщо необхідно збирати інформацію з джерел, що містять результати реальних фізичних вимірювань, потрібно вирішувати, як привести інформацію до потрібного формату в залежності від конкретної задачі.

Ще одним фактором є розмір вибірки експериментальної інформації. На дані можуть накладатися обмеження, зумовлені доступною пам'яттю для обробки даних. Це пов'язано з тим, що у роботі нейронної мережі значний час витрачається на обчислення помилки. Під час цього процесу вихідні дані порівнюються з даними, які були обчислені мережею. Залежно від представлення даних можуть знадобитися додаткові процедури, такі як згорткові мережі, які дозволяють одночасно зменшити та зберегти інформацію.

Наприклад, у даній роботі були використані експериментальні дослідження, подані у вигляді графіків залежності коефіцієнта пропускання від частоти випромінювання. Щоб зберегти інформативність щодо всіх змін в графіках, було зроблено припущення, що 20 точок буде достатньо, які представляють графіки. Графіки експериментальних досліджень, які були подані у вигляді масиву з 20 точок можна побачити на рисунку 2.13.



а)



б)

Рисунок 2.13 – Графіки експериментальних досліджень, які були використані для тестування нейронної мережі. (а) Для структури, яка представлена на рис. 2.9 при $h = 100$ нм. (б) Для структури, яка представлена на рис. 2.9 при $h = 180$ нм

Звичайно, для різних даних ці параметри є можливість корегувати. Значення у двадцять були встановлені допущенням про оптимальну кількість точок.

Важливим кроком для використання даних для навчання та тестування штучної нейронної мережі є масштабована обробка цих даних. Існує декілька видів масштабування, два найчастіше використовуваних це “min-max масштабування” та “стандартизація”.

Перший вид обробки масштабує значення вхідних даних до проміжку від нуля до одиниці. Формула для проведення min-max масштабування:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (2.15)$$

де x – значення величини що потрібно масштабувати

x_{scaled} – нове значення величини

x_{min} – мінімальне значення величини x у вибірці даних

x_{max} – максимальне значення величини x у вибірці.

Це масштабування потрібне для спрощення обчислень всередині нейронної мережі. Таке масштабування збільшує якість передбачення нейронної мережі[91]

Стандартизація також масштабує параметри розподілу, але вже не значення величини, а їх статистичні параметри розподілу. У випадку стандартизації після

масштабування середнє значення вибірки стає рівним нулю, а значення стандартного відхилення дорівнює одиниці. Формула для стандартизації величини x_i :

$$z_i = \frac{x_i - \bar{X}}{\sigma}, \quad (2.16)$$

де x_i – i -те значення величини x

$\bar{X}$ – середнє значення величини x у вибірці

σ – значення стандартного відхилення величини x у вибірці

z_i – стандартизоване значення величини x .

Цей тип масштабування не підходить для типу даних в цьому дослідженні, це пов'язано з малою кількістю даних.

Також потребують кодування дані для задання умов експериментальних досліджень.

В якості умов експерименту буде задано 3 величини, а саме кут поляризації, кут опромінення та показник заломлення діелектрику. Всі 3 умови можуть приймати різні значення, тому, для покращення якості передбачення та спрощенню обчислювань всередині нейронної мережі, ці значення будуть масштабовані та кодовані, а саме:

- тип поляризації, експериментальні дослідження проводились з крос поляризацією та взаємною поляризацією. Це бінарний тип даних, він може приймати лише 2 значення. Найоптимальніший спосіб закодувати такий тип це призначити одному випадку значення “0”, іншому “1”. В даному випадку для крос поляризації було призначено значення “1”, а випадку взаємної поляризації було призначено значення “0”;
- кут опромінення може приймати значення в межах $[20^\circ, 90^\circ]$. Такі дані потребують масштабування до меж $[0,1]$. Це масштабування буде робитися

за допомогою min-max масштабування. Мінімальне значення кута опромінення у виборці дорівнює 20° , а максимальне дорівнює 90° . Таким чином підставляючи ці значення у формулу масштабування (2.15):

$$\alpha_{scaled} = \frac{\alpha_{опромінення} - 20}{90 - 20} = \frac{\alpha_{опромінення} - 20}{70}, \quad (2.17)$$

де $\alpha_{опромінення}$ - значення кута опромінення

α_{scaled} - масштабоване значення кута опромінення;

- показник заломлення діелектрика підкладки може приймати значення в межах $[1.3, 2.0]$. Такі дані також потребують масштабування, як і випадку даних куту опромінення, до меж $[0, 1]$. Мінімальне значення показника заломлення діелектрику дорівнює 1.3, а максимальне значення показника заломлення дорівнює 2.0. Підставляючи ці значення у формулу min-max масштабування (2.15) отримуємо формулу для масштабування показнику заломлення:

$$n_{scaled} = \frac{n - 1.3}{2.0 - 1.3} = \frac{n - 1.3}{0.7}, \quad (2.18)$$

де n – показник заломлення діелектрика підкладки

n_{scaled} - масштабований показник заломлення діелектрика.

Для спрощення обробки даних та підготовки їх до навчання, умови експерименту будуть записані в імені відповідного файлу зі структурою та відповідними експериментальними даними у вигляді точок.

Сама інформація про умови експерименту буде представлена у вигляді тривимірного вектора:

$$cond = [pol, \alpha, n], \quad (2.19)$$

Тобто, якщо в умовах експерименту буде зазначено, що тип поляризації був взаємним, кут опромінення був 70° , а показник заломлення діелектрику 1.5, то не масштабований вектор умов експерименту буде:

$$cond = [0, 70, 1.5], \quad (2.20)$$

Як я зазначив вище, такі дані потребують масштабування для оптимального навчання та тестування нейронної мережі, якщо використати формули (2.17) та (2.18) до цього вектору умов експерименту, то отримаємо, що:

$$cond_{scaled} = [0, 0.714, 0.14], \quad (2.21)$$

Такий вектор буде вважатися готовим представленням умов експерименту для штучної нейронної мережі. Виходячи з формату такого представлення, єдиний варіант архітектури, яка зможе прийняти на вхід таку розмірність, буде повнозв'язна нейронна мережа.

Інший варіант представлення умов експерименту, це додати ці умові відповідно до кожного (x,y,z) як додаткові канали у 3D структурі разом з каналами котрі являються представленням характеристик матеріалів.

Але, у такого підходу виникає зразу декілька важливих недоліків:

- ми збільшуємо кількість інформації, яка подається на вхід нейронної мережі у $Width * Height * Length$ разів, аніж у випадку використання вектора умов експерименту. Це призведе до вимушеного збільшення ємності(кількості параметрів) штучної нейронної мережі. Додавання цих трьох каналів до вхідних даних збільшить об'єм вхідного структурного масиву у 2 рази, що призведе до уповільнення навчання, збільшить потребу у RAM та VRAM, та уповільнить швидкість прямого проходження нейронної мережі;

- додавання умов експерименту як додаткових каналів у структурному масиві вимагатиме значного дублювання інформації в даних. Звичайно умови експерименту є константними в кожній точці структури. Штучна нейронна мережа має внутрішні механізми ігнорування надлишкової інформації. Ці механізми корисні за умови, що ми не знаємо чи є в наших даних надлишок інформації. В цьому дослідженні, ми знаємо, що додаткові умови структурно-константі, а саме, тип поляризації, кут опромінення та показник заломлення діелектрику постійні на всій площині метаматеріалу, тому нема потреби у створенні структурованої залежності від умов експерименту, на відміну від структурно-матеріальної залежності.

У випадку, якщо б ми працювали з матеріалом на який діють різні умови експерименту, при чому в один й той самий час, такий спосіб представлення інформації о експериментальних умовах мав би підґрунтя.

Висновки до розділу 2

У даному розділі було розглянуто кілька ключових аспектів, пов'язаних з підготовкою даних для використання в глибокому навчанні для передбачення характеристик електромагнітних метаматеріалів.

Джерелом даних для навчання можуть статті у журналах або математичне моделювання. Такі дані повинні містити параметри матеріалів, частотні характеристики, інформацію про структуру та склад метаматеріалів.

Структурна інформація повинна бути представлена для штучної нейронної мережі. Чотиривимірний масив буде оптимальним контейнером для зберігання структурної інформації.

Для кодування інформації о умовах експериментальних досліджень вистачить векторного представлення.

Загальною метою цього розділу було розкрити різні аспекти збору, обробки та представлення даних про електромагнітні метаматеріали з огляду на використання їх у глибокому навчанні. Це є важливим етапом у підготовці даних для успішного навчання нейронної мережі та передбачення характеристик метаматеріалів на основі їх структури та складу.

РОЗДІЛ 3.

АРХІТЕКТУРА ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ПРОГНОЗУАННЯ ВЛАСТИВОСТЕЙ ЕЛЕКТРОМАГНІТНИХ МАТЕРІАЛІВ

3.1 Постановка задачі

Повний цикл навчання містить в собі:

- Конвеєр даних. Тобто, повністю спроектований цикл обробки даних від файлу з необробленими даними, до представлення даних для нейронної мережі у вигляді тензорів
- Коректно створені вибірки даних для тренування та тестування штучної нейронної мережі. При роботі зі штучними нейронними мережами, або класичними алгоритмами машинного навчання, є так зване правило “70/20/10”, або, при надмалій кількості даних “80/20”[92]. Це правило означає, що на тренування йде 80% від загальної вибірки даних, та 20% йде на тестувальну вибірку. Це робиться для того, щоб перевірити узагальнюючу можливість штучної нейронної мережі. Тобто, ми перевіряємо якість передбачення штучної нейронної мережі на даних, яка вона не “бачила”. В цій роботі, із-за над малої кількості даних, в деяких тестах це співвідношення не буде виконуватися, в таких випадках співвідношення буде 70/30 або 90/10
- Спроектвана архітектура штучної нейронної мережі. Сучасні архітектури штучних нейронних мереж зазвичай являють собою комбінацію під-архітектур[93-96]. Навіть стара та одна з найперших архітектур LeNet, що відкрила шлях розвитку архітектур згорткових нейронних мереж, яка була

використана для класифікації зображень, побудована на базі двох блоків: згортковий блок для визначення та обробки візуальних ознак, та перетворення його у вектор розмірності 128, та блок повнозв'язної штучної нейронної мережі[34]. Ці два блоки можна поділити на 2 логічні частини штучної нейронної мережі. Це поділення допоможе інтуїтивно зрозуміти як відбувається проектування штучної нейронної мережі. У випадку архітектури LeNet, перший згортковий блок являє собою візуальний препроцесор, який дозволяє обробити та стиснути обсяг інформації. Таку оброблену та скомпресовану інформацію, яка являє собою 128-мірний вектор, далі передають на другий блок, який вже являє собою механізм прийняття рішень, який по скомпресованій та обробленій інформації визначає що саме було зображено на картинці, яку подали на вхід нейронній мережі. Перед інженером машинного навчання, або дослідник завжди стоїть першочергова задача по первинній обробці даних. Ця обробка містить в собі аналіз самих даних. Результатом цього аналізу це відповіді на питання:

- яка природа цих даних?
- Чи є в даних часові, структурні, семантичні залежності?
- Чи можливий первинна обробка даних(стиснення та обробка класичними методами машинного навчання) для покращення роботи штучної нейронної мережі
- Яка розмірність цих даних і як вона вплине на швидкість навчання та тестування?
- Яка кількість даних буде в тренувальній вибірці і чи можна штучно збільшити різноманіття?

Відповіді на ці питання дають змогу підготувати первинну архітектуру штучної нейронної мережі. Наприклад, якщо типом даних є картинка, 3D структура, цифровий або аналоговий сигнал, який містить локальні залежності, то первинним блоком архітектури буде скоріш за все блок заснований шарах згорткової мережі з

ядрами згортки 1D для аналогових та цифрових сигналів, ядрами згортки 2D для зображень, ядрами згортки 3D для об'ємних структур. Для даних з часовою залежністю можуть бути використані також рекурентні нейронні мережі, зазвичай вони використовуються для обробки одновимірних типів даних, як сигнал, або часовий ряд. Але, з мого досвіду, може використовуватися комбінація згорткового блоку та рекурентного. Така комбінація добре працює коли вхідні дані являють собою відеоматеріал. Згортковий блок оброблює незалежно кадр за кадром, генеруючи вектор представлення розмірністю:

$$visual\ representation = [features, time], \quad (3.1)$$

де, *features* – розмірність вектору представлення кадрової інформації

time – кількість кадрів у поданому на вхід відеоматеріалі.

Такий вектор представлення подається на вхід вже рекурентному блоку штучної нейронної мережі, який має можливість знаходити та узагальнювати залежність у часі. Результатом обробки рекурентного блоку вже буде один n-мірний вектор, який буде зберігати в собі інформацію як часову, так й структурну. Такий вектор зручно подати на вхід блоку прийняття рішень, який базується на простій повнозв'язній штучній нейронній мережі. Таким чином, задача класифікації відеоматеріалу буде вирішуватися за допомогою штучної нейронної мережі, яка побудована на трьох класичних блоках штучної нейронної мережі, а саме CNN(згортковий блок), RNN(рекурентний блок) та FCNN(повнозв'язний блок). Ці блоки можуть бути замінені на більш складні, наприклад рекурентний блок можна замінити на блок з трансформером. Це ускладнить навчання та може зменшити швидкість навчання штучної нейронної мережі, але покращити точність класифікації. Блок згортки можна замінити на Visual Transformer, що також може покращити якість класифікації, але суттєво зменшити пропускну здатність нейронної мережі[97]. Використання більш складних та сучасних блоків штучної

нейронної мережі завжди призводить до збільшення ємності архітектури, що з одного боку може покращити якість роботи, але з іншого боку, при занадто великій ємності призводить до перенавчання та погіршення якості класифікації, якщо даних буде замало. Перед інженером машинного навчання завжди стоїть питання “швидкість або якість” і в залежності від задачі інженер проектує архітектуру, вибираючи з блоків, або навіть модифікуючи існуючі.

1. Обрані метрики для тестування штучної нейронної мережі та функція помилки. Для отримання об’єктивної інформації о якості передбачень штучної нейронної мережі повинні бути обрані метрики, котрі кількісно будуть сигналізувати про те, чи погано, або чи добре була навчена штучна нейронна мережа. Наприклад для задачі класифікації зазвичай використовують метрику Ассигасу(точність), якщо в даних є нерівномірність між “Істинними” та “Хибними” даними додають ще метрики Recall(повнота) та Precision(точність). Для розуміння що таке “Істинні” та “Хибні” дані, розглянемо задачу класифікацію не бракованих деталей. В такому випадку, ми обираємо що, “Хибними” даними будуть такі, які являють собою браковану деталь, а “Істинними” які являють собою не браковану деталь. Зазвичай відсоток браку не значний, нехай він буде 5%. В такому разі, якщо ми збираємо дані з виробництва, ми можемо очікувати значне зміщення “Істинних” даних. Що призводить до того, що ми можемо замість штучної нейронної мережі завжди маркувати будь-яку деталь як “не брак”, що дасть нам метрику Ассигасу більшу за 50%. Вибір метрик дуже важлива частина розробки повного конвеєру навчання штучної нейронної мережі. Також важливою частиною є вибір функції помилки якою будуть оновлюватися ваги штучної нейронної мережі. До функції помилки ставляться такі умови:

- Функція втрат повинна бути не від'ємною для всіх можливих вихідних значень моделі
- Функція втрат повинна бути неперервною, а також її похідна повинна бути визначеною в більшості точок
- Для ефективного навчання нейронної мережі функція втрат повинна бути диференційованою. Це дозволяє використовувати градієнтні методи для оновлення параметрів мережі на основі похідних функції втрат.
- Здатність виразити різницю між прогнозованими і справжніми значеннями: Функція втрат повинна виявляти різницю між прогнозованими значеннями моделі та справжніми значеннями в навчальних даних. Це дозволяє нейронній мережі коригувати свої параметри так, щоб наближатися до оптимальних рішень.
- Монотонність: Інтуїтивно, якщо втрата зменшується, коли модель наближається до правильних відповідей, це може сприяти більш ефективному навчанню. Таким чином, монотонна функція втрат може полегшити процес оптимізації.
- Чутливість до помилок: Функція втрат повинна відреагувати на невірні прогнози моделі та величину цих помилок. Це допомагає забезпечити відповідний вплив невірних прогнозів на процес навчання.
- Спроможність враховувати особливості задачі: Функція втрат повинна бути вибрана з урахуванням конкретної задачі та цілей моделі. Наприклад, для задач класифікації можуть використовуватися різні функції втрат, такі як категоріальна перехресна ентропія або втрата Х'юбера для регресії.

Таким чином були розглянуті основні пункти первинного проєктування конвеєру штучної нейронної мережі. В даному описані не були описані такі під етапи як: вибір регуляризації даних, вибір гіперпараметрів, або навіть grid-search(перебору гіперпараметрів) для пошуку найкращої комбінації налаштувань, перебір параметрів архітектури (шарів, ширини, кількості блоків), вибору типів зв'язків між блоками (одинарний послідовний, або ж зв'язок зі skip-connection), наявність augmentation(штучного збільшення кількості даних внаслідок нелінійної зміни вхідних даних, такі як оберти, віддзеркалення, створення локальних занулень даних), наявність або відсутність планувальників оптимізації (Cosine, Decay, Annealing, Plateau), вибір оптимізатора(звичайний градієнтний стохастичний спуск, Adam, RAdam, NAdam, Nesterov, RMSProp, etc), наявність алгоритмів розумного сортування даних перед подачею їх на тренування.

Це не всі пункти, що вирішуються при проєктуванні та навчанні штучної нейронної мережі, але проблема їх вирішення заздалегідь в тому, що такі пункти вирішуються вже при навчанні, та тільки методом експериментів. Кожне навчання потребує тонке налаштування всіх цих параметрів та пунктів

3.2 Алгоритм програмного забезпечення з підготовки, кодування та представлення даних про метаматеріали та умови їх властивостей

Як було вже зазначено, перший етап конвеєру це проєктування циклу обробки даних починаючи від відкриття файлу з диска, закінчуючи поданням готового представлення на вхід штучної нейронної мережі.

Через те, що файли рендеру з програмного забезпечення *Blender* не мають можливості бути зчитаними, був обраний проміжний тип файлів *.xyzrgb*, який може

бути зчитаний як текстовий файл мови програмування *Python*. Далі на рис. 3.1 буде показан приклад такого типу даних.

```

1 111.5000 57.5000 5.0000 0 -36 45500000
2 111.5000 57.5000 5.0000 0 -36 45500000
3 111.5000 92.5000 5.0000 0 -36 45500000
4 111.5000 92.5000 5.0000 0 -36 45500000
5 76.5000 57.5000 5.0000 0 -36 45500000
6 76.5000 57.5000 5.0000 0 -36 45500000
7 76.5000 92.5000 5.0000 0 -36 45500000
8 76.5000 92.5000 5.0000 0 -36 45500000
9 95.5000 57.5000 5.0000 0 -36 45500000
10 95.5000 57.5000 5.0000 0 -36 45500000
11 92.5000 57.5000 5.0000 0 -36 45500000
12 92.5000 57.5000 5.0000 0 -36 45500000
13 92.5000 62.5000 5.0000 0 -36 45500000
14 81.5000 62.5000 5.0000 0 -36 45500000
15 76.5000 57.5000 5.0000 0 -36 45500000
16 92.5000 57.5000 5.0000 0 -36 45500000
17 81.5000 87.5000 5.0000 0 -36 45500000
18 76.5000 92.5000 5.0000 0 -36 45500000
19 95.5000 62.5000 5.0000 0 -36 45500000
20 106.5000 62.5000 5.0000 0 -36 45500000
21 111.5000 57.5000 5.0000 0 -36 45500000
22 95.5000 57.5000 5.0000 0 -36 45500000
23 111.5000 92.5000 5.0000 0 -36 45500000
24 106.5000 87.5000 5.0000 0 -36 45500000
25 76.5000 92.5000 5.0000 0 -36 45500000
26 81.5000 87.5000 5.0000 0 -36 45500000
27 111.5000 57.5000 6.0000 0 -36 45500000
28 111.5000 92.5000 6.0000 0 -36 45500000
29 111.5000 92.5000 6.0000 0 -36 45500000
30 111.5000 57.5000 6.0000 0 -36 45500000
31 76.5000 57.5000 6.0000 0 -36 45500000
32 76.5000 92.5000 6.0000 0 -36 45500000
33 76.5000 92.5000 6.0000 0 -36 45500000
34 76.5000 57.5000 6.0000 0 -36 45500000

```

Рисунок 3.1 – Приклад структурних даних, що були переформатовані у формат *.xyzrgb*

Такий формат даних в не обробленому вигляді не підходить для представлення штучної нейронної мережі, тому постає задача у створенні алгоритму кодування сирих даних у готове представлення для штучної нейронної мережі.

Іншою проблемою є те, що таких рядків, як на рис. 3.1 мільйони, тому цей алгоритм повинен швидко та оптимально конвертувати та кодувати ці дані.

Для розв’язання цієї проблеми була використана бібліотека *Pandas*, для роботи з табличними даними. Такий файл, може бути представлений як таблиця формату CSV(Comma-Separated Values), де замість ком – пробіли. Бібліотека *Pandas* має оптимізовані алгоритми, що дозволяють ефективно виконувати пошук, сортування, видалення та модифікацію значень у внутрішньому форматі бібліотеки *DataFrame*, в який конвертуватиметься файл формату *.xyzrgb*.

Інша задача, це пошук оптимального контейнера для зберігання форматованої та закодованої інформації о структурі метаматеріалу. Мною було обране представлення структурної та матеріальної інформації у вигляді 4D тензору:

$$struct = (x, y, z, n) \quad (3.2)$$

де x, y, z – структурні координати

n – канали з інформацією о матеріалі, по аналогії з кольоровими каналами у RGB картинках та структурах.

Так як мною були обрані 3 характеристики матеріала, що будуть цифровим представленням, а саме: діелектрична проникність, магнітна проникність та питомий опір, то цих каналів буде три. Тобто, підставляючи у формулу (3.2), в нашому випадку розмірність тензора структурної та матеріальної інформації буде:

$$struct = (x, y, z, 3) \quad (3.3)$$

Таким контейнером для зберігання був обраний *NumPy* масив. Це масив – основний формат зберігання даних, який підтримується двома основними бібліотеками

глибокого навчання *PyTorch* від *Meta* та *Tensorflow* від *Google*. Причиною вибору такого формату даних три пункти:

- підтримка формату даних основними бібліотеками навчання, що дає змогу вибору більш слушної бібліотеки для цієї задачі;
- оптимізоване зберігання масиву у пам'яті, що зменшує об'єм RAM потрібної для роботи конвеєру;
- інструменти для обробки даних в такому масиві. Алгоритми та математичні операції з масивами *NumPy* реалізовані на низькорівневих мовах програмування таких як *Fortran* та *C++*.

Надалі всі контейнери для зберігання представлень структурних, експериментальних даних та умов експериментів будуть *NumPy* масивами.

Алгоритм обробки формату даних *DataFrame* буде полягати в ітеративному переборі від початку та кінця визначених меж структури. Проблеми такого алгоритму полягає в тому, що його асимптотична складність big-O анотації буде виражена як $O(x * y * z)$, де x, y, z , - кількість пікселів по кожній з координат.

Нажаль, такий алгоритм складно оптимізувати, тому що сама задача це перебрати всі можливі точки у просторі структури. Єдиним шляхом пришвидшення повного циклу перетворення *.xyzrgb* у *NumPy* масив, це пришвидшити пошук елементів у *DataFrame*. Так як операції пошуку у такому форматі були оптимізовані, ми можемо скористатися об'єднанням умов пошуку по таблиці:

```
material = df[(df.x == x) & (df.y == y) & (df.z == z)],
```

де *material* – значення трьох обраних параметрів за координатами x, y, z

& - логічна операція “AND”

df – *DataFrame* бібліотеки *Pandas*

Далі буде розглянуто блок схеми на рис. 3.2 такого алгоритму:

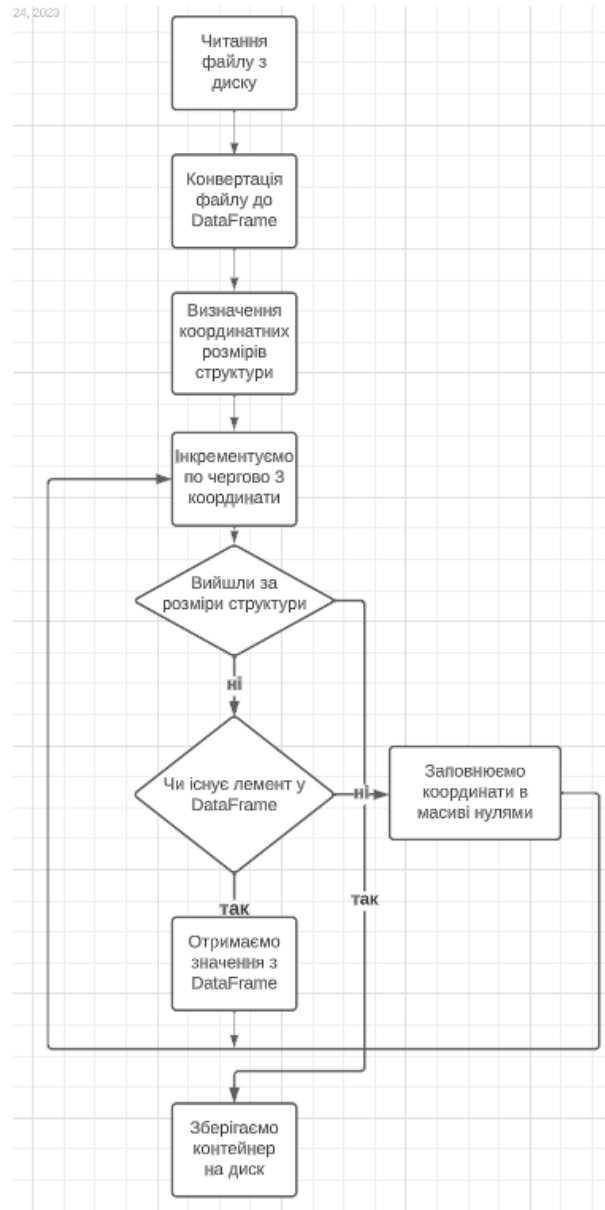


Рисунок 3.2 – Блок схема алгоритму обробки формату *.xyzrgb* у NumPy масив

З блок-схеми на рис. 3.2 видно, що алгоритм ітерується по всім координатам визначеним в файлі формату *.xyzrgb* який був конвертований у формат зберігання

даних *DataFrame*, робить перевірку, чи існує така трійка координат у *DataFrame*, і якщо існує, то отримує трійку значень характеристик, що кількісно відображає матеріал, в іншому випадку, якщо така трійка координат не існує в *DataFrame*, відносить до такої трійки координат пустоту, що кодується характеристиками (0,0,0). Тобто, простір, де матеріал був не вказаний вважається пустотою. Коли алгоритм, перетинає просторову межу структури, тобто виходить за визначені координати у *DataFrame*, він вважає, що обробка закінчена і зберігає файл у проміжний файл на диск.

Механізм проміжного збереження потрібен, тому що сам процес обробки одного *.xyzrgb* файлу займає від 3 до 5 годин, в залежності від заданих розмірів. Для того, щоб не втратити результат роботи, якщо програма завершиться із-за помилки або нестачі пам'яті, вона й зберігає проміжні конвертації,

Наступним етапом обробки структурних даних є масштабування. В цьому дослідженні використовуються такі характеристики матеріалу як діелектрична провідність, магнітна провідність та питомий опір. Значення цих величин відрізняється на декілька порядків, самі значення суттєво виходять за оптимальні межі [0, 1]. Тому наступною задачею стоїть у виборі способу масштабування.

Як було зазначено в цій роботі таких методів існує декілька. Мною було обрано алгоритм масштабування min-max. Формула масштабування min-max для структур буде виглядати як:

$$structure_{scaled} = \frac{structure - min_{structure}}{max_{structure} - min_{structure}}, \quad (3.4)$$

де $structure_{scaled}$ — масштабована структура

$min_{structure}$ — мінімальне значення в каналах представлення структури

$max_{structure}$ — максимальне значення в каналах представлення структури.

Алгоритм масштабування буде представлений на рис. 3.3 у вигляді блок-схеми:



Рисунок 3.3 – Блок-схема масштабування структурної інформації о метаматеріалі

Після масштабування значення характеристик матеріалів буде знаходитися в проміжку $[0, 1]$.

Наступним типом даних, що потребує попередньої обробки для представлення у потрібному, для навчання штучної нейронної мережі, форматі є експериментальні дані, передбаченням яких буде навчатися нейронна мережа. Тобто, використовуючи визначення машинного навчання, це будуть “*label*” або розмітка.

На диску такі дані зберігаються як текстовий документ, в якому на кожній строчці визначена пара x та y , що показано на рис. 3.4.

1	0.2;0.9854
2	0.21;0.9902
3	0.22;0.9805
4	0.23;0.9805
5	0.24;0.9805
6	0.25;0.9805
7	0.26;0.9707
8	0.27;0.9659
9	0.28;0.9659
10	0.29;0.9659
11	0.3;0.9415
12	0.31;0.9415
13	0.32;0.9463
14	0.33;0.922
15	0.34;0.8585
16	0.35;0.7951
17	0.36;0.7463
18	0.37;0.7659
19	0.38;0.8439
20	0.39;0.8829
21	0.4;0.8293
22	0.41;0.6049
23	0.42;0.4878
24	0.43;0.5805
25	0.44;0.8341
26	0.45;0.9512
27	0.46;0.9512
28	0.47;0.9512
29	0.48;0.9659
30	0.49;0.9805
31	0.5;0.9659
32	0.51;0.961
33	0.52;0.9659
34	0.53;0.9659
35	0.54;0.961
36	0.55;0.961
37	0.56;0.961
38	0.57;0.9659
39	0.58;0.961
40	0.59;0.961
41	0.6;0.9756

Рисунок 3.4 – Приклад текстового документа з експериментальними даними

Такі дані є більш простими і не потребують додаткової конвертації у проміжні формати як *DataFrame*. Реалізація цього алгоритму не буде містити а ні проміжних конвертацій, а ні проміжного збереження на диск.

Нижче буде представлено алгоритм у вигляді блок-схеми на рис. 3.5.



Рисунок 3.5 – Блок-схема обробки експериментальних даних

Як можна побачити з блок-схеми на рис. 3.5, до експериментальних даних використовується min-max масштабування. Форматом зберігання оброблених даних був обраний масив *NumPy*.

Експериментальні дані містять в собі 40 пар точок x та y . Нашою задачею є прогнозування як x так і y . В такому випадку, вихідний шар нейронної мережі містив би в собі 80 нейронів. З огляду на кількість даних в вибірці та експериментальними дослідженнями, 80 нейронів не є оптимальною кількістю. Тому, було обрано, що штучна нейронна мережа матиме 40 нейронів в останньому шарі. Тобто, 20 точок x та 20 точок y . Для цього, на етапі обробки, алгоритм відсіює кожну другу точку.

Для збереження у файли тренувальної вибірки, був обраний бінарний формат пакета *NumPy* .prz. Цей контейнер дозволяє структуровано зберігати дані у вигляді словника. Це дає змогу розгалузити структурні та експериментальні дані для навчання. Вони зберігатимуться за ключами “*struct*” та “*points*”. Блок-схема цього алгоритму представлена на рис. 3.6.



Рисунок 3.6 – Блок-схема об’єднання структурної інформації та експериментальних даних

Вище представлена блок-схема об'єднання обробленої структурної інформації та експериментальних, та збереження їх у одному файлі-контейнері формату *.npz*

Це дасть змогу строго і однозначно співставити пару *структура-експериментальні дані*. Іншою перевагою використання вбудованого формату зберігання даних *.npz*, це можливість напряму зчитувати дані з диска під час навчання не використовуючи проміжні конвертації з примітивних типів даних та оптимізованих бібліотеки *NumPy* та бібліотеки глибокого навчання.

Наступний тип даних, що потребує попередньої обробки це – умови проведення експериментальних досліджень. Ця інформація, що показана на рис. 3.7 занесена у ті ж самі файли, що й експериментальні дані.

```
polarization = cross-pol
angle = 90 degree
ns = 2.0
number of plates = 2
```

Рисунок. 3.7 – Формат запису умов проведення дослідження

Такий тип даних, як і експериментальні дані, не потребує алгоритмічно надскладної обробки. В першу чергу, для обробки потрібно створити скрипт обробки даних з файлу, написати алгоритм масштабування даних базуючись на *min-max*. Та вибрати формат зберігання даних. Алгоритм для обробки параметрів електромагнітного випромінення представлений на рис. 3.8.

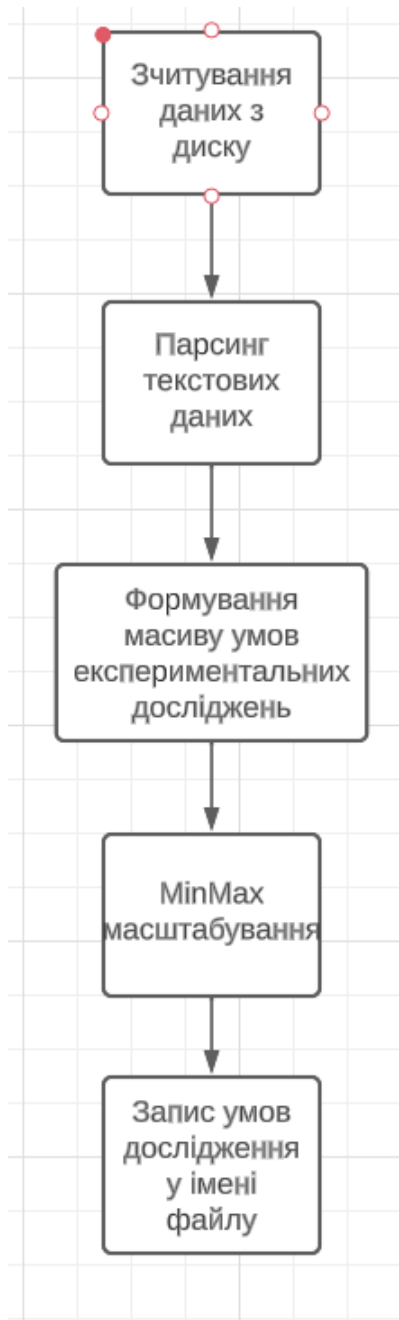


Рисунок 3.8 – Блок-схема обробки даних про параметри електромагнітного випромінювання.

Через те, що в якості умов експериментального дослідження були обрані три величини: тип поляризації, кут опромінення та показник заломлення діелектрику, то було обрано записувати в ім'я файлу ці умови. Це зручний формат, так як:

- такий формат дає змогу розгалузити експерименти з метаматеріалами. Так як унікальних структур у цьому дослідженні всього 5, то треба мати унікальне ім'я файлу під кожний експеримент;
- додавання умов експериментальних досліджень дає змогу візуально оцінити вибірку даних та власноруч поділити на тренувальні та тестувальні вибірки;
- прив'язка умов експерименту до представлення структурної інформації, що дає змогу рівнозначно співставити дані.

Таким чином було обрано саме запис умов експериментальних досліджень у ім'я файлу. Наприклад: `figure1A-1_co-pol_90_2.0_1.npz`, де `figure1A` – кодова назва метаматеріалу, `co-pol` – взаємна поляризація, `90` – кут опромінення, `2.0` – показник заломлення діелектрику, `1` – кількість пластин.

Останній параметр використовуватися не буде, але він був залишений для повної інформації о умовах проведення експериментальних досліджень.

Далі буде представлена повна схема первинного конвеєру для обробки даних починаючи від завантаження “сирих” файлів з диска, закінчуючи готовими для використання в навчанні файлів. Таким чином був спроектований алгоритм повного конвеєру обробки даних, який представлений на рис. 3.9.

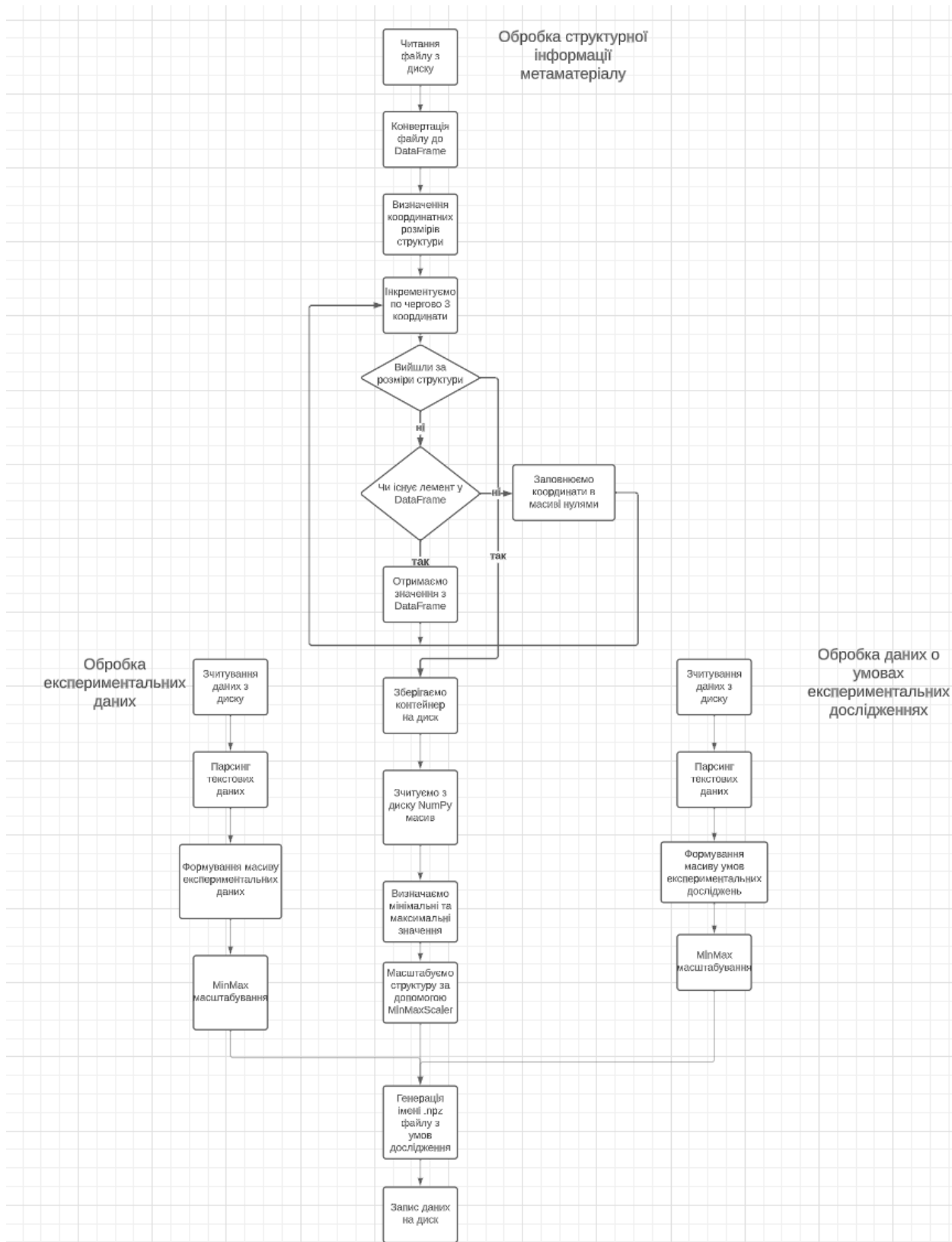


Рисунок 3.9 – Блок-схема конвеєру обробки даних для навчання ШНМ

На рис. 3.9 відображений повний конвеєр обробки даних, що є поєднанням у єдину схему алгоритмів з рис. 3.2, рис. 3.5, рис. 3.6, рис. 3.8

3.3 Метод закріплення інформації про умови експерименту до структури та складу метаматеріалів

Тут буде розглянуто декілька методів закріплення інформації про умови експерименту до структури та складу метаматеріалів. Один із ключових аспектів у вивченні електромагнітних метаматеріалів полягає в тому, як ефективно і точно інтегрувати умови експерименту та фізичні параметри матеріалів у нейронну мережу. Цей пункт присвячений пошуку оптимальних стратегій кодування інформації про умови експерименту для подальшого використання в навчанні нейронної мережі.

Важливість цієї задачі полягає в тому, що фізичні властивості, які характеризують умови експерименту, можуть значно впливати на електромагнітні властивості метаматеріалів. Врахування цієї інформації може покращити точність прогнозування та роботу нейронної мережі на реальних даних. В цьому пункті будуть розглянуті різні аспекти, пов'язані зі способами включення умов експерименту до даних

Підходи, представлені у цьому розділі, мають на меті забезпечити оптимальне представлення інформації про умови експерименту, що дозволить покращити якість прогнозування властивостей метаматеріалів та дозволить нейронній мережі враховувати різноманітні аспекти фізичного експерименту під час навчання.

Перш ніж буде розглянуто декілька способів використання додаткової інформації для покращення якості передбачення, будуть сформовані умови проведення цього дослідження:

- мала вибірка даних для навчання;
- комбінація умов велика, а саме значення комбінацій цих умов дорівнює 32;
- не має ліміту на пропускну можливість штучної нейронної мережі;
- попередня інформація в виборці не використовується як додаткова.

Існує декілька способів використання додаткової інформації для покращення якості передбачення, класифікації, тощо.

Перший такий спосіб, який буде розглянутий це створення “*model zoo*” або ж зоопарку моделей для кожного випадку[98][99].

Такий спосіб може бути корисним в умовах:

- кількість даних у навчальній вибірці достатньо велика для розділення цього датасету на під-вибірки на кожну комбінацію умов;
- комбінація умов не має велике різноманіття;
- є ліміт на пропускну можливість штучної нейронної мережі. Через те, що єдина архітектура штучної нейронної мережі для всіх умов буде значно більша ніж архітектура штучної нейронної мережі для окремого випадку, за рахунок можливості зменшення ємності із-за зменшення різноманіття даних;
- попередня інформація в даних не використовується як додаткова.

Опираючись на ці умови, можна сказати, що такий метод використання додаткової інформації о умовах проведення експериментальних досліджень не співпадають з умовами навчання штучної нейронної мережі в цьому дослідженні. Тому цей метод не буде використовуватися при прогнозуванні властивостей.

Другий метод використання додаткової інформації для покращення якості прогнозування штучної нейронної мережі, це використання рекурентних штучних нейронних мереж, прикладом якої є комірka LSTM, що показана на рис. 3.10.

Цей метод базується на архітектурній можливості рекурентних штучних нейронних мереж приймати на вхід вектору попереднього контексту:

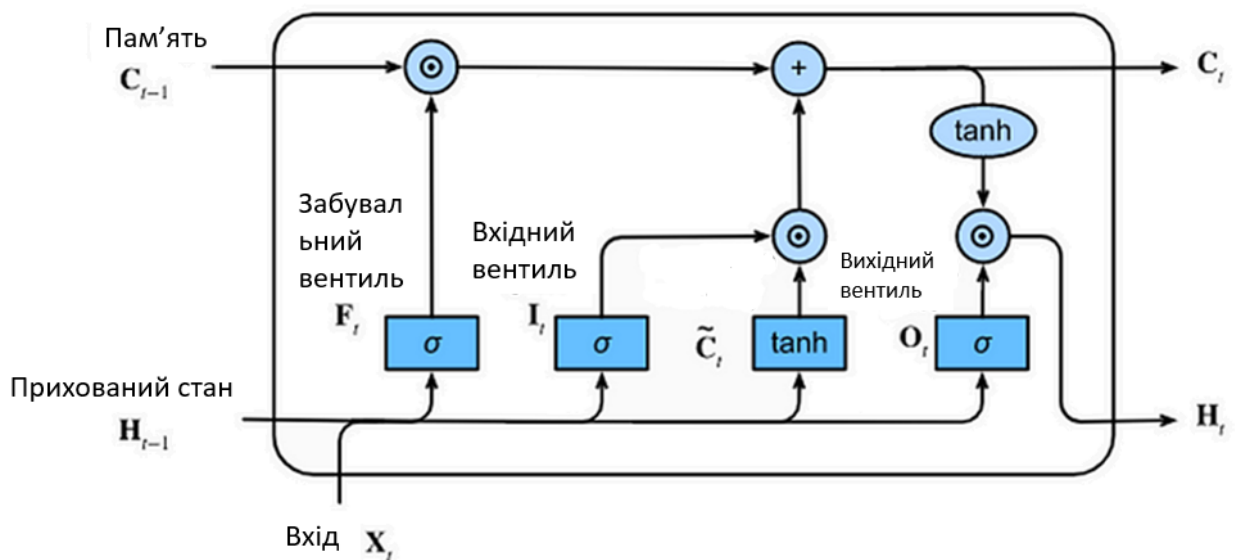


Рисунок 3.10 – Комірka LSTM на якій базуються рекурентні нейронні мережі[100]

На рис. 3.10 такий вектор попереднього контексту позначений H_{t-1} , та оброблений вектор H_t поточного контексту. Легко побачити, що використовуються вектор попереднього контексту у комірці LSTM у всіх операціях.

Але умова яка накладається на використання рекурентних мереж, це однорідність даних, та їх послідовна залежність у часі. Цим умовам дані про метаматеріали не відповідають, через те, що:

- інформація про умови дослідження не є структурно тотожними до інформації о структурі;
- між умовами експериментального дослідження та структурою метаматеріалу нема часової залежності

Отже, такий варіант використання додаткової інформації не підходить для покращення прогнозування штучної мережі в цьому дослідженні

Третій спосіб закріплення додаткової інформації називається атрибутивним способом закріплення [101].

Такий метод припускає додавання додаткової інформації у вхідний тензор як, наприклад додаткових каналів у структуру, буцім то 2D або 3D формату. Також, якщо вхідний тип даних являє собою вектор, можливе додавання додаткових точок у цей вектор. Розглянемо декілька прикладів атрибутивного закріплення додаткової інформації.

Нехай вхідним типом даних є плоске зображення RGB формату кодування кольорів. Це означає, що розмірність вхідного тензора буде:

$$img = (height, width, channels), \quad (3.5)$$

де *height* – висота зображення у пікселях

width – ширина зображення в пікселях

channels – канали кодування кольорів, у випадку RGB ця розмірність буде дорівнювати трьом.

Через те, що атрибутивний спосіб закріплення інформації можливий як кодування додаткових умов у зображення, з математичної точки зору можливе кодування у розмірності:

$$\text{context conditions} = (\text{height}, \text{width}, \text{conditions}), \quad (3.6)$$

де *context conditions* – тензор представлення умов

height – висота мапи представлення додаткових умов у пікселях

width – ширина мапи представлення додаткових умов у пікселях

conditions – канали інформації о додаткових умовах закодовані у мапу представлення.

Таким чином, можна побачити, що закріплення додаткових умов обмежене розмірністю вхідного зображення, так як для атрибутивного закріплення буде використана операції конкатенації вхідного тензору з мапою представлення додаткових умов у результуючий вхідний вектор:

$$\text{attributed img} = \text{Concatenate}(\text{img}, \text{context conditions}, \text{axis} = 3) \quad (3.7)$$

$$\text{attributed img} = (\text{height}, \text{width}, \text{channels} + \text{condtions})$$

де *attributed img* – вхідний тензор представлення зображення з атрибувано закріпленими умовами

Concatenate – операція конкатенації тензорів

axis – параметр конкатенації який вказує розмірність по якій така конкатенації буде відбуватися.

Так як первинний блок штучної нейронної мережі для обробки зображень є блок згорткових операцій, то й така контекстна інформація повинна мати структурну залежність з вхідним зображенням.

Як приклад таких контекстних умов може бути семантична карта, в якій для кожного пікселя існує відповідне значення класу. Наприклад: одяг, людина, тварина, рослина, будівля.

Також таким чином можна задати карту глибини. Тобто існує інформація о дистанції від об'єктиву до об'єкту на зображенні. Схожим чином може бути закодован рівень освітленості на зображенні

Тобто, всі умови мають очевидний структурний зв'язок з вхідним зображенням, такі умови також не будуть абсолютно дубльованими, як наприклад інформація о типі камери якою було зроблено зображення. Якщо б було потрібно кодувати таку інформацію атрибутивним способом, вимоги до розмірності тензору призвели б до того, що контекстний вектор мав би *height*widths* дуплікованих даних.

Іншим варіантом буде атрибутивне закріплення інформації для формату вхідних даних у вигляді вектору.

Розглянемо задачу, в якій вхідним масивом є вектор розмірністю n . По аналогії з розглянутим вище способом атрибутивного закріплення:

$$\text{input vector} = (n) \quad (3.8)$$

$$\text{attributed input vector} = \text{Concatenate}(\text{input vector}, \text{context condit}$$

$$\text{attributed input vector} = (n + \text{condtions})$$

де *input vector* – вхідний вектор ознак

attributed input vector – вхідний вектор з атрибутовано закріпленими умовами

context conditions – вектор представлення умов

Якщо уважно подивитися на порядок операцій та розмірність атрибутованого вхідного вектора, можна легко побачити, що такий підхід просто додає *conditions* кількість додаткових умов.

Такий підхід не ставить вимоги до розмірності вектора представлення додаткових вимог окрім самого типу даних, не ставить вимоги до якоїсь структурної залежності. Такий спосіб атрибутованого закріплення підходить для використання в цьому дослідженні, але не способом закріплення умов до вхідного тензору, а спосіб закріплення додаткових умов у проміжних шарах повнозв’язної штучної нейронної мережі на етапі “*прийняття рішень*” базуючись на візуальному представленні з згорткового блоку нейронної мережі. Такий спосіб існує, розглянемо його нижче.

Четвертий відомий спосіб використання умовної додаткової інформації для покращення якості прогнозування це використання контекстного вектору та вбудовування його у архітектуру штучної нейронної мережі[102].

Зазначимо умови для використання такого підходу:

- довільна кількість комбінацій додаткових умов. Так як цей вектор буде додатково оброблюватися окремою гілкою штучної нейронної мережі, не існує жорсткої умови у співставленні умова-вхідні дані. Така незалежна гілка має можливість незалежно знаходити вплив умов на всі дані;
- умови можуть являти собою довільний тип даних, буцім то текст, аудіо, аналоговий та цифровий сигнал, зображення та відео. Ці умови досягається тим, що на кожний такий тип даних можливо спроектувати таку штучну нейронну мережу, яка робить перетворення *вхідні дані – вектор контексту*;
- такі додаткові умовами можуть мати як часовий зв'язок, так і не часовий. Наприклад, якщо стоїть задача обробки відеоматеріала, додатковими умовами можуть бути як освітлення на кадрі, так й наприклад, кут зйомки;
- нема обмеження на пропускну можливість штучної нейронної мережі. Через те, що для перетворення додаткових умов у формат вектору, може знадобитися використання значних за розміром блоків сучасних архітектур.

Порівнюючи умови використання такого методу покращення якості прогнозування з умовами дослідження в цій роботі, можна зробити висновок, що такий метод підходить.

Таким чином, об'єднуючи третій та четвертий спосіб закріплення, буде запропоновану архітектуру штучної нейронної мережі, в якій буде незалежний від представлення додаткових умов експериментального дослідження блок згорткової нейронної мережі для виявлення залежностей всередині структури метаматеріалу та перетворення їх до вектора.

Другий блок штучної нейронної мережі буде являти собою повнозв'язну штучну нейронну мережу, яка буде приймати рішення о прогнозування властивостей метаматеріалів.

Третій блок штучної нейронної мережі буде являти собою повнозв'язну штучну нейронну мережу для обробки додаткових умов та представлення їх у вектор представлення розмірністю *conditions*, який буде атрибутивно закріплюватися з векторним представленням структурної інформації.

3.4 Архітектура згорткової нейронної мережі для прогнозування коефіцієнту пропускання в залежності структури, складу та умов вимірювання властивостей матеріалів

Як вже було описано, архітектура згорткової нейронної мережі для прогнозування коефіцієнта пропускання в залежності структури, складу та умов вимірювання властивостей матеріалів буде складатися з трьох блоків.

Блок згорткової нейронної мережі для виявлення структурних залежностей та конвертації знайдених залежностей у векторне представлення.

Існує багато видів архітектур які відрізняються як глибиною архітектури, кількістю згорткових фільтрів та розмірами таких фільтрів, так і типом зв'язку між згортковими блоками, як наприклад в архітектурі ResNet на рис. 3.11.

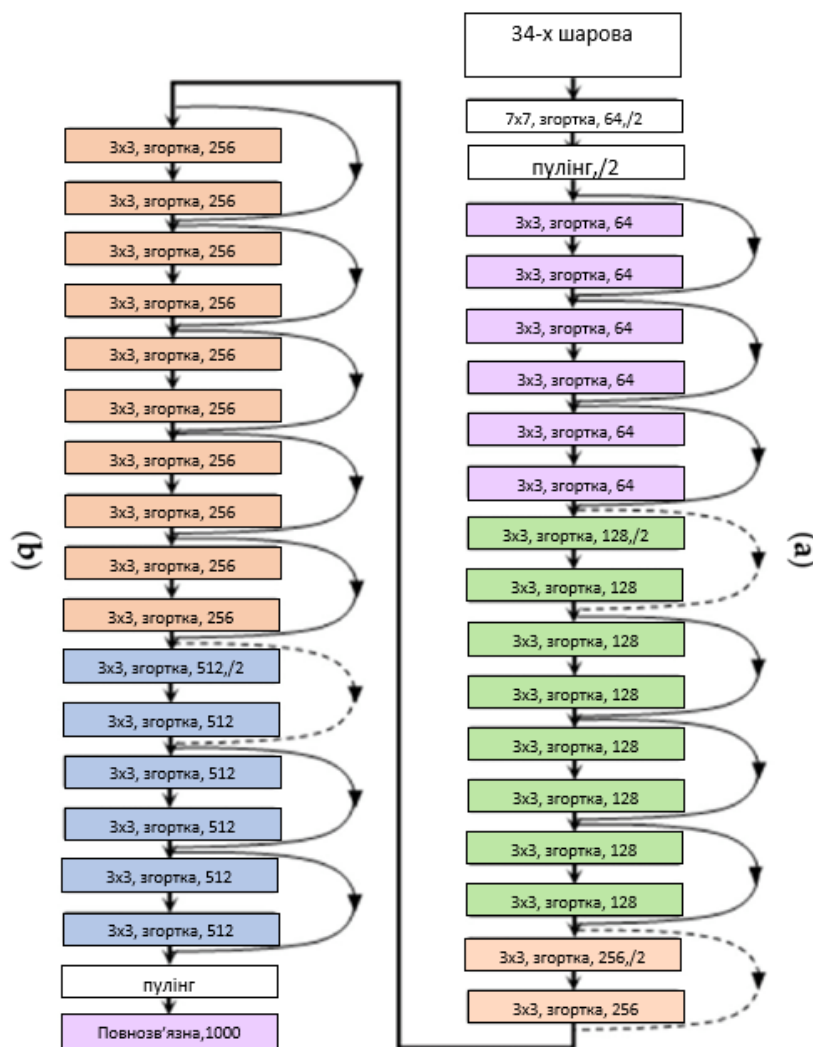


Рисунок 3.11 – Графічне представлення штучної згорткової нейронної мережі типу ResNet[103]

Особливістю такої архітектури є те, що між блоками шарів згорткової нейронної мережі використовуються так звані “*skip-connections*”, які математично описуються:

$$x = \text{convolutional block}(x \text{ prev}) + x \text{ prev},$$

де x – оброблені дані на даному блоці згорткок

convolutional block – дві згорткові операції на даному блоці

x_{prev} – дані, що були оброблені на попередньому блоці.

Такий тип зв'язку між блоками згорткової нейронної мережі був запропонований та реалізований для вирішення проблеми “затухання градієнтів”. Ця проблема виникає при використанні надглибоких архітектур (мається на увазі кількість шарів). Ця проблема виникає із-за над глибокої залежності у обчислюванні градієнтів на кожному шарі штучної нейронної мережі. З'являється експоненційне зменшення значень градієнта що впливає на алгоритм зворотного розповсюдження помилки. Такі “*skip-connections*”, розділяють значення градієнту помилки шару на 2 складові, що дає змогу нівелювати експоненційне зменшення помилки. Це продемонстровано на рис. 3.12.

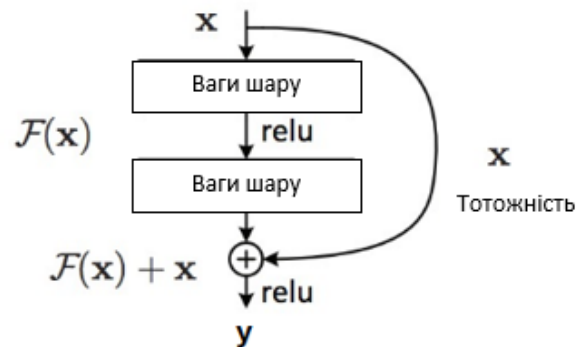


Рисунок 3.12 – вихідний блок нейронної зі “*skip-connection*”, де $F(x)$ – оброблені шаром мережі дані x , y – вихідні дані, які використовуються для розрахунку функції помилки $y = x + F(x)$, relu – функція активації нейронної мережі[103]

На прикладі спрощеної архітектури з рис. 3.12 розпишемо часткову похідну функції помилки від x використовуючи ланцюгове правило диференціювання:

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial x} \quad (3.9)$$

Значення похідної функції $y'(x) = 1 + F'(x)$, тому, підставляючи у формулу (3.19) отримуємо:

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial x} = \frac{\partial E}{\partial y} * \left(1 + \frac{\partial F}{\partial x}\right) = \frac{\partial E}{\partial y} + \frac{\partial E}{\partial y} * \frac{\partial F}{\partial x'} \quad (3.10)$$

де $\frac{\partial E}{\partial y}$ – значення часткової похідної функції помилки від вихідних даних

$\frac{\partial F}{\partial x}$ – значення часткової похідної функції блоку шарів штучної нейронної мережі від вхідних даних.

За аналогії з цим типом зв'язку у згорткових нейронних мережах існує також аналогічний зв'язок між блоками в рекурентних нейронних мережах спроектованих на базі комірок LSTM та GRU. Якщо звернутися до схеми LSTM з рис. 3.10, на ньому легко побачити такий маршрут який позначається значеннями векторів C_{t-1} та C_t [100]. Не дивлячись на те, що зазвичай рекурентні блоки нейронної мережі схематично виглядають як один блок, на практиці реалізації такий блок “розгортається” у n блоків відповідній довжині часової характеристики вхідних даних. Значення n – для деяких типів даних може сягати сотень, або тисяч.

Такий тип зв'язку став стандартом в сучасних архітектурах штучних нейронних мережах, так як глибина архітектур збільшується разом з обчислювальними можливостями. Але, збільшення глибини штучної нейронної мережі призводить до збільшення ємності, котре веде до тенденції перенавчання мережі при постійному значенні розмірів вибірки.

В випадку цього дослідження кількість даних, що були зібрані, є малою, тому нема ніякого сенсу до збільшення глибини архітектури та використання такого типу зв'язку.

З розвитком наноелектроніки почали створюватися компактні носимі девайси, що мають змогу локально обчислювати штучні нейронні мережі. Разом з цим виник попит на згорткові архітектури, що математично оптимізовані для використання на носимому пристрої. Такі архітектури це баланс між якістю передбачення, класифікації, детекції, тощо, та швидкістю, використанням батареї та RAM девайсу.

Зазвичай такі архітектури не мають строго математичного обґрунтування чому саме така архітектура була спроектована. Гіперпараметри таких архітектури підбираються або звичайним сітковим перебором, або ж знаходяться більш складними алгоритмами, як наприклад NAS – нейронний архітектурний пошук[104][105].

Із-за зазначених вище умов проектування таких архітектур, їх графічне представлення не виглядає інтуїтивно зрозумілими та легко пояснювальними. Так як основне поле використання таких архітектур це компактні девайси, то використання їх як бази для власної архітектури для прогнозування властивостей метаматеріалів не буде.

До початку експоненційного росту обчислювальних можливостей, кількості даних для навчання, використовувалися класичні архітектури штучних нейронних мереж, одним із прикладів є LeNet, яка предствалена на рис. 3.13. Стратегією проектування таких архітектур було пошарове накладання операцій згорток.



Рисунок 3.13 – схематичне зображення архітектури штучної нейронної мережі LeNet[34]

На рис. 3.13 зображено приклад такої класичної архітектури. На ній Зображення(28 x 28) – це вхідне зображення, Згортка – операція згортки, (n) – кількість фільтрів в операції згортки, stride – крок операції згортки, доп – довжина масиву нулів, яким доповнюється зображення для збереження розмірності даних всередині штучної нейронної мережі, Пулінг – операція пулінгу даних методом середнього значення, що зменшує розмірність даних всередині штучної нейронної мережі, Повнозв'язна(n) – шар повнозв'язної нейронної мережі з n – кількістю нейронів.

Можна побачити, що така архітектура простіша, аніж архітектури, що були приведені вище. Таким самим чином буде спроектований блок згорткової нейронної мережі для прогнозування коефіцієнта пропускання в залежності структури, складу та умов вимірювання.

Так як вхідні дані мають 3D структуру плюс інформаційні канали, то використання з очевидних причин 2D операцій згортки не є можливим. Для такого типу даних використовуються 3D ядра згортки, що зображені на рис. 3.14.

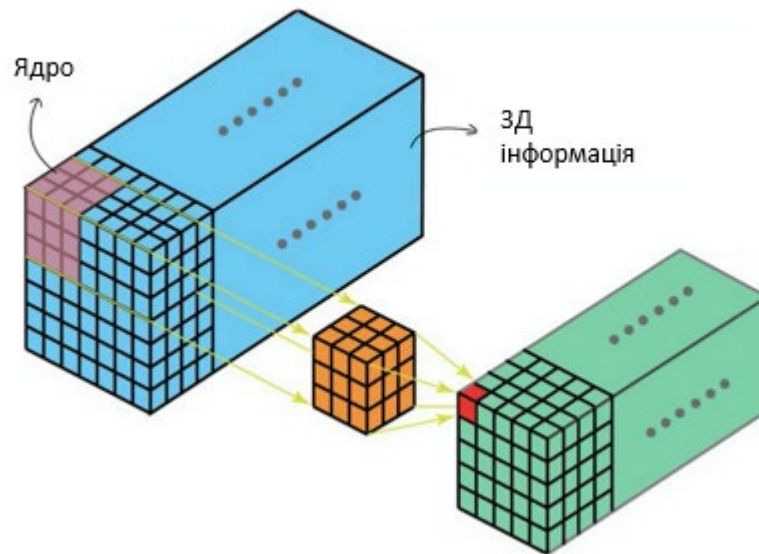


Рисунок 3.14 – візуалізація використання 3D згортки[106]

Легко побачити з рис. 3.14, що операція згортки відбувається таким самим чином, як і для 2D випадку, відмінність лише у додаванні додаткової розмірності. На рис N зображено приклад використання повністю симетричного ядра згортки, але, ядро згортки може бути будь-якої форми. Не симетричне ядро згортки буде використовуватися в запропонованій архітектурі.

Наступним елементом згорткового блоку штучної нейронної мережі є шари нормалізації даних. Нормалізація даних в штучних нейронних мережах є важливим кроком для поліпшення швидкості та якості навчання мережі. Вона допомагає зробити процес навчання більш стійким до різних масштабів вхідних функцій та активацій внутрішніх шарів мережі. Основні види нормалізації даних включають:

BN (batch normalization), тобто *пакетна нормалізація*. Ця нормалізація є однією з найпоширеніших форм нормалізації. Вона застосовується до виходів активацій внутрішніх шарів для кожного пакету вхідних даних. Математично пакетна нормалізація описується системою[107]:

$$\left\{ \begin{array}{l} \mu_B = \frac{1}{m} \sum_{i=1}^m X_i \\ \sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (X_i - \mu_B)^2, \\ \hat{X}_i = \frac{X_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \\ Y_i = \gamma \hat{X}_i + \beta \end{array} \right. \quad (3.11)$$

де X_i - вхідні дані

$\hat{X}_i$ - центровані та масштабовані вхідні дані

Y_i - дані, що пройшли шар пакетної нормалізації

γ та β – лінійний коефіцієнт та зсув пакетної нормалізації які також навчаються під час тренування штучної нейронної мережі

ε – малий параметр для стабілізації розрахунків для випадку стандартного відхилення, яке дорівнюватиме нулю.

Це допомагає стабілізувати розподіл активацій та зменшити проблему градієнтного зникнення. Як виходить з формули (3.11), BN виконує нормалізацію та центрування активацій, а потім застосовує масштабування та зміщення з параметрами, які навчаються під час процесу навчання.

Іншими типами нормалізації є Instance Normalization, Group Normalization та Layer Normalization [108-110]. Всі вони використовуються такі ж самі формули (3.11) для пакетної нормалізації, змінюється лише кількість даних та елементи для яких рахується статистика. Наприклад, для Group Normalization, вхідний пакет розбивається на підгрупи даних для яких окремо рахується статистика. Instance Normalization рахує статистику окремо для кожного елементу в пакеті.

Вибір типу нормалізації залежить від типу даних, що будуть подаватися на вхід штучної нейронної мережі. Якщо, наприклад, дані в середині пакету суттєво відрізняються один від іншого, зазвичай використовують Instance Normalization. Такий самий випадок із зібраними даними для цього дослідження.

Тому, типом нормалізації в штучній нейронній мережі для прогнозування коефіцієнта пропускання в залежності структури, складу та умов вимірювання властивостей метаматеріалів буде Instance Normalization.

Наступним кроком для проектування штучної нейронної мережі буде вибір функції активації.

Функції активації є ключовими компонентами нейронних мереж, оскільки вони надають нелінійність та здатність моделювати складні взаємозв'язки між вхідними та вихідними даними. Вони дозволяють мережі здійснювати апроксимацію більш різноманітних функцій, які важко або навіть неможливо виразити за допомогою простих лінійних зв'язків. Основні види функцій активації будуть представлені на рис. 3.15 – 3.17.

Логістична функція. Ця функція обмежена в діапазоні $(0,1)$, що робить її придатною до використання в задачах бінарної класифікації. Математично логістична функція описується як:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (3.11)$$

Побудуємо графік згідно формули (3.11), отримуємо:

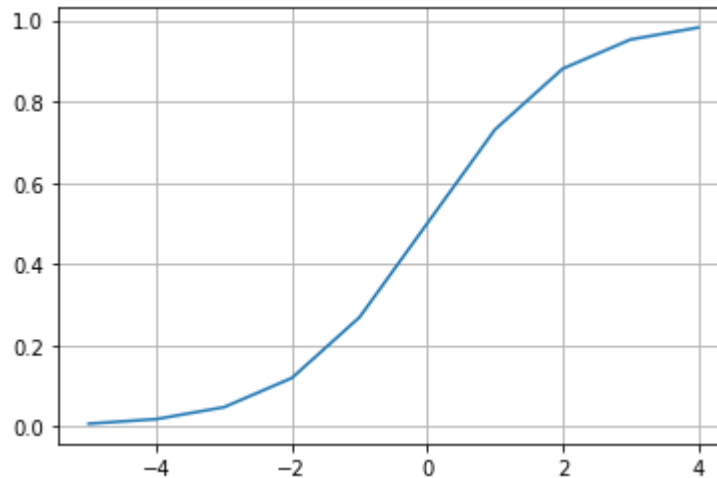


Рисунок 3.15 – Графік логістичної функції

В цій роботі не стоїть задачі бінарної класифікації, та нема потреби використання властивості “насичення”, які впливають з рис. 3.15, тому і використовуватися така функція не буде.

Наступною вже класичною функцією активації є гіперболічний тангенс. Область значень цієї функції знаходиться у проміжку (-1,1). Це означає, що така функція активації підтримує можливість роботи з від’ємними даними. Математично гіперболічний тангенс описується як:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (3.12)$$

Якщо побудувати графік гіперболічного тангенсу згідно формули (3.12), то отримуємо:

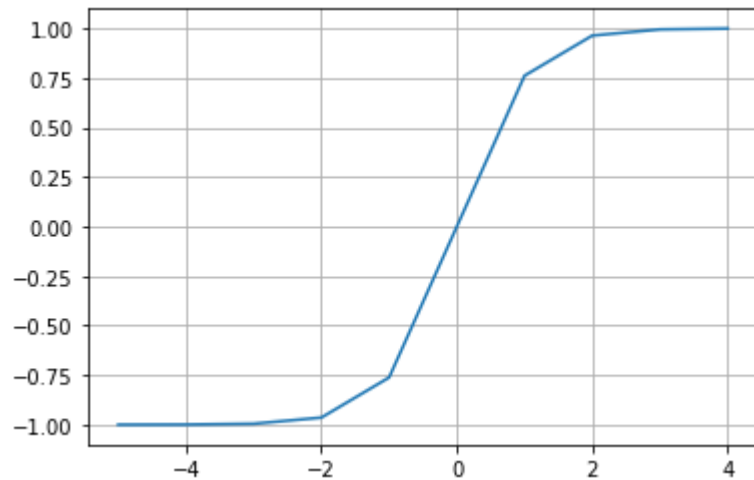


Рисунок 3.16 – Графік гіперболічного тангенсу

Але, у такої функції є також проміжок насичення, який для задачі прогнозування коефіцієнтів пропускання буде погіршувати якість роботи штучної нейронної мережі.

В сучасних архітектурах штучних нейронних мереж використовується таке сімейство функцій активації як ReLU.

Математично функція активації описується як:

$$ReLU(x) = \max(0, x), \quad (3.12)$$

Графік такої функції згідно формули (3.12):

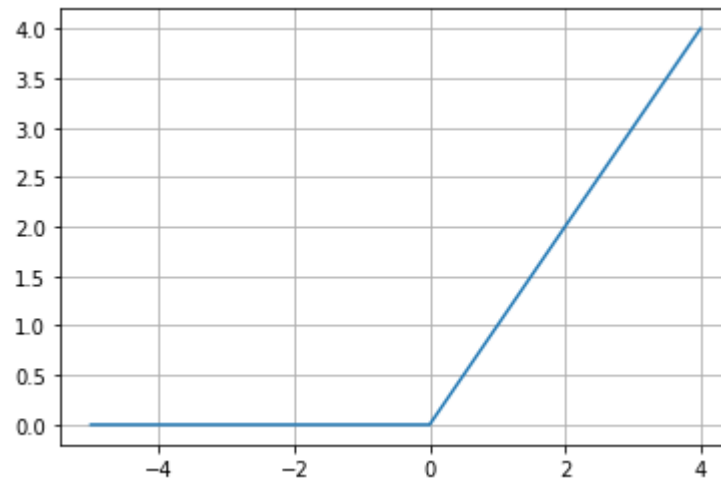


Рисунок 3.17 – Графік функції ReLU

Тобто згідно рис. 3.17, ця функція перетворює в нуль будь-які значення, що менші за 0, та залишає такими самими значення в іншому випадку.

Існує також декілька модифікацій цієї функції як LeakyReLU, як замість занулення використовує маленький лінійний коефіцієнт.

Шляхом експериментів мною було визначено, що в архітектурі штучної нейронної мережі для прогнозування коефіцієнтів пропускання буде оптимально використовувати класичну реалізацію функції активації ReLU.

В якості шарів для обробки умов експериментальних досліджень та прийняття рішень о коефіцієнтах пропускання метаматеріалу буде використовуватися повнозв'язна штучна нейронна мережа з функцією активації ReLU.

Експериментальні умови будуть кодуватися в штучну нейронну мережу за допомогою атрибутивного закріплення контекстного вектора.

Таким чином на рис. 3.18, представлена архітектура штучної мережі для прогнозування коефіцієнта пропускання в залежності структури, складу та умов вимірювання властивостей метаматеріалів.

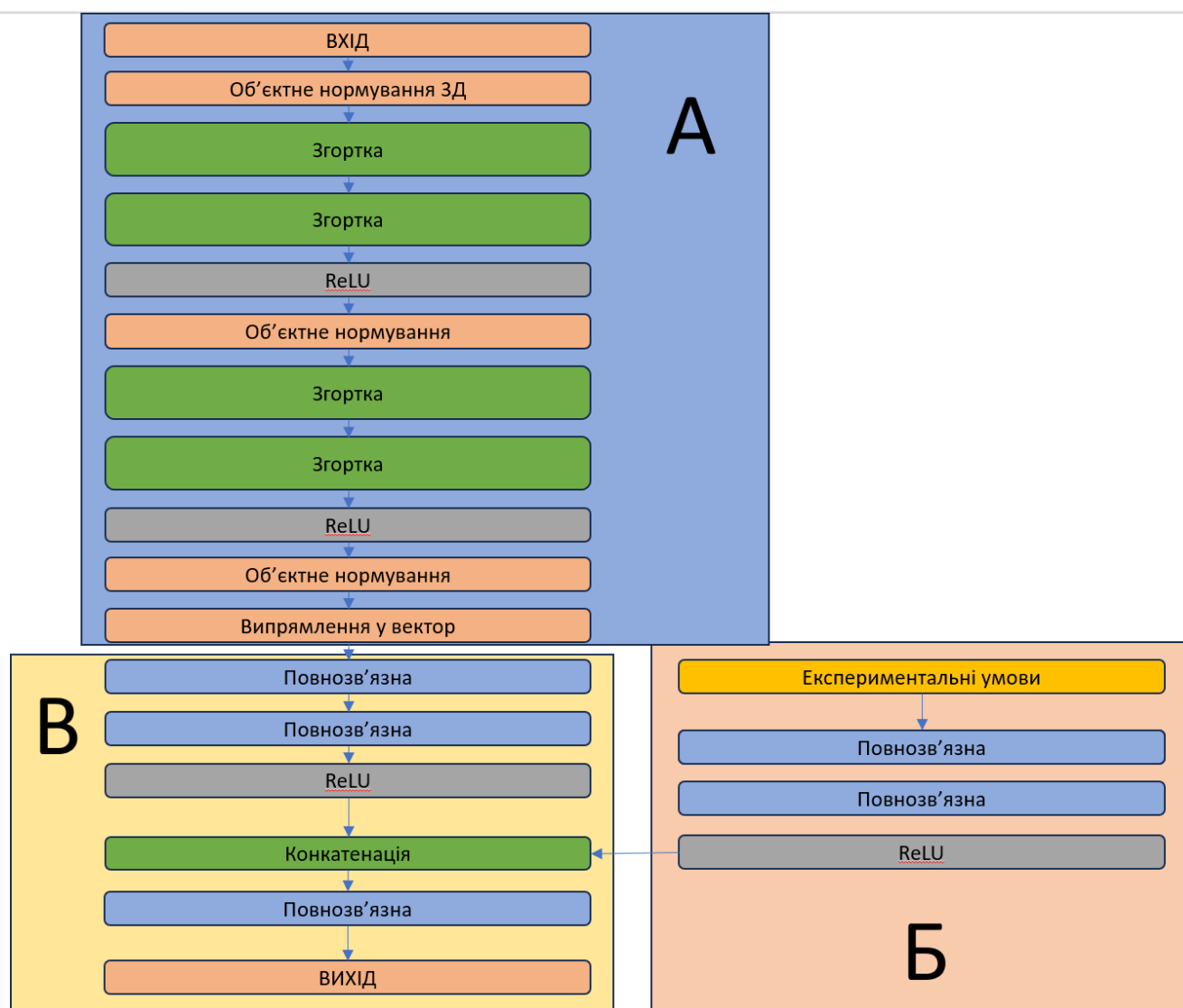


Рисунок 3.18 – Графічне представлення архітектури для прогнозування коефіцієнта пропускання в залежності від умов експериментальних досліджень

На цій схемі:

А) - блок згорткової нейронної мережі для виявлення структурних залежностей в метаматеріалі та представленні його у вигляді вектора за допомогою операції Flatten (операція яка випрямляє 2D, 3D тип даних у вектор);

Б) - блок обробки та представлення у вигляді вектора інформації о умовах проведення дослідження;

С) - блок прийняття рішення о значеннях коефіцієнту пропускання на основі атрибутованого контекстним вектором з блоку Б.

Для імплементації штучної нейронної мережі не вистачає гіперпараметрів архітектури, які являють собою: кількість фільтрів згортки та розміри ядра згортки, тип доповнення гулями, розмірність повнозв'язних шарів, та інше. Ці гіперпараметри будуть визначені в наступному розділі.

Висновки до розділу 3

В цьому розділі була поставлені задачі по розробленню конвеєр обробки вхідних даних, методу закріплення інформації про умови експерименту до структури та складу метаматеріалів, та проектування архітектури штучної нейронної мережі для прогнозування коефіцієнту пропускання в залежності структури, складу та умов вимірювання властивостей метаматеріалів.

Був запропонований та описаний конвеєр обробки вхідних даних, який містить в собі:

- структурні дані метаматеріалів. Перетворення формату .xyzrgb у *NumPy* масив з масштабуванням за допомогою min-max алгоритму;

- експериментальні дані. Перетворення текстового формату у *NumPy* масив з масштабуванням за допомогою min-max алгоритму;
- умови експериментальних досліджень. Їх масштабування;
- спосіб зберігання оброблених даних на диску/

Існує декілька методів закріплення інформації, а саме “зоопарк моделей”, “атрибутивне закріплення”, “контекстний вектор”.

“Зоопарк моделей” не підходить для цієї задачі, так як існує значна кількість комбінацій умов експериментальних досліджень, що призведе до потреби тренувати нові моделі під нові комбінації умов.

Використання “атрибутивного закріплення” буде надлишковим в цій задачі, так як вхідний тип даних має просторі залежності, а експериментальні умови – ні. Таким чином, використання такого методу потребуватиме внесення не структурної інформації до структурної, що призведе до дуплікації вхідних даних у $\text{length} \times \text{width} \times \text{height}$ разів, що для тривимірного зображення $> 1\,000\,000$ разів.

Сучасні архітектури згорткових нейронних мереж мають значну глибину шарів. Це пов’язано з тим, що зі збільшенням початкової вибірки даних існує потреба у збільшенні ємності штучної нейронної мережі. В такі архітектури закладений механізм компенсації зникання, та вибуху градієнтів. Використання таких архітектур на малій кількості даних, як в цьому дослідженні, призведе до перенавчання моделі.

Оптимальною буде архітектура LeNet з меншою кількістю шарів і параметрів. Така модель буде мати меншу схильність до перенавчання.

Вхідні дані потребують нормалізації, для покращення якості передбачення. Існують декілька алгоритмів нормалізації, а саме стандартизації та мін-макс масштабування. Для цієї задачі був обраний алгоритм мін-макс масштабування для вхідних даних.

Додаткові умови експериментальних досліджень також потребують нормалізації. Для підтримки однорідності даних у штучній нейронній мережі мін-макс масштабування буде оптимальним вибором.

Одним з шарів в штучній нейронній мережі є функція активації, яка вносить нелінійність у штучну нейронну мережу. Існує значна кількість різних видів таких функцій активацій для різних задач, таких як прогнозування, класифікація та генерація. Функція активації ReLU підходить по всім критеріям для цього дослідження.

Процес побудови штучної нейронної мережі включає в себе і підбір гіперпараметрів архітектури, таких як кількість шарів, кількість параметрів у шарів, типи регуляризації

РОЗДІЛ 4.

ПРОГНОЗУВАННЯ КОЕФІЦІЄНТУ ПРОПУСКАННЯ ЕЛЕКТРОМАГНІТНИХ МЕТАМАТЕРІАЛІВ

4.1 Постановка задачі

Основною задачею цього розділу буде проведення навчання штучної нейронної мережі для прогнозування коефіцієнта пропускання в залежності структури, складу та умов вимірювання властивостей метаматеріалів, архітектура якої була спроектована у розділі 3.

Для проведення дослідження потрібно сформувати повний конвеєр навчання штучної нейронної мережі, який буде містити:

- метрику для аналізу якості процесу навчання;
- оптимізатор – алгоритм градієнтного спуску та його гіперпараметри;
- поділені вибірки основного датасету на тренувальні та навчальні;
- обрані гіперпараметри архітектури штучної нейронної мережі, а саме: кількість фільтрів у згортковому блоку, тип доповнення нулями, тип регуляризації для уникнення ефекту перенавчання, кількість згорткових та повнозв'язних шарів, кількість нейронів у повнозв'язних шарах у блоках кодування умов експериментальних досліджень та прийняття рішень, розміри ядер згортки у згорткових операціях, крок згорткових операцій, розмір ядер пулінгу для зменшення розмірності даних всередині штучної нейронної мережі;

- методи регуляризації навчання за межами штучної нейронної мережі для уникнення ефекту перенавчання та їх параметри;
- візуалізації графіків значень функції помилки від кроку епохи;
- візуалізація результатів прогнозування для аналізу якості роботи натренованої штучної нейронної мережі.

Потрібно дати оцінку результатів згідно проведеного аналізу отриманих результатів, запропоновані можливі вирішення тих чи інших аспектів, які негативно вплинули на результат навчання.

Треба запропонувати варіанти розв'язку проблеми зі збором даних, пришвидшення цього процесу та збільшення поля збору даних, вирішення проблеми нестачі так званої “розмітки” до існуючих електромагнітних метаматеріалів.

Також потрібно розглянути можливі архітектурні зміни для покращення якості прогнозування та алгоритми навчання, які можуть вплинути на результат прогнозування.

Існує потреба в проведенні огляду перспективи створення оберненого алгоритму прогнозування структури електромагнітного метаматеріалу, його складу та умов вимірювання в залежності від необхідних властивостей. Будуть запропоновані можливі варіанти вирішення такої проблеми, оглянуті можливі архітектури та підходи для генерації структур електромагнітних матеріалів для отримання потрібних властивостей. Будуть запропоновані вирішення проблем збору даних, типи навчань, для яких не потрібно збирати так звану “розмітку”.

Потрібно розглянути часові проміжки для проектування таких підходів, їх імплементація, складнощі, які можуть з'явитися на різних етапах будування конвеєру

4.2 Навчання нейронної мережі та тестування нейронної мережі

Як було визначено в постановці задачі, для проведення навчання потрібно визначити всі гіперпараметри конвеєру навчання, які складаються з:

- гіперпараметри архітектури штучної нейронної мережі;
- функція помилки та метрики для аналізу якості прогнозування;
- гіперпараметри циклу навчання;
- оптимізатор градієнтного спуску та його гіперпараметри;
- розділення датасету на тестувальні та тренувальні вибірки.

Визначимо та опишемо гіперпараметри архітектури штучної нейронної мережі. Зазвичай гіперпараметри штучної нейронної мережі визначаються емпірично. Тобто, в процесі навчання дослідник вивчає та аналізує отримані результати впродовж циклів навчання, та вносить різні архітектурні та гіперпараметричні зміни. Але, існують вже відомі та дослідження емпіричні правила задання гіперпараметрів, на яких базується первинна архітектура:

1) Зі збільшенням кількості шарів операцій згортки збільшується і кількість фільтрів. Таке правило пов'язують з тим, що згорткова штучна нейронна мережа за своєю природою першими шарами знаходить глобальні признаки, тобто, якщо стоїть задача розпізнавання користувачів за їх обличчям, то на перших шарах згорткова нейронна мережа буде знаходити прості форми овалу лиця, а вже на наступних знаходитиме ніс, очі та інше.

З рис 3.11 видно, що кількість фільтрів з наступними блоками зростає, а саме на перших шарах(позначені фіолетовим кольором) кількість фільтрів 64, на

наступних(позначені зеленим кольором) – 128, після цього(позначені червоним) йдуть фільтри на 256, а останній блок згорток вже використовує кількість фільтрів 512.

2) Розмір ядра згортки повинен зменшуватись з кількістю доданих шарів. Це не таке жорстке як правило 1), і в сучасних архітектурах його скоріше не використовують, а ніж використовують, але, якщо подивитися на рис 4.1, можна побачити, що перший шар має згорткове ядро розміром 7x7 пікселів, а наступні ядра мають розмір 3x3 пікселів. Це правило обумовлено такою ж самою ідеєю про те, що згорткова нейронна мережа виділяє спочатку глобальні признаки, а вже потім часткові.

3) Кількість штучних нейронів у шарах повнозв'язної штучної нейронної мережі мають збільшуватися до серединних шарів та зменшуватися до фінального виходу архітектури. Це емпіричне правило, але теоретично воно обумовлене тим, що для якісної внутрішньої репрезентації інформації потрібне великий латентний простір, та поступове зменшення розмірності для ефективного кодування інформації перед прийняттям рішення.

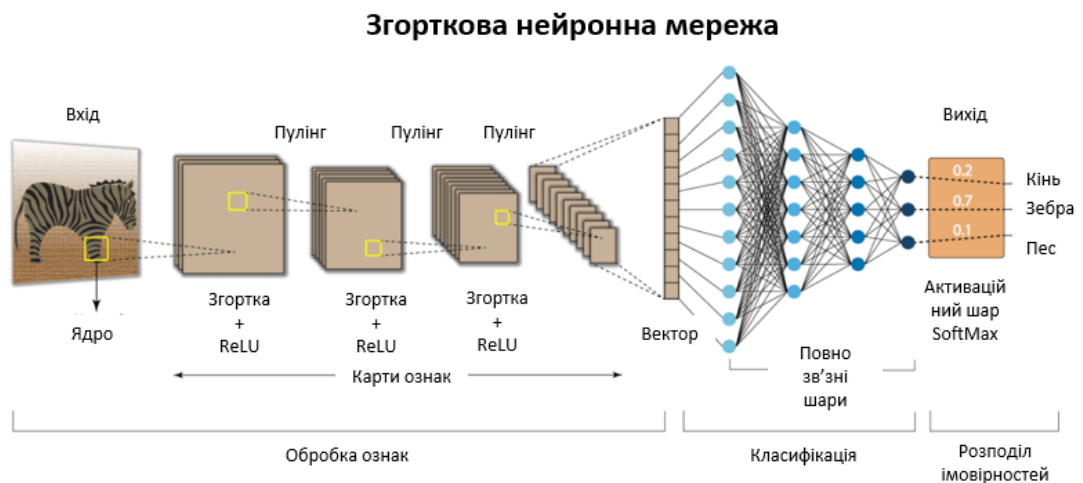


Рисунок 4.1 – Приклад архітектури з використанням повнозв'язних шарів[111]

З рис. 4.2 – видно, як вектор візуальної репрезентації зображення поступає на розширений шар повнозв'язної штучної нейронної мережі, та зменшення кількості штучних нейронів біля вихідного шару.

Опираючись на ці правила та шляхом експериментальних циклів навчання та аналізу отриманих результатів, були задані такі гіперпараметри архітектури штучної нейронної мережі

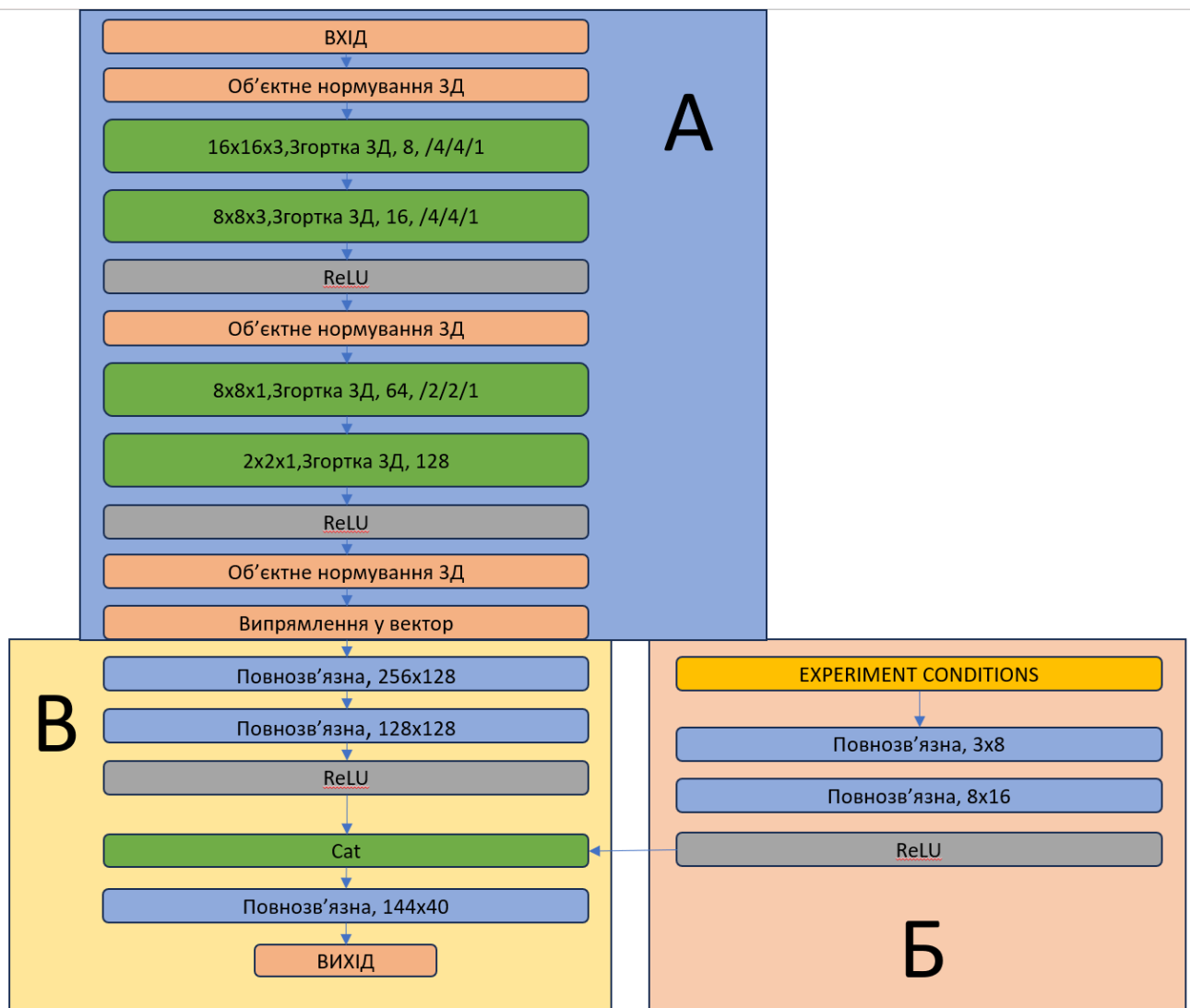


Рисунок 4.2 – Архітектура штучної нейронної мережі з урахуванням гіперпараметрів архітектури

Як видно з рис. 4.2 для проектування архітектури використовувалися емпіричні правила 1), 2), 3). Такий набір параметрів був отриманий шляхом експериментів.

Таблиця 4.1

Згортковий блок А	Повнозв'язний блок В	Повнозв'язний блок кодування умов досліджень С	Значення функції середньоквадратичної помилки
Кількість фільтрів: [8,16,64,128] Розмір ядра згортки [16x16x3,8x8x3,8x8x1,2x2x1]	Ширина шарів: [256,128,144]	Ширина шарів: [8,16]	0,025
Кількість фільтрів: [16,32,128,128] Розмір ядра згортки [8x8x3,4x4x3,4x4x1,2x2x1]	Ширина шарів: [512,256,160]	Ширина шарів: [16,32]	0,047
Кількість фільтрів: [8,8,16,64] Розмір ядра згортки [16x16x3,4x4x3,4x4x1,2x2x1]	Ширина шарів: [256,256,160]	Ширина шарів: [16,32]	0,089
Кількість фільтрів: [16,32,64,128] Розмір ядра згортки [16x16x3,16x16x3,4x4x1,4x4x1]	Ширина шарів: [256,128,160]	Ширина шарів: [8,16,32]	0,038
Кількість фільтрів: [8,16,64,128] Розмір ядра згортки [16x16x3,8x8x3,8x8x1,2x2x1]	Ширина шарів: [128,128,80]	Ширина шарів: [8,16]	0,041

Перебір оптимальних параметрів та порівняння помилки прогнозування описаний в таблиці 4.1

В якості регуляризації був обраний шар Dropout, принцип роботи якого показаний на рис. 4.3, який випадково “вимикає” від навчання нейронні зв’язки всередині мережі.

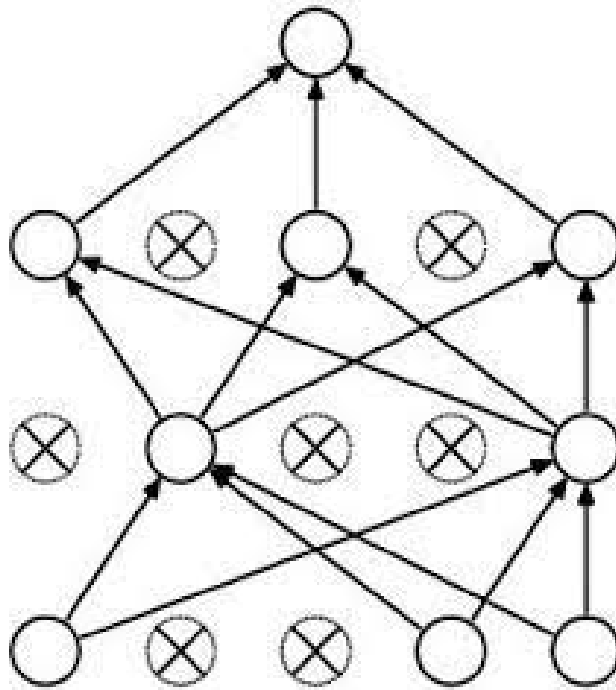


Рисунок 4.3 – візуалізація роботи Dropout на прикладі повнозв’язної нейронної мережі[112]

Як видно на рис. 4.3 частина штучних нейронів перестає бути зв’язаною. Гіперпараметр такого підходу – вірогідність відключення кожного нейронну у шарі. Для запропонованої архітектури нейронної мережі цей параметр дорівнює 35%. Використання цього шару “змушує” нейронну мережу перевикористовувати закріплені зв’язки, що дає змогу уникати перенавчання.

Для уникнення ефекту перенавчання під час циклу навчання, було використано деякі прості техніки.

Перша техніка регуляризації циклу навчання носить назву “Data shuffling”, тобто “перемішування даних у датасеті”. Це досить простий й ефективний спосіб, який базується на алгоритмі, що представлений блок-схемою 4.4.

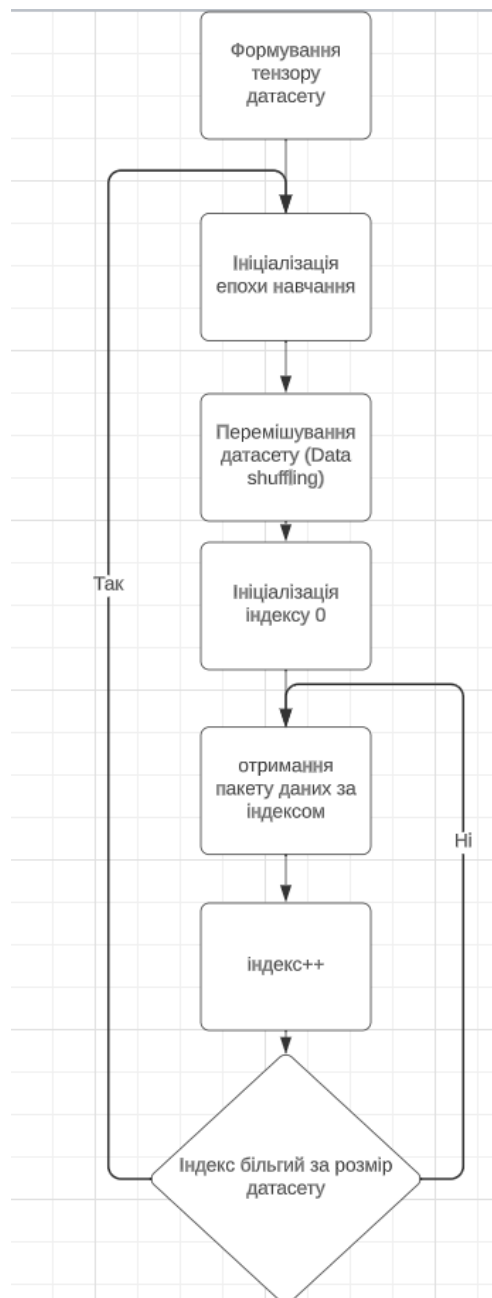


Рисунок 4.4 – Блок-схема алгоритму DataShuffling

На блок-схемі з рис. 4.4 процес перемішування даних відбувається у блоці “Перемішування датасету (Data shuffling)”. Ця операція відбувається після кожної епохи навчання. Це робиться для того, щоб нейронна мережа не запам’ятовувала порядок даних у датасеті.

Наступний метод регуляризації називається Early Stopping. Цей метод полягає у тому, що впродовж навчання фіксуються епохи при яких не відбулося зниження значень функції помилок на тестовому датасеті[113]. Через те, що перенавчання проявляється у зниженні значень функції помилки на навчальному датасеті та збільшення або не змінність значення функції помилки на тестовому датасеті, то такий алгоритм дозволяє відстежити початок процесу перенавчання, та зберігати фіксувати ваги моделі ще не перенавченної штучної нейронної мережі.

Вибір алгоритму оптимізації зазвичай базується на експериментальних запусках навчання. Так як простір значень функції помилки займає n -мірний простір, він не подається комплексному аналізу для вибору оптимізатора, тому простіше перебрати існуючі оптимізатори для визначення оптимальнішого під конкретний датасет та архітектуру штучної нейронної мережі.

Емпіричним шляхом було визначено, що найкращим оптимізатором для задачі прогнозування коефіцієнтів пропускання для зібраних в цьому дослідженні даних є NAdam.

NAdam (Nesterov-accelerated Adaptive Moment Estimation) є оптимізатором, який поєднує дві популярні техніки оптимізації: Nesterov Accelerated Gradient (NAG) та Adam. Він поєднує переваги обох підходів, щоб покращити швидкість та ефективність навчання нейронних мереж[114].

Для розуміння роботи оптимізатора NAdam, визначимо кроки обчислень за оновлення вагів в алгоритмі Adam:

1. Обчислення експоненційно зваженого середнього градієнта m_t :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (4.1)$$

2. Обчислення експоненційно зваженого середнього квадрату градієнта v_t :

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (4.2)$$

3. Обчислення зміщених оцінок m_t, v_t :

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (4.3)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

4. Оновлення вагів штучної нейронної мережі:

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{v}_t} + \varepsilon} \hat{m}_t \quad (4.4)$$

де m_{t-1} - значення експоненційно зваженого середнього градієнта на минулому кроці

β_1, β_2 - змінні коефіцієнти, що виступають гіперпараметрами алгоритму

g_t - градієнт відносно параметрів нейронної мережі на ітерації t

α - крок навчання

ε - малий додаток для чисельної стійкості

θ_t - значення вагів штучної нейронної мережі на ітерації t .

Nesterov Accelerated Gradient (NAG) додає зміщення параметрів на певну відстань в напрямку моменту градієнту перед вирахуванням градієнта. Визначення моменту зсуву:

$$\tilde{m}_t = (1 - \beta_1) \cdot \nabla J(\theta_{t-1}) + \beta_1 \cdot \nabla J(\theta_{t-2}) \quad (4.5)$$

Таким чином, результуючою формулою оновлення вагів штучної нейронної мережі використовуючи оптимізатор NAdam буде:

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{v}_t} + \varepsilon} (\beta_1 \tilde{m}_t + (1 - \beta_1) g_t) \quad (4.6)$$

NAdam використовує зсунуті оцінки моменту та другого порядку, а також зсунутий градієнт моменту для покращення ефективності. Ця комбінація допомагає швидше збігатися під час навчання та підвищує стійкість оптимізаційного процесу[114].

Метрикою якості навчання було обрано MSE(mean squared error), або середньоквадратичну похибку. Розрахунок середньоквадратичної похибки відбувається за формулою:

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad (4.7)$$

де N - кількість даних у вибірці

$\hat{y}$ та y_i - пара значень прогнозованих та реальних даних.

Ця метрика зручна ще й тим, що вона відповідає критеріям, які ставляться до функції помилки для навчання штучної нейронної мережі.

Таким чином, ця ж функція була обраною як функція помилки для навчання штучної нейронної мережі.

Для проведення дослідження було обрано 3 види розбивки повного датасету. Це було зроблену для двох цілей:

- дослідити можливість штучної нейронної мережі узагальнювати дані для експериментальних досліджень з різним проміжком частоти опромінення;
- провести 3 дослідження на різній розбивці даних на тренувальну та тестувальну вибірки. Це потрібно для більш комплексного та об'єктивного аналізу можливості прогнозування експериментальних досліджень з вибіркою малої кількості.

Розглянемо три розбивки даних на підвибірки для кожного експерименту з навчанням:

Дані, частота випромінювання на які лежить у проміжку $[0.2, 0.8]$ ТГц. Для цієї підвибірки були обрані дані для тренування та тестування та створена таблиця для візуального представлення:

Таблиця 4.2

Тренувальні дані	Тестувальні дані
figure1A-1_co-pol_90_2.0_1.npz figure1A-2_cross-pol_90_2.0_1.npz figure1B-1_co-pol_90_2.0_1.npz figure5-1_co-pol_90_2.0_2.npz	figure1B-2_cross-pol_90_2.0_1.npz

figure5-2_cross-pol_90_2.0_2.npz	
----------------------------------	--

Співвідношення тренувальних до тестувальних даних в цій вибірці дорівнює 83%/17% відповідно. Проблема такої малої кількості даних є в тому, що не можна гарантувати те, що штучна нейронна мережа навчиться добре узагальнювати отриману інформацію з тренувального датасету, але в межах відомих структур, які нейронна мережа отримувала під час навчання, можна буде проаналізувати результати навчання та тестування

Наступна вибірка даних зроблена з даних частота випромінювання якої лежить в межах [136, 371] ТГц. Для візуалізації розподілу даних у цій вибірці була створена таблиця:

Таблиця 4.3

Тренувальні дані	Тестувальні дані
figure2-1_cross-pol_90_2.0_1.npz	figure2-3_cross-pol_30_2.0_1.npz
figure2-2_cross-pol_20_2.0_1.npz	figure3-10_cross-pol_90_1.5_2.npz
figure2-4_cross-pol_40_2.0_1.npz	
figure2-5_co-pol_90_2.0_1.npz	
figure3-11_cross-pol_90_1.3_2.npz	
figure3-12_cross-pol_90_1.5_2.npz	
figure3-1_cross-pol_90_2.0_1.npz	
figure3-2_cross-pol_90_2.0_2.npz	
figure3-4_cross-pol_90_2.0_2.npz	
figure3-5_cross-pol_90_2.0_2.npz	
figure3-6_cross-pol_90_2.0_2.npz	
figure3-7_cross-pol_90_2.0_2.npz	

figure3-8_cross-pol_90_2.0_2.npz	
figure3-9_cross-pol_90_1.73_2.npz	
figure4-1_cross-pol_90_2.0_1.npz	
figure4-2_co-pol_90_2.0_1.npz	

Співвідношення тренувальних до тестувальних даних в цій вибірці дорівнює 87.5%/12.5% відповідно. В цього проміжку даних було зібрано більше даних, що дає змогу більш об'єктивно оцінити спроможність нейронної мережі узагальнювати знання отримані з навчальної вибірки.

Третя підвибірка даних буде включати в себе всі дані обох проміжків частоти випромінювання:

Таблиця 4.4

Тренувальні дані	Тестувальні дані
figure1A-1_co-pol_90_2.0_1.npz	figure1B-1_co-pol_90_2.0_1.npz
figure1A-2_cross-pol_90_2.0_1.npz	figure3-8_cross-pol_90_2.0_2.npz
figure1B-2_cross-pol_90_2.0_1.npz	figure2-3_cross-pol_30_2.0_1.npz
figure2-1_cross-pol_90_2.0_1.npz	
figure2-2_cross-pol_20_2.0_1.npz	
figure2-4_cross-pol_40_2.0_1.npz	
figure2-5_co-pol_90_2.0_1.npz	
figure3-10_cross-pol_90_1.5_2.npz	
figure3-11_cross-pol_90_1.3_2.npz	
figure3-12_cross-pol_90_1.5_2.npz	
figure3-1_cross-pol_90_2.0_1.npz	
figure3-2_cross-pol_90_2.0_2.npz	

figure3-4_cross-pol_90_2.0_2.npz	
figure3-5_cross-pol_90_2.0_2.npz	
figure3-6_cross-pol_90_2.0_2.npz	
figure3-7_cross-pol_90_2.0_2.npz	
figure3-9_cross-pol_90_1.73_2.npz	
figure4-1_cross-pol_90_2.0_1.npz	
figure4-2_co-pol_90_2.0_1.npz	
figure5-1_co-pol_90_2.0_2.npz	
figure5-2_cross-pol_90_2.0_2.npz	

Така вибірка даних дасть змогу дослідити можливість штучної нейронної мережі одночасно узагальнювати залежність коефіцієнту пропускання від структури та умов дослідження для різних частот випромінення.

Згідно таблиць 4.1 - 4.3 було проведено три навчання, результати яких показані на рис. 4.5 - 4.13.

Для першої вибірки даних з таблиці 4.2 отримані були такі результати:

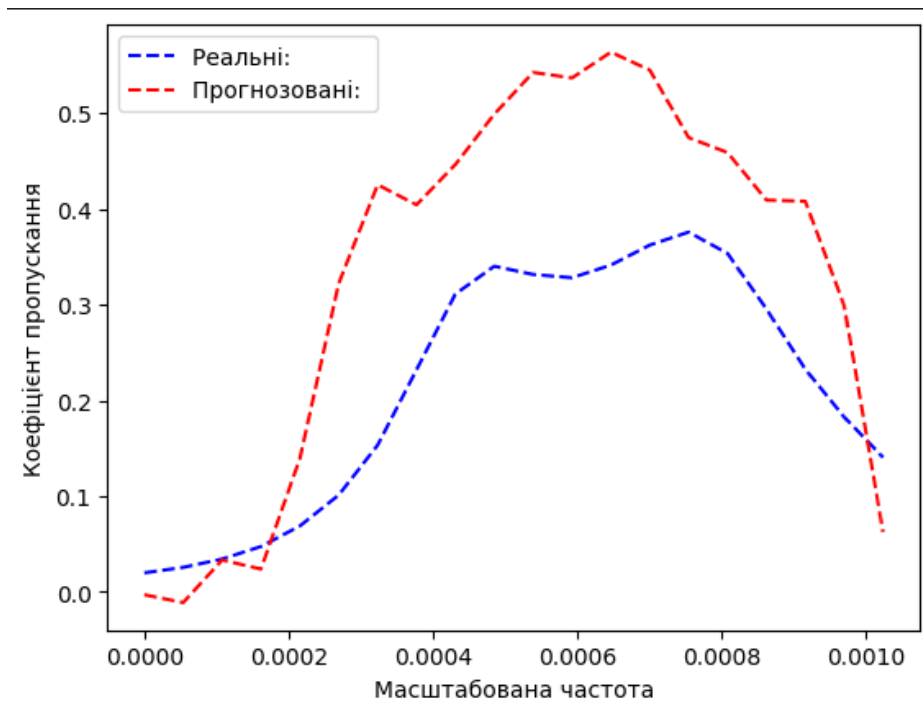


Рисунок 4.5 – Порівняння прогнозованого коефіцієнта пропускання та експериментально отриманого

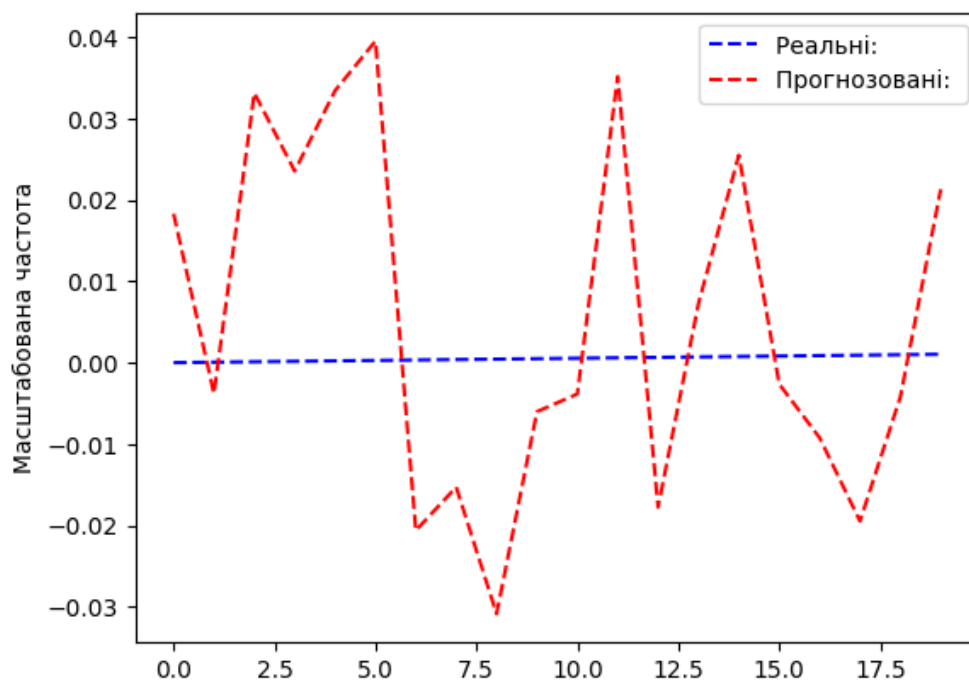


Рисунок 4.6 – порівняння прогнозованої частоти опромінення та експериментально встановленої

Можна побачити, що не зважаючи на малу кількість даних, штучна нейронна мережа змогла достатньо узагальнити інформації з тренувальної вибірки для прогнозування тренду для тестувальної вибірки. Звісно, такі значення передбачення не можуть вважатися точними.

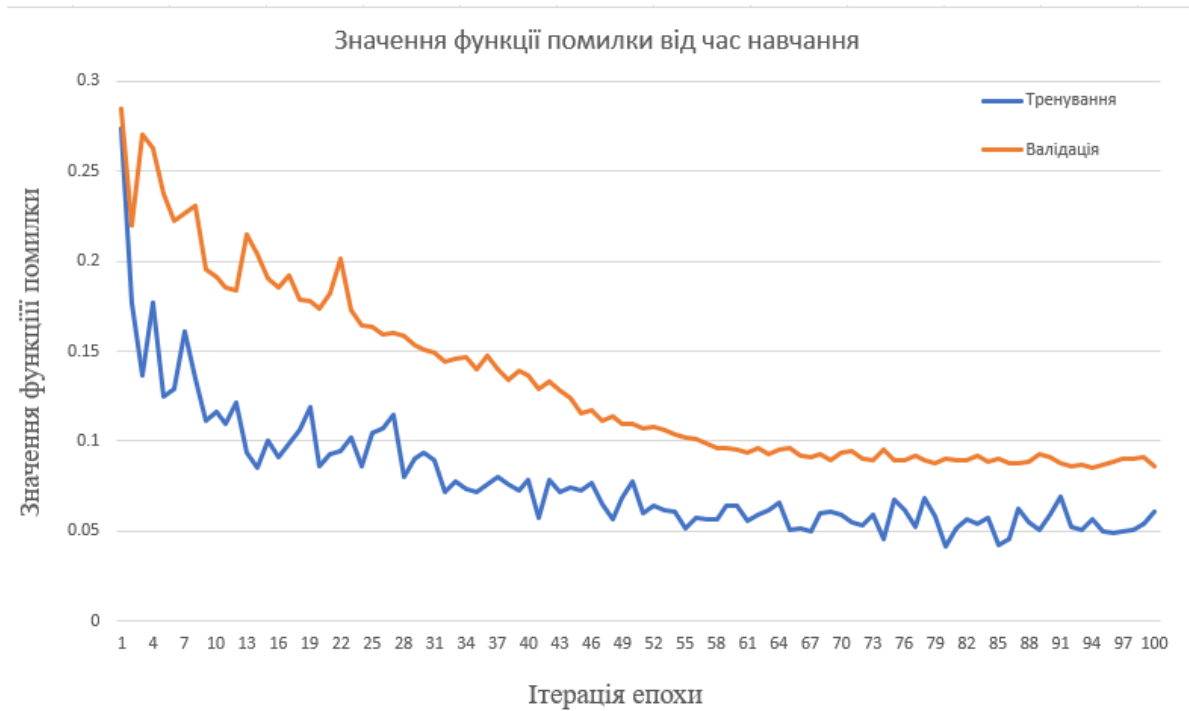
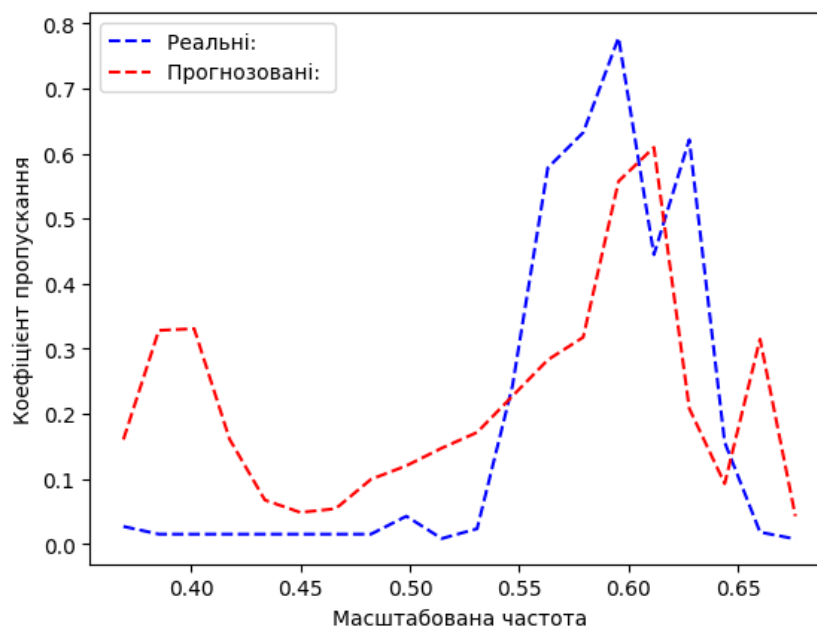


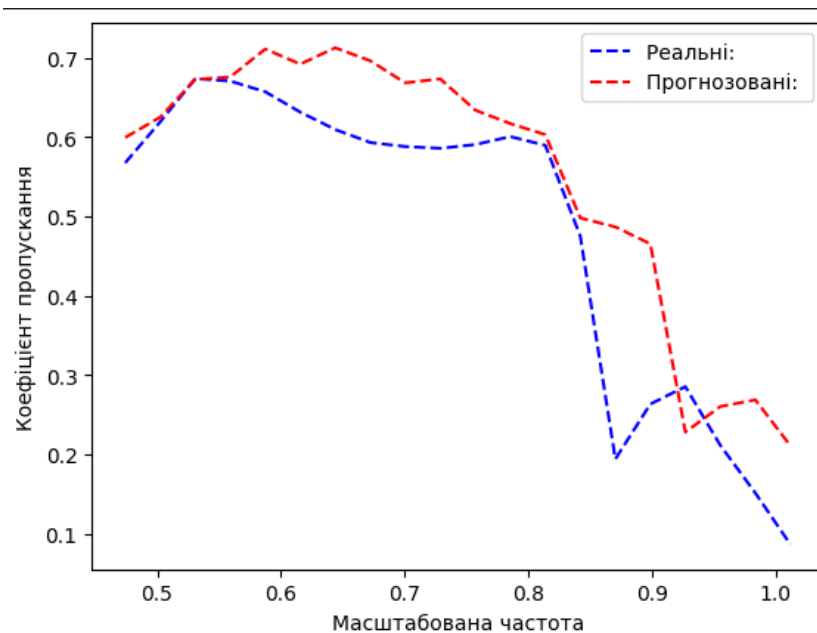
Рисунок 4.7 – Значення функції помилки під час навчання для тренувальної та тестувальної вибірок

Як видно з рис. 4.7 значення функції помилки для тренувальної та тестувальної вибірки мають спадаючий тренд, що свідчить про процес навчання.

Для даних другої вибірки з таблиці 4.3 було отримано такі результати:

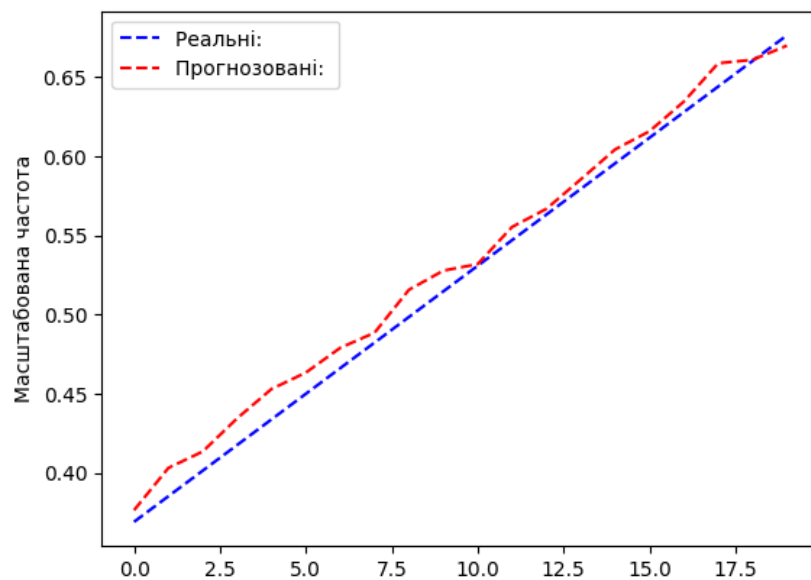


а)

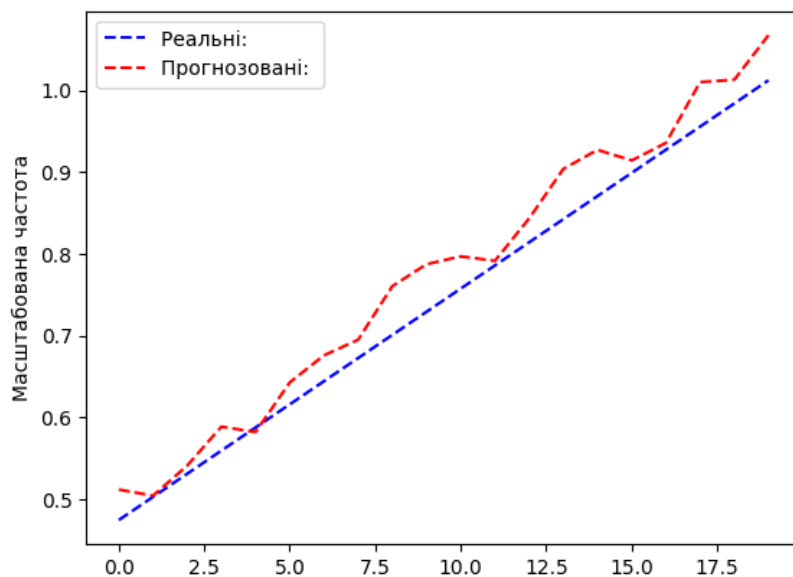


б)

Рисунок 4.8 – порівняння прогнозованого коефіцієнту пропускання та експериментально отриманого для структур з тестувальної вибірки а) figure3-10_cross-pol_90_1.5_2.npz, б) figure2-3_cross-pol_30_2.0_1.npz



а)



б)

Рисунок 4.9 – порівняння прогнозованої частоти опромінення та експериментально встановленої для структур з тестувальної вибірки а) figure3-10_cross-pol_90_1.5_2.npz, б) figure2-3_cross-pol_30_2.0_1.npz

Легко побачити, що збільшення даних для тренування призвело до того, що прогнозовані значення коефіцієнту пропускання лежать ближче до реальних, котрі були отримані під час експериментально дослідження.

Також, можна побачити для що для цього проміжку частоти опромінення нейронна мережа прогнозує більш стабільні та близькі до реальних значення

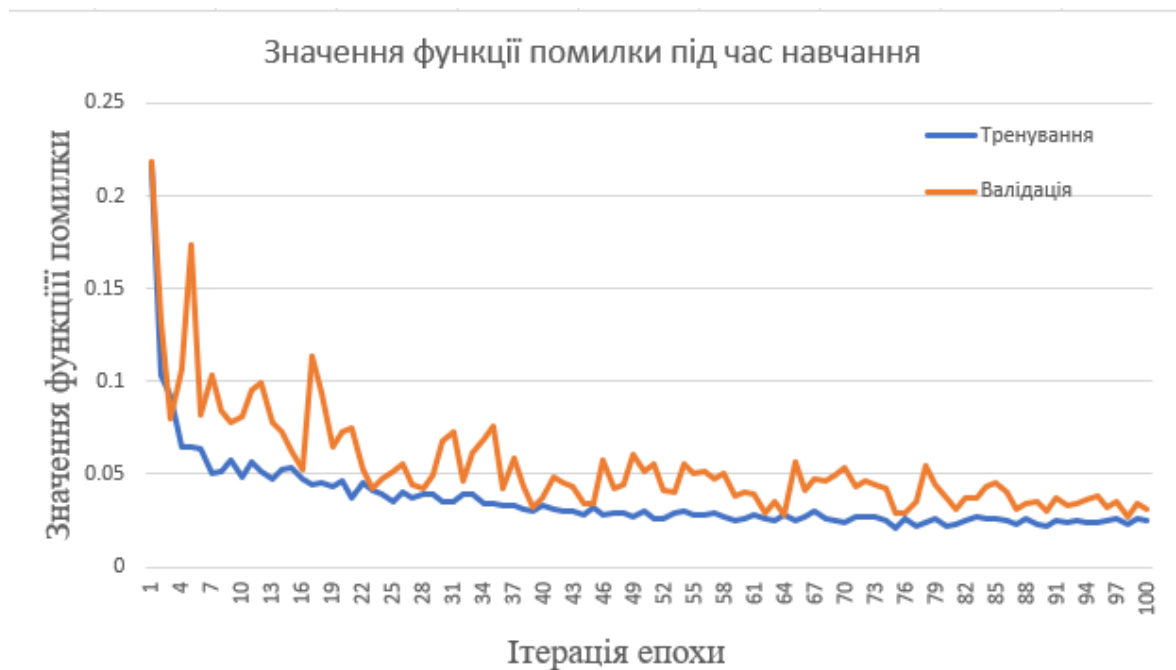
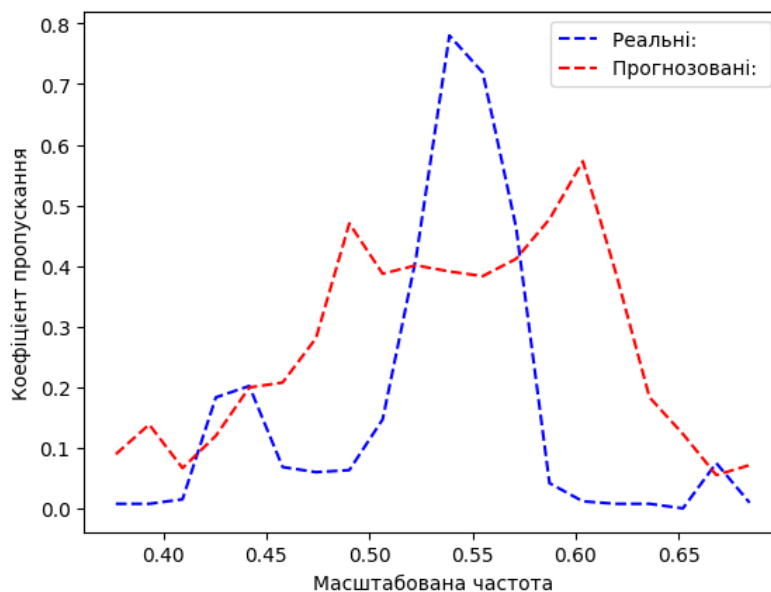


Рисунок 4.10 – Значення функції помилки під час навчання для тренувальної та тестувальної вибірок

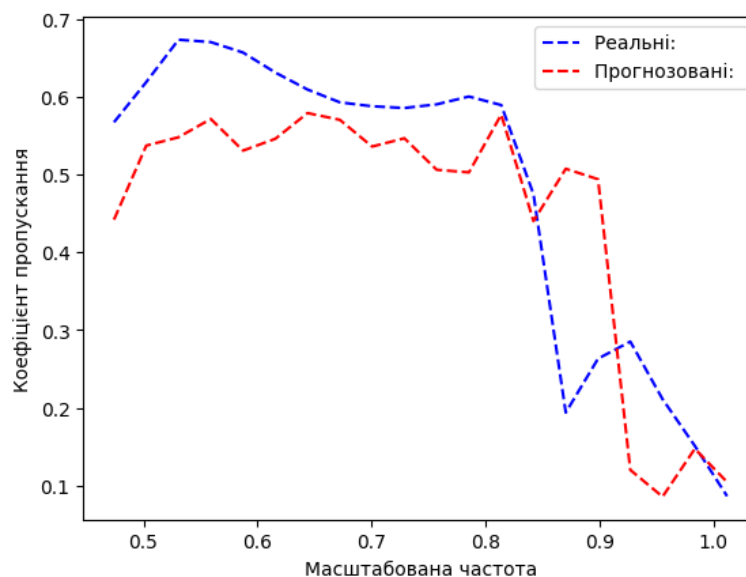
З графіку на рис. 4.10 видно, що у порівнянні з графіком рис 4.7 значення функції помилки на тестувальному датасеті більш наближене до тренувальної вибірки, що означає покращення узагальнюючої можливості штучної нейронної мережі. Також,

більш нестабільні значення для графіка з рис 4.10 у порівнянні з графіком з рис. 4.7 можуть бути пояснені збільшенням вибірки та різноманіття значень у даних.

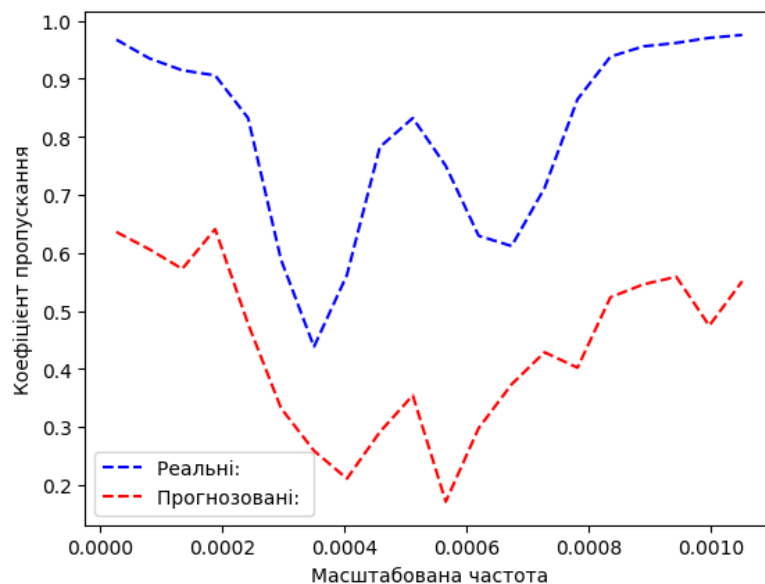
Для третьої вибірки даних з таблиці 4.4 були отримані такі результати:



а)

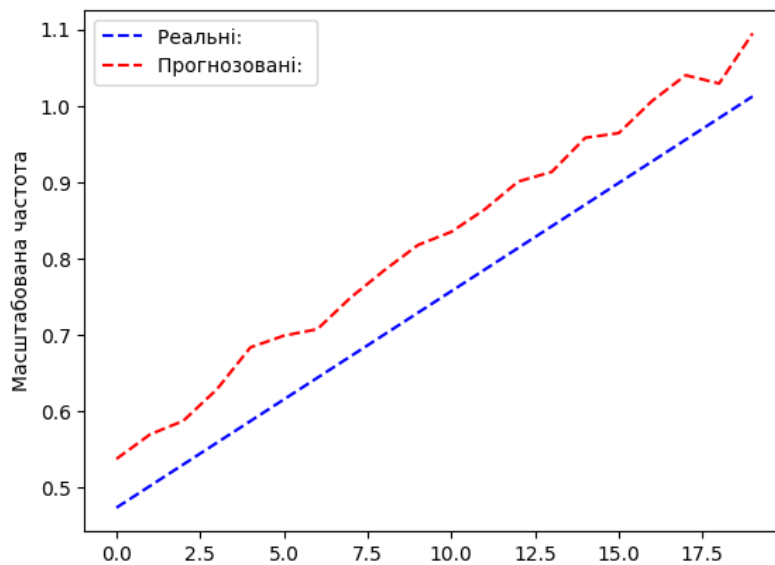


б)

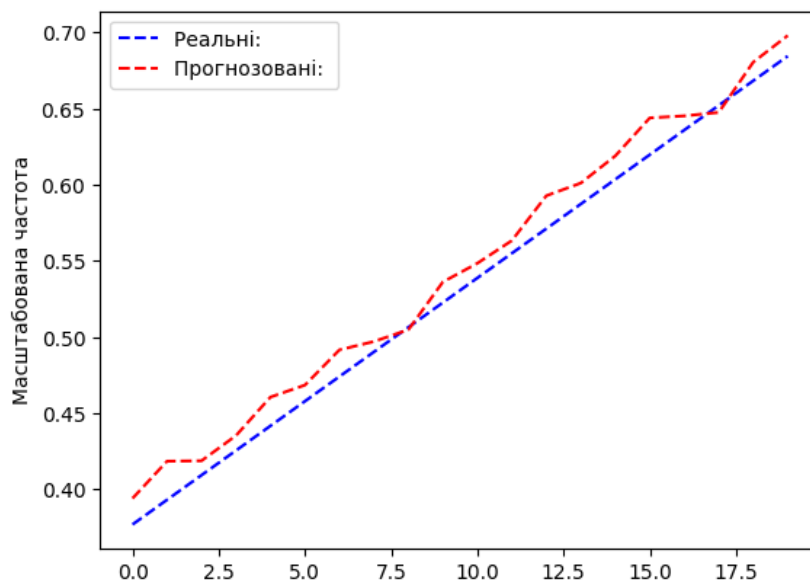


в)

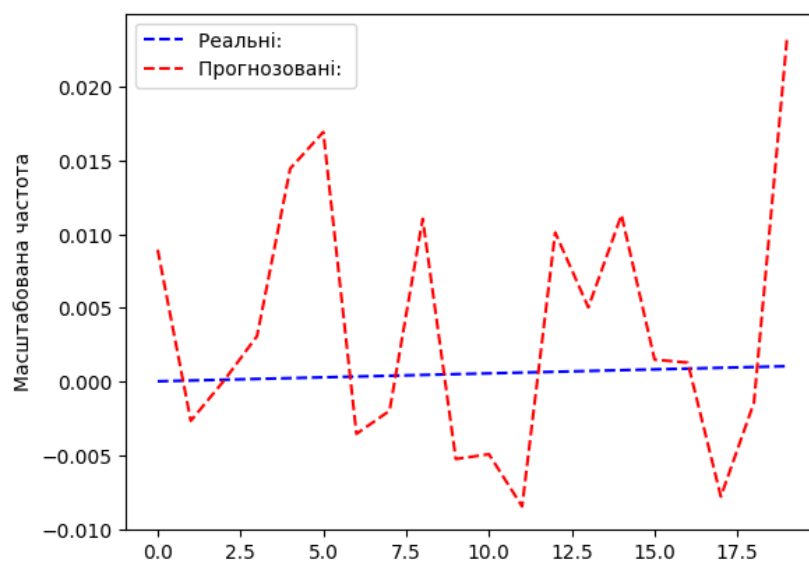
Рисунок 4.11 - Порівняння прогнозованого коефіцієнту пропускання та експериментально отриманого для структур з тестувальної вибірки а) figure3-8_cross-pol_90_2.0_2.npz, б) figure2-3_cross-pol_30_2.0_1.npz, в) figure1B-1_co-pol_90_2.0_1.npz



а)



б)



в)

Рисунок 4.12 - Порівняння прогнозованої частоти опромінення та експериментально встановленої для структур з тестувальної вибірки а) figure3-8_cross-pol_90_2.0_2.npz, б) figure2-3_cross-pol_30_2.0_1.npz, в) figure1B-1_co-pol_90_2.0_1.npz

Можна побачити, що змішування даних не призвело до покращення якості прогнозування коефіцієнту пропускання. Значення на графіку коефіцієнту пропускання лежать не близько до прямої реальних значень, а скоріш прогнозують тренд значень .



Рисунок 4.13 – Значення функції помилки під час навчання для тренувальної та тестувальної вибірок

Причиною таких результатів є нестабільне навчання, яке було викликане змішуванням даних з різною частотою опромінення.

4.3 Аналіз та шляхи покращення отриманих результатів

Архітектура штучної нейронної мережі дозволяє додавати “голови” з додатковими характеристиками для прогнозування. Такі голови були додані у показані на рис. 4.14.

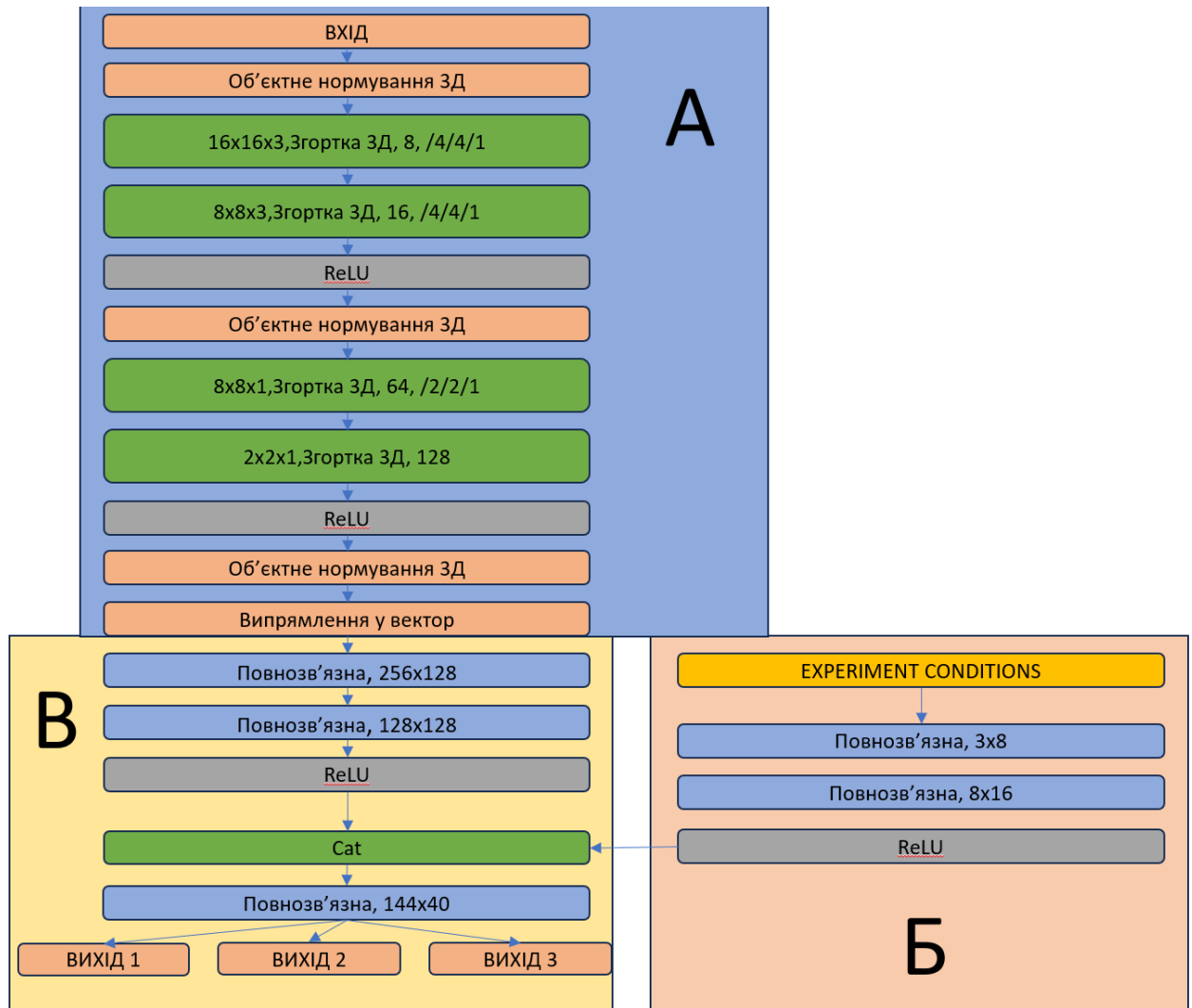


Рисунок 4.14 – Розширена архітектура для передбачення коефіцієнту пропускання, поглинання та відбивання

На рис. 4.14 зображена архітектура штучної нейронної мережі для передбачення коефіцієнтів пропускання, відбивання та поглинання. Такий тип архітектур легко модифікується під задачі, що потребують розв'язку. В даній роботі розглядався коефіцієнт пропускання із-за обмеженості в даних, що були зібрані в роботах [17][18][88-90].

Аналізуючи графіки на рис. 4.7, рис. 4.10, рис. 4.13, легко побачити, що друга підвбірка даних, частота випромінення яких лежить в межах одного діапазону, кількість даних велика, відносно першої вибірки, має більш кращий, з точки зору значень функції помилки, процес навчання. Це підтверджуються візуальним порівнянням реальних експериментальних значень коефіцієнта проникнення та прогнозованих штучною нейронною мережею.

Об'єднання даних з різним проміжком частоти випромінення для збільшення кількості даних не дало приросту по якості прогнозування, це видно з графіків рис. 4.8 б) та рис. 4.11 б), а навпаки погіршило якість прогнозування коефіцієнту проникнення.

За результатами першої вибірки, візуалізація прогнозування штучної нейронної мережі продемонстрована на рис. 4.5, можна стверджувати, що не дивлячись на однаковий проміжок частоти випромінення для експериментальних даних, результати прогнозування не такі коректні як для другої вибірки. Це пов'язано з кількістю даних, що були зібрані для цього проміжку.

Для коректного об'єднання вибірок, треба зробити такі ж шляхи закріплення інформації о частоті випромінення, як це було зроблено для кута опромінення, типу поляризації та показнику заломлення діелектрика. Розбиваючи зібрані дані для цього дослідження на підвбірки в залежності від частоти опромінення, ми таким чином закріплюємо інформацію о випроміненні методом “зоопарку моделей”, що був оглянутий в розділі 3.

Наступним шляхом покращення якості прогнозування був би метод закріплення даних про частоту випромінення методом контекстуального вектора, або атрибутивним алгоритмом. Теоретично так само можна задавати значення частоти випромінення методом контекстуальним методом, об'єднуючи вектор репрезентації з внутрішнім представленням всередині штучної нейронної мережі.

Іншим шляхом покращення результатів є більш ефективний збір даних разом з розміткою. Найбільше часу зайняв процес рендеру даних у програмному забезпеченні *Blender*. Так як процес рендеру відбувався ручним методом за описом структури метаматеріалу, можна розробити алгоритм автоматичного рендеру по заданій деяким форматом схемі. Ця схема може мати табличний формат, в якому будуть вказано кількість площин одного матеріалу, їх розміри, та проміжок у просторі, де вони починаються та закінчуються. Таким чином, існує можливість створення алгоритму автоматичного рендеру структур метаматеріалів одразу в *NumPy* масив мовою програмування *Python*, уникаючи проміжних форматів та ручного рендеру.

Інший спосіб покращити якість передбачення – збільшенні кількості даних. Основним блоком згорткової нейронної мережі для прогнозування коефіцієнта пропускання в залежності структури, складу та умов вимірювання властивостей метаматеріалів є згортковий блок представлення структурної інформації. Вже існують багато методів перенавчання штучних нейронних мереж, котрі в собі несуть інформацію. Одним із таких способів є Transfer Learning (навчання передачею)[115]. Цей метод можна описати такими шляхами:

1. Беремо модель навчену на більш глобальній задачі
2. Заморожуємо перші N шарів, які відповідають за можливість знаходження загальних залежностей
3. Тренуємо останні M шарів на більш вузькоспеціалізованій задачі

Такий підхід дає кращі результати прогнозування, класифікації, детекції, тощо[115]. Єдине обмеження цього підходу є те, що глобальна модель, яка потрібна для донавчання, повинна бути навчена на даних, які повинні бути надвибіркою цільових даних. Тобто, якщо є задача розпізнавання рідкісних видів котів, можна взяти модель, що була навчена на вибірці с тваринами, або на вибірці зі звичайними котами. Для задачі прогнозування властивостей електромагнітних метаматеріалів, можна

натренувати якусь глобальну модель на різних типах метаматеріалів. І вже її використовувати для донавчання.

Інша проблема, яка виникає при зборі даних – це відсутність “розмітки” для структур, тобто не завжди можна знайти потрібні експериментальні дослідження. Для розв'язку такої проблеми існує окремий клас процесів навчання як Self-supervised learning. З назви цього класу навчань, випливає те, що розмітка для даних отримується з самих даних.

Self-supervised learning - це підхід до навчання глибоких нейронних мереж, де модель намагається використовувати внутрішню структуру вхідних даних для створення власної умовної навчальної задачі. Замість того, щоб мати експертно позначені мітки для великої кількості даних, як це робиться у багатьох навчаннях з учителем, self-supervised learning використовує внутрішню кореляцію або залежності в даних для створення задач, які модель може вирішувати[116].

Один із загальних методів self-supervised learning - це прогнозування контексту (context prediction). В основі цього методу лежить ідея заснована на передбаченні частини даних залежно від інших частин. Наприклад, у текстовому контексті модель може навчатися передбачати наступне слово в реченні, засновуючись на попередніх словах. Також, відео можна навчати за допомогою передбачення наступного кадру з попередніми кадрами. Ці внутрішні залежності дозволяють моделі вивчати важливі репрезентації даних.

Загальні етапи self-supervised learning включають:

- створення завдання самонавчання - вибір завдання, яке модель буде намагатися вирішити, використовуючи вхідні дані. Це може бути передбачення частини даних на основі інших частин, реконструкція даних з шумом та інші;

- побудова моделі - створення архітектури нейронної мережі для вирішення створеного завдання самонавчання. Ця модель може бути багат шаровою нейронною мережею або іншою архітектурою, залежно від вхідних даних та завдання;
- навчання моделі - навчання моделі на вхідних даних, виконуючи підходящі операції для створеного завдання. Оскільки дані у самонавчанні є непозначеними, модель намагається здогадатися про залежності між різними частинами даних;
- фінтюнінг або використання репрезентацій - однією зі стратегій після навчання є використання отриманих репрезентацій для вирішення інших завдань. Наприклад, можна використовувати отримані репрезентації для класифікації або сегментації.

Загальний принцип self-supervised learning полягає у використанні природних залежностей в даних для внутрішнього самонавчання, що допомагає створити високоякісні репрезентації даних без потреби в ручних мітках.

Одним із відомих алгоритмів self-supervised learning є Wav2Vec. Wav2Vec є підходом до self-supervised learning, спрямованим на навчання корисних репрезентацій з аудіоданих. Цей підхід розроблений для обробки мовлення, де модель намагається вивчити високоякісні акустичні репрезентації з реальних аудіоданих без потреби в ручних мітках або перекладах[116-119].

Основною ідеєю Wav2Vec є використання задачі передбачення квантованого аудіосигналу. Аудіосигнал можна розділити на короткі фрагменти, які потім квантуються на кілька можливих значень. Модель намагається відновити квантовані значення аудіосигналу, використовуючи лише попередні значення. Таким чином,

Wav2Vec навчається розрізняти важливі акустичні особливості та патерни в аудіоданих, щоб вирішити цю передбачувальну задачу.

Архітектура Wav2Vec використовує згорткові шари для аналізу акустичних властивостей даних на різних рівнях деталей. Також вона включає пулінг та стохастичне пригнічення для створення більш досконалих репрезентацій аудіо сигналу. Після навчання на завданні передбачення квантованого сигналу, модель може бути використана для створення акустичних репрезентацій, які є корисними для різних задач обробки мовлення, таких як розпізнавання мови, синтез мовлення, класифікація аудіо тощо.

Основними перевагами підходу Wav2Vec є:

- відсутність потреби в ручних мітках - під час навчання використовуються лише аудіо дані без потреби в експертних анотаціях, що робить навчання більш доступним і ефективним;
- загальність репрезентацій - отримані репрезентації можуть бути використані для різних завдань обробки мовлення без перенавчання моделі;
- наявність високоінформативних репрезентацій - відповідно до досліджень, репрезентації, отримані від Wav2Vec, можуть містити значущі акустичні характеристики мовлення.

Загалом, Wav2Vec підтверджує важливість використання самонавчання для створення цінних репрезентацій даних і знаходить своє використання в області обробки мовлення.

Таким чином, існує можливість створення аналогічного методу генерації репрезентацій структур метаматеріалів.

Обмеження використання self-supervised підходів є значна кількість даних у датасеті. Тобто, перед тим, як використовувати такі підходи, треба провести глобальний збір інформації о структурах метаматеріалів і генерації їх структурного представлення. Наприклад, для wav2vec підхода було зібрано понад 10 тисяч часів записаної мови, в той же час, на той момент існували вибірки всього лише на сотні часів записаної мови. Тобто, якщо зберігати таку пропорцію, виходить, що для self-supervised алгоритму треба зібрати понад 600 структур метаматеріалів, що неможливо зробити ручним методом із-за часових обмежень.

Інший спосіб покращення якості прогнозування коефіцієнта пропускання є використання сучасних архітектур штучних нейронних мереж. Перспективним типом, для дослідження таких структурних даних як метаматеріали є трансформери. Основною властивістю яких є врахування глобальної структури даних, на відміну від згорткових шарів, контекстне вікно яких обмежено розміром згортки.

С самого початку, основною задачею під яку були спроектовані трансформери, це обробки мови, тексту, речень. Такий тип даних має лише часову залежність, тому класичні трансформери не підійдуть до задачі обробки структур метаматеріалів із-за 3D залежностей в їх структурі. В роботі [120] була запропонована реалізація трансформерів для тривимірних типів даних, що підійде для задачі прогнозування коефіцієнта пропускання. Обмеженням для використання таких трансформерів є їх алгоритмічна складність. Швидкість інференсу і навчання суттєво залежить від розмірності вхідних даних. Наприклад для задачі з лише часовою залежністю, складність трансформерів являється $O(n^2)$. Також суттєвою проблемою використання трансформерів є те, що для їх навчання потребується набагато більша вибірка даних, ніж зібрана для цього дослідження.

Тому, аналізуючи всі можливі шляхи покращення результатів в цьому дослідженні, можна сказати, що основними шляхами є розробка ефективного та автоматичного

алгоритму прямого рендеру візуального представлення структури та складу метаматеріалу напряму у *NumPy* масив для пришвидшення створення тренувальних вхідних даних. Другим шляхом є створення автоматичного, або напівавтоматичного збору експериментальних досліджень властивостей метаматеріалів.

4.4 Перспективи створення оберненого алгоритму для прогнозування структури, складу та умов вимірювання метаматеріалів в залежності від необхідних властивостей

Оберненою задачею до задачі прогнозування властивостей метаматеріалів за їх структурою є задача генерації структури та складу та умов дослідження за заданими властивостями.

Результати цього дослідження доводять те, що штучні нейронні мережі можуть та знаходять залежність між структурою, складом метаматеріалів умовами їх дослідження та їх властивостями. Це дає можливість припускати можливість оберненої задачі.

Історично, перші задачі по обробці зображень мали такий самий тип залежності, тобто структура-властивості, як наприклад задача розпізнавання образів на картинці.

Вже існує декілька підходів у вирішенні задач з типом залежності властивість-структура для зображень, аудіо, або текстового матеріалу.

Саме збільшення обчислювальної можливості, збільшення кількості даних дало можливість вирішення задач генерації картинок за їх описом. Наприклад, штучна нейронна мережа, для задачі розпізнавання образів на картинці в середньому матиме декілька десятків мільйонів параметрів, як наприклад класичний ResNet-34, який має 63.5 мільйона параметрів був натренований на даних ImageNet, який містить в собі 14

мільйонів картинок, для тренування якого можна використовувати одну відеокарта Nvidia V100[121]. В протипагу цьому, сучасна архітектура для генерації зображень Stable Diffusion першої версії має тільки один мільярд параметрів, що тренуються (без врахування мовної моделі), датасет для тренування якої був зібраний з двох мільярдів картинок, а для тренування нейронної мережі було використано 256 найсучасніших графічних прискорювачів A100[122]. Ціна тренування склала за словами дослідників 600 000\$. Деякі сучасні закриті архітектури мають понад два мільярда параметрів, а інформації о кількості графічних прискорювачів для тренування таких моделей не були опубліковані.

Іншою проблемою генеративних моделей, які генерують за залежністю “властивість-структура” є те, що вони до сих пір не є стабільними у якості генерування не дивлячись на обсяг тренувальних даних та розмірів моделей. Все ще проводяться дослідження з покращення стабільності генерування. Один із прикладів сучасної архітектури показан на рис. 4.15.

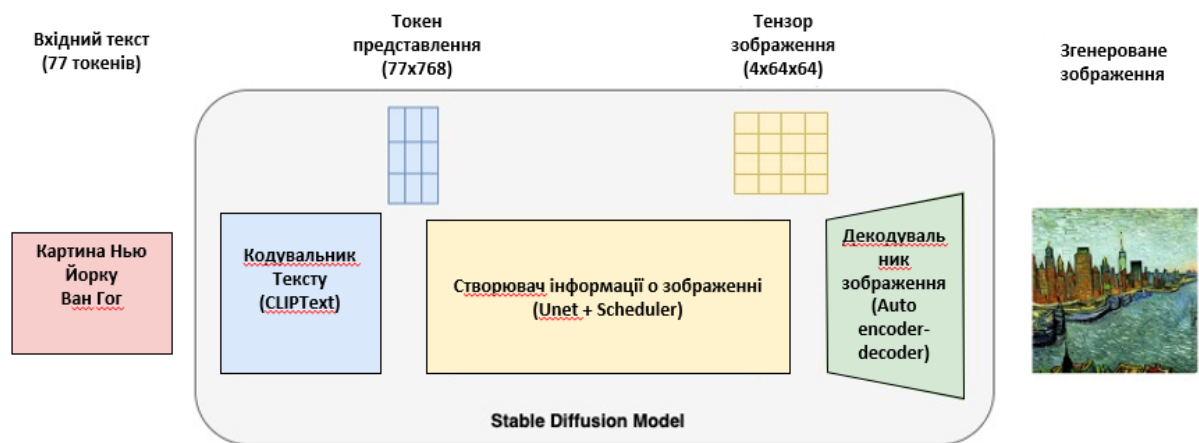


Рисунок 4.15 – схематичне спрощення представлення архітектури Stable Diffusion[122]

Аналізуючи схематичне зображення архітектури Stable diffusion з рис. 4.15, можна виділити перелік блоків, що потрібні для генерації зображення:

- текстовий кодувальник (голубий колір) - це нейронна мережа яка кодує текст представлений літерами у вектори, зі збереженням семантичної інформації;
- модель для об'єднання процесу дифузії та семантично закодованого тексту (жовтий колір) - ця модель потрібна для первинної генерації зображення яка використовує процес дифузії та контекстну інформацію з семантичного закодованого тексту з мережі;
- модель з знешумлення картинки що використовує семантично закодований текст (зелений колір) - ця модель потрібна для виправлення помилок генерації та прибирання випадково залишеного шуму після процесу дифузії, вона також використовує закодовану інформацію мережею.

Таким чином, можна запропонувати схему генерації структурної інформації за властивостями метаматеріалу, яка показана на рис. 4.16.

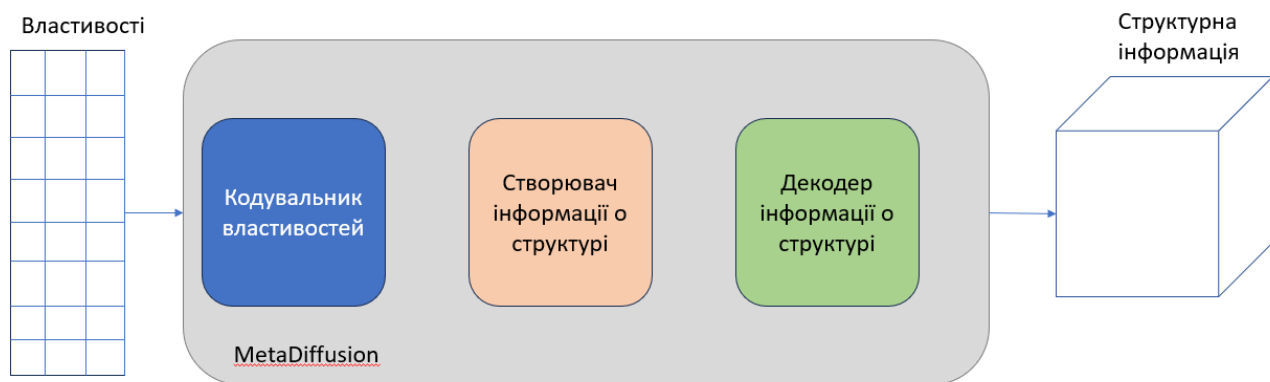


Рис. 4.16 – Схема блоків можливої штучної нейронної мережі для генерації структури метаматеріалу за заданими властивостями

Якщо для задачі прогнозування структури метаматеріалів запропонувати таку ж саму схему рис. 4.16, то окрім збору надзвичайної кількості даних в форматі “структура-властивості”, потрібний збір даних для кодування властивостей метаматеріалів, для тренування мережі, що здатна ефективно кодувати задані властивості та їх комбінаційні залежності.

3D структури мають додаткову розмірність, що допускає потребу у збільшенні кількості параметрів штучної нейронної системи, для якісної генерації структур за бажаними властивостями.

Ще, одним обмеженням до використання такого не стабільного підходу є те, що якість генерації штучної нейронної мережі за описом картинки можна перевіряти (тобто перевірити якість і точність генерації) на око. Перевірити співставлення структури метаматеріалу і його властивостями на око не вийде. Єдиним варіант первинної перевірки є використанням запропонований нейронній мережі для прогнозування властивостей за структурою, яка була натренована на окремій вибірці.

Тому, на мою думку, задача генерації структури метаматеріалу за його властивостями перспективна задача, яка має багато кроків для вирішення, перш ніж буде створюватися конвеєр для генерації. Більш доцільно пройти такий же шлях як проходили архітектури для класифікації зображення, а саме:

- поступове збільшення кількості даних у навчальній вибірці;
- поступове збільшення розмірів моделей, винаходження нових блоків штучної нейронної мережі;
- поступове покращення якості передбачення та заснування нових підзадач, нейронні мережі з яких можуть використовуватися без попереднього навчання для генерації структур метаматеріалів.

Таким чином однією з задач в цій роботі полягало у створенні підходу до збільшення кількості даних для навчання, що й вирішує перший пункт.

Висновки до розділу 4

Штучна нейронна мережа має можливість знаходження структурних залежностей в топологічному представленні інформації о складі метаматеріалу під час тренування.

Змішування даних з різним частотним діапазоном, для збільшення розмірів навчальної вибірки, призводить до збільшення середньоквадратичної помилки у 3 рази (0.024). Також, використання одного діапазону частот призводить до стабільного навчання, згідно графіку залежності значення функції помилки від номера епохи.

Використання додаткових умов експериментального дослідження збільшило кількість даних у навчальній вибірці до 30 екземлярів з 10.

Використання вибірки для навчання такого розміру призводить до ефекту перенавчання штучної нейронної мережі. Цей ефект може бути компенсований або збільшенням кількості даних для навчання, або використання суттєво інших підходів глибокого навчання. Як наприклад transfer learning із більших моделей, котрі були навчені методом self-supervised або unsupervised.

Подальший розвиток використання глибокого навчання для прогнозування властивостей метаматеріалів відкриє можливості до вирішення оберненої задачі. А саме, генерація структури метаматеріалів за заданими властивостями. Для вирішення цієї задачі потрібно збільшити кількість даних для навчання, та адаптувати існуючі архітектури для генерації зображень.

Збільшення кількості даних для навчання методом кодування параметрів опромінення поверхні метаматеріалу збільшує можливість розв'язку оберненої задачі.

ОСНОВНІ НАУКОВІ ТА ПРАКТИЧНІ РЕЗУЛЬТАТИ І ВИСНОВКИ

1) Розроблений метод кодування структурної інформації і інформації про склад метаматеріалу успішно кодує таку інформацію у чотиривимірну матрицю яка описує розташування кожного матеріалу у тривимірному просторі.

2) Доведено, що існує можливість визначати та вивчати структурні залежності метаматеріалів використовуючи операцію тривимірної згортки. Блоки такої згортки можуть бути базовими для наступних ітерацій навчання штучної нейронної мережі при збільшеній кількості даних. Таким чином існує можливість аналізувати метаматеріали з абсолютно різною просторовою структурою метаматеріалів.

3) Розроблений метод збільшення кількості даних шляхом кодування параметрів опромінення, збільшив кількість даних з 10-ти пар структура-властивість до 28 пар структура-властивість. Середньоквадратична похибка прогнозування зменшилася з 0.051 до 0.025. Введення додаткових умов також розширило межі використання такої штучної нейронної мережі, та додало гнучкість у процес підготовки, так як використання контекстного вектора для закріплення умов експериментальних досліджень не потребує навчання додаткових моделей під кожні умови

4) Розроблений метод збільшення кількості даних для навчання штучних нейронних мереж створює можливість розв'язання оберненої задачі проектування структури метаматеріалу за потрібними властивостями та параметрами опромінення такого метаматеріалу.

5) Було запроновано вирішення оберненої задачі за допомогою згорткової штучної нейронної мережі, а також використанні уніфікованих даних для навчання методом кодування параметрів опромінення у штучну нейронну мережу

СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ

- [1] Lopez, Maël & Dupont, Thomas & Panneton, R.. (2023). A mass-spring analogy for modeling the acoustic behaviour of a metamaterial. INTER-NOISE and NOISE-CON Congress and Conference Proceedings. 265. 6295-6301. 10.3397/IN_2022_0935.
- [2] Porfyrakis, Polykarpos & Tsitsas, Nikolaos. (2019). Nonlinear electromagnetic metamaterials: Aspects on mathematical modeling and physical phenomena. *Microelectronic Engineering*. 216. 111028. 10.1016/j.mee.2019.111028.
- [3] M. Maasch, “Tunable microwave metamaterial structures,” Ph.D. dissertation, Electron. Eng., Technische Universität Darmstadt, 2016. [Online]. Available: Proquest. Accessed on: 08.02.2016.
- [4] X. Cai, R. Zhu and G. Hu. “Experimental study for metamaterials based on dielectric resonators and wire frame,” *Metamaterials*, vol. 2, pp. 220-226. Dec. 2008. doi:10.1016/j.metmat.2008.08.001.
- [5] F. Demore, G. Rizzi, M. Collet, P. Neff and A. Madeo, “Unfolding engineering metamaterials design: Relaxed micromorphic modeling of large-scale acoustic meta-structures,” *Journal of the Mechanics and Physics of Solids*, vol. 168, no. 104995, Nov. 2022. doi:10.1016/j.jmps.2022.104995.
- [6] L. Yanglong and W. Yan, “Structural optimization of metamaterials based on periodic surface modeling,” *Computer Methods in Applied Mechanics and Engineering*, vol. 395, no. 115057, May 2022. doi:10.1016/j.cma.2022.115057.
- [7] W. Pabst, S. Hříbalová and T. Uhlířová. “Quasi-laminate and quasi-columnate modeling of dielectric and piezoelectric properties of cubic-cell metamaterials,” *Journal of the European Ceramic Society*, vol. 42, no. 3, pp. 1396-1406, Apr. 2022. doi:10.1016/j.jeurceramsoc.2021.12.013.

- [8] Z. Shaoyu, Z. Yingyan, Z. Yihe, Z. Wei, Y. Jie and K. Sritawat, "Genetic programming-assisted micromechanical models of graphene origami-enabled metal metamaterials," *Acta Materialia*, vol. 228, no. 117791, Apr. 2022. doi:10.1016/j.actamat.2022.117791.
- [9] P. Kumar, T. Ali and M. Pai, "Electromagnetic Metamaterials: A New Paradigm of Antenna Design," *IEEE Access*, vol. 9, pp. 18722-18751, Jan. 2021, doi:10.1109/ACCESS.2021.3053100.
- [10] Y. Jin, L. He, Z. Wen, B. Mortazavi, H. Guo, D. Torrent, B. Djafari-Rouhani, T. Rabczuk, X. Zhuang, and Y. Li, "Intelligent on-demand design of phononic metamaterials," *Nanophotonics*, vol. 11, no.3, pp. 439-460, Jan. 2022, doi:10.1515/nanoph-2021-0639.
- [11] R. Trivedi, N. Rai and K. Bharadwaj, "Metamaterial: Materials with Exceptional Properties," *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT). ETRASC*, vol. 2, no. 3, pp. 367-373, Jul. 2018, **doi:10.17577/IJERTCONV2IS03058.**
- [12] M. Bessa, P. Glowacki and M. Houlder, "Bayesian Machine Learning in Metamaterial Design: Fragile Becomes Supercompressible," *Advanced Materials*, vol. 31, no. 48, pp. 1904845, Oct. 2019, doi: 10.1002/adma.201904845.
- [13] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," *Science*, vol. 361, no. 6400, pp. 360-365, Jul. 2018, doi:10.1126/science.aat2663.
- [14] T. Xie and J. Grossman, Jeffrey, "Crystal Graph Convolutional Neural Networks for Accurate and Interpretable Prediction of Material Properties," *Physical Review Letters*, vol. 120, no. 14, pp. 145301, Apr. 2018, doi:10.1103/PhysRevLett.120.145301.
- [15] X. Zhou, Q. Xiao and H. Wang, "Metamaterials Design Method based on Deep Learning Database," in 2021 International Conference on Advanced Technologies and Applications of Modern Industry (ATAMI 2021), pp. 12

- [16] S. Lim, S. Young and R. Patton, “An analysis of image storage systems for scalable training of deep neural networks,” in *An analysis of image storage systems for scalable training of deep neural networks*.
- [17] L. Cong, Y. Srivastava and R. Singh, “Inter and intra-metamolecular interaction enabled broadband high-efficiency polarization control in metasurfaces,” *Applied Physics Letters*, vol. 108, no. 1, pp. 011110, Jan. 2016, doi:10.1063/1.4939564.
- [18] F. Ding, Z. Wang, S. He, V. Shalaev and A. Kildishev, “Broadband High-Efficiency Half-Wave Plate: A Super-Cell Based Plasmonic Metasurface Approach,” *ACS Nano*, vol. 9, no. 4, pp. 4111-4119, Mar. 2015, doi:10.1021/acsnano.5b00218.
- [19] Ma, Chunping & Chang, Yilong & Wu, Shuai & Zhao, Ruike. (2022). Deep Learning-Accelerated Designs of Tunable Magneto-Mechanical Metamaterials. *ACS Applied Materials & Interfaces*. 14. 10.1021/acsami.2c09052.
- [20] I. Sajedian, J. Kim and J. Rho, “Finding the optical properties of plasmonic structures by image processing using a combination of convolutional neural networks and recurrent neural networks,” *Microsyst Nanoeng*, vol. 5, no. 27, Jun. 2019, doi:10.1038/s41378-019-0069-y.
- [21] Khatib, Omar & Ren, Simiao & Malof, Jordan & Padilla, Willie. (2021). Deep Learning the Electromagnetic Properties of Metamaterials—A Comprehensive Review. *Advanced Functional Materials*. 31. 10.1002/adfm.202101748.
- [22] S. So and J. Rho, “Designing nanophotonic structures using conditional deep convolutional generative adversarial networks,” *Nanophotonics*, vol. 8, no. 7, Jun. 2019, pp. 1255-1261, [doi:10.1515/nanoph-2019-0117](https://doi.org/10.1515/nanoph-2019-0117).
- [23] Liu, Le & Xie, Long-Xiang & Huang, Weichun & Zhang, Xiu & Lu, Ming-Hui & Chen, Yan-Feng. (2022). Broadband acoustic absorbing metamaterial via deep learning approach. *Applied Physics Letters*. 120. 251701. 10.1063/5.0097696.

- [24] Zeng, Qingliang & Zhao, Zeang & Lei, Hongshuai & Wang, Panding. (2022). A Deep Learning Approach for Reverse Design of Gradient Mechanical Metamaterials. *International Journal of Mechanical Sciences*. 10.1016/j.ijmecsci.2022.107920.
- [25] W. Ma and Y. Liu, “A data-efficient self-supervised deep learning model for design and characterization of nanophotonic structures,” *Science China Physics, Mechanics and Astronomy*, vol. 63, no. 284212, Jun. 2020, doi:10.1007/s11433-020-1575-2.
- [26] A. Mall, A. Patil, D. Tamboli, A. Seth and A. Kumar, “Fast design of plasmonic metasurfaces enabled by deep learning,” *Journal of Physics D: Applied Physics*, vol. 53, no. 49, Aug. 2020, doi:[10.1088/1361-6463/abb33c](https://doi.org/10.1088/1361-6463/abb33c).
- [27] M. Tahersima, K. Kojima, T. Koike-Akino, D. Jha, B. Wang, C. Lin and K. Parsons, “Deep Neural Network Inverse Design of Integrated Photonic Power Splitters,” vol. 9, no. 1, pp. 1368, Feb. 2019, doi:10.1038/s41598-018-37952-2.
- [28] T. Qiu, X. Shi, J. Wang, Y. Li, S. Qu, Q. Cheng, T. Cui and S. Sui, “Deep Learning: A Rapid and Efficient Route to Automatic Metasurface Design,” *Advanced Science*, vol. 6, no. 12, pp. 1900128, Jun. 2019, doi:10.1002/advs.201900128.
- [29] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303-314, Dec. 1989, doi:10.1007/BF02551274.
- [30] C. Gurbuz, F. Kronowetter, C. Dietz, M. Eser, J. Schmid and S. Marburg, “Generative adversarial networks for the design of acoustic metamaterials,” *The Journal of the Acoustical Society of America*, vol. 149, pp. 1169-1174, Feb. 2021, doi:10.1121/10.0003501.
- [31] M. Zozyuk, D. Koroliouk, P. Krysenko, A. Yurikov and Y. Yakymenko, “Prediction of characteristics using a convolutional neural network based on experimental data on the structure and composition of metamaterials,” *STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING*, vol. 11, no. 3, pp. 777–787, Jun. 2023, doi:10.19139/soic-2310-5070-1707.

- [32] [3] P. I. Krysenko, M. O. Zoziuk "Using Information about Experimental Conditions to Predict Properties of Metamaterials," *Microsystems, Electronics and Acoustics*, vol. 28, no. 3, 2023, pp. 287808.1-287808.7, Dec. 2023. doi: 10.20535/2523-4455.me.287808.
- [33] Manry, Michael & Fitzner, Jack & Klemer, David & Perceptron, Multi-layer. (2000). Efficient Nonlinear Mapping Using The Multi-Layer Perceptron.
- [34] Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. (1998). "Gradient-based learning applied to document recognition" *Proceedings of the IEEE*. **86** (11): 2278–2324. doi:10.1109/5.726791. S2CID 14542261.
- [35] Ahlawat, Samit. (2022). Recurrent Neural Networks. 10.1007/978-1-4842-8835-1_4.
- [36] Dolgikh, Serge. (2023). From Data to Model: Evolutionary Learning with Generative Neural Systems. 10.1007/978-3-031-27524-1_71.
- [37] Laredo, David & Ma, Shangjie & Leylaz, Ghazaale & Schütze, Oliver & Sun, J. Q.. (2020). Automatic model selection for fully connected neural networks. *International Journal of Dynamics and Control*. 8. 1-17. 10.1007/s40435-020-00708-w.
- [38] Engheta, Nader; Richard W. Ziolkowski (June 2006). *Metamaterials: Physics and Engineering Explorations*. Wiley & Sons. pp. xv, 3–30, 37, 143–50, 215–34, 240–56. ISBN 978-0-471-76102-0.
- [39] Zharov, Alexander A.; Zharova, Nina A.; Noskov, Roman E.; Shadrivov, Ilya V.; Kivshar, Yuri S. (2005). "Birefringent left-handed metamaterials and perfect lenses for vectorial fields". *New Journal of Physics*. **7** (1): 220. arXiv:physics/0412128. Bibcode:2005NJPh....7..220Z. doi:10.1088/1367-2630/7/1/220.
- [40] Bowers J. A.; Hyde R. A. et al. "Evanescent electromagnetic wave conversion lenses I, II, III" US Patent and Trademark Office, Grant US-9081202-B2, 14 juli 2015, U.S. Patent 9,081,202

- [41] Yu, Peng; Wu, Jiang; Ashalley, Eric; Govorov, Alexander; Wang, Zhiming (2016). "Dual-band absorber for multispectral plasmon-enhanced infrared photodetection" (PDF). *Journal of Physics D: Applied Physics*. 49 (36): 365101. Bibcode:2016JPhD...49J5101Y. doi:10.1088/0022-3727/49/36/365101. ISSN 0022-3727. S2CID 123927835.
- [42] Liang Ge, Jining Li, Kai Zhong, Yuye Wang, Degang Xu, Jianquan Yao. "A thermocouple based on wideband hybrid metamaterial absorber for mid-infrared photo-thermoelectric detector", *Optics Communications*, Volume 545, 2023, 129714, ISSN 0030-4018,
- [43] Wu, B.-I.; W. Wang, J. Pacheco, X. Chen, T. Grzegorzczuk and J. A. Kong; Pacheco, Joe; Chen, Xudong; Grzegorzczuk, Tomasz M.; Kong, Jin Au (2005). "A Study of Using Metamaterials as Antenna Substrate to Enhance Gain". *Progress in Electromagnetics Research*. 51: 295–28. doi:10.2528/PIER04070701.
- [44] Wei Wang, Wenhao Wang, Fengping Yan, Zhi Wang, Haisu Li, Siyu Tan, Xuemei Du, Numerical research on supercavity sensing driven by guided resonance in a terahertz all-metal metamaterial absorber, *Optics & Laser Technology*, Volume 169, 2024, 110113, ISSN 0030-3992,
- [45] Moon, Gwiyeong & Choi, Jong-Ryul & Lee, Changhun & Oh, Youngjin & Kim, Kyunghwan & Kim, Donghyun. (2020). Machine learning-based design of meta-plasmonic biosensors with negative index metamaterials. *Biosensors and Bioelectronics*. 164. 112335. 10.1016/j.bios.2020.112335.
- [46] Gao, Huixuan & Liang, Yuzhang & Yu, Li & Shuwen, Chu & Cai, Lina & Wang, Fei & Wang, Qiao & Peng, Wei. (2021). Bifunctional plasmonic metamaterial absorber for narrowband sensing detection and broadband optical absorption. *Optics & Laser Technology*. 137. 106807. 10.1016/j.optlastec.2020.106807.

- [47] Monfared, Yashar & Qasymeh, Montasir. (2021). Graphene-assisted infrared plasmonic metamaterial absorber for gas detection. *Results in Physics*. 23. 103986. 10.1016/j.rinp.2021.103986.
- [48] Dan, Yihang & Zhang, Tian & Sun, Xiaojuan & Dai, Jian & Xu, Kun. (2022). Multifunctional plasmonic waveguide system based on coding metamaterials and inverse design. *Optics & Laser Technology*. 156. 108478. 10.1016/j.optlastec.2022.108478.
- [49] Feng, Mingming & Zhang, Xijian & Zhang, Baoqing & Li, Ke & Li, Zhaolin & Wang, Yiming & Wang, Qingpu & Guo, Qinglei & Zhang, Jiawei & Qin, Weidong & Song, A. & Zhang, Yifei. (2023). Polarization-selective modulation of meandered-line metamaterials with graphene surface plasmonics. *Optics and Lasers in Engineering*. 169. 107693. 10.1016/j.optlaseng.2023.107693.
- [50] Li, Jitao & Yue, Zhen & Li, Jie & Zheng, Chenglong & Liu, Jingyu & Yang, Fan & Li, Hui & Zhang, Yating & Zhang, Yan & Yao, J.. (2022). Wavefront-controllable all-silicon terahertz meta-polarizer. *Science China Materials*. 10.1007/s40843-022-2126-0.
- [51] Cao, Shuhua & Wang, Qi & Gao, Xufeng & Zhang, Shijie & Hong, Ruijing & Zhang, Dawei. (2021). Electrically-modulated infrared absorption of graphene metamaterials via magnetic dipole resonance. *Physica E: Low-dimensional Systems and Nanostructures*. 137. 115078. 10.1016/j.physe.2021.115078.
- [52] Long, Linshuang & Taylor, Sydney & Ying, Xiaoyan & Wang, Liping. (2019). Thermally-switchable spectrally-selective infrared metamaterial absorber/emitter by tuning magnetic polariton with a phase-change VO₂ layer. *Materials Today Energy*. 13. 214-220. 10.1016/j.mtener.2019.05.017.
- [53] Rinkevich, A. & Demokritov, So & Samoilovich, M.I. & Nemytova, O.V.. (2015). Interaction of microwave electromagnetic waves with 3D magnetic metamaterials. *Photonics and Nanostructures - Fundamentals and Applications*. 15. 10.1016/j.photonics.2015.03.004.

- [54] Lei, Zhenkuang & Tan, Guoguo & Man, Qikui & Ning, Mingqiang & Wang, Shengding & Chen, Shuwen & Yang, Zhen & Li, Run-Wei. (2020). A Flexible Metamaterial Based on Liquid Metal Patterns Embedded in Magnetic Medium for Lightweight Microwave Absorber. *Materials Research Bulletin*. 137. 111199. 10.1016/j.materresbull.2020.111199.
- [55] Galea, Russell & Dudek, Krzysztof & Farrugia, Pierre-Sandre & Mangion, Louis & Grima, Joseph & Gatt, Ruben. (2021). Reconfigurable magneto-mechanical metamaterials guided by magnetic fields. *Composite Structures*. 280. 114921. 10.1016/j.compstruct.2021.114921.
- [56] Peng, Yuping & Wang, Qi & Xu, Yunchao & Shan, Dongyong & He, Longhui & Cao, Yanming. (2023). Optically transparent and mechanically stretchable fractal-structured wave-absorbing metamaterial in low frequency range. *Journal of Alloys and Compounds*. 961. 171100. 10.1016/j.jallcom.2023.171100.
- [57] Zhu, Ruxian & Xu, Wei & Chen, Leyi & Shaolong, Tang. (2023). Boosting multi-mode strong coupling in a hybrid vertical split-ring resonator metamaterial for magnetoplasmonic sensing. *Optics Communications*. 549. 129862. 10.1016/j.optcom.2023.129862.
- [58] O. Khatib, S. Ren, J. Malof and W. Padilla, “Deep Learning the Electromagnetic Properties of Metamaterials—A Comprehensive Review,” *Advanced Functional Materials*, vol. 31, no. 31, pp. 2101748, Aug. 2021, doi:10.1002/adfm.202101748.
- [59] A. Hammond and R. Camacho, “Designing integrated photonic devices using artificial neural networks,” *Optics Express*, vol. 27, no. 21, pp. 29620-29638, Sep. 2019, doi:10.1364/OE.27.029620.
- [60] Y. Xu, X. Zhang, Y. Fu and Y. Liu, “Interfacing photonics with artificial intelligence: an innovative design strategy for photonic structures and devices based on artificial neural networks,” *Photonics Research*, vol. 9, no. 4, pp. 135-152, Mar. 2021, doi:10.1364/PRJ.417693.

- [61] Kim, Jong & Messerschmidt, Marc & Graves, William. (2022). Enhancement of Partially Coherent Diffractive Images Using Generative Adversarial Network. *AI*. 3. 274-284. 10.3390/ai3020017.
- [62] An, Sensong & Fowler, Clayton & Zheng, Bowen & Shalaginov, Mikhail & Tang, Hong & Li, Hang & Zhou, Li & Ding, Jun & Agarwal, Anuradha & Baleine, Clara & Richardson, Kathleen & Gu, Tian & Hu, Juejun & Zhang, Hualiang. (2019). A Deep Learning Approach for Objective-Driven All-Dielectric Metasurface Design. *ACS Photonics*. XXXX. 10.1021/acsp Photonics.9b00966.
- [63] W. Kong, J. Chen, Z. Huang and D. Kuang, “Bidirectional cascaded deep neural networks with a pretrained autoencoder for dielectric metasurfaces,” *Photonics Research*, vol. 9, no. 8, pp. 1607-1615, Jul. 2021, doi:10.1364/PRJ.428425.
- [64] Peurifoy, John & Shen, Yichen & Jing, Li & Yang, Yi & Cano-Renteria, Fidel & Delacy, Brendan & Joannopoulos, John & Tegmark, Max & Soljačić, Marin. (2018). Nanophotonic Particle Simulation and Inverse Design Using Artificial Neural Networks. *Science Advances*. 4. eaar4206. 10.1126/sciadv.aar4206.
- [65] Shi, Ruiyang & Huang, Jie & Li, Shulun & Niu, Lingfeng & Yang, Junbo. (2022). Forward Prediction and Inverse Design of Nanophotonic Devices Based on Capsule Network. *IEEE Photonics Journal*. 14. 1-7. 10.1109/JPHOT.2022.3182050.
- [66] Chen, Yingshi & Zhu, Jinfeng & Xie, Yinong & Feng, Naixing & Liu, Qing. (2020). Smart inverse design of graphene-based photonic metamaterials by an adaptive artificial neural network Supplementary Information.pdf.
- [67] I. Sajedian, J. Kim and J. Rho, “Finding the optical properties of plasmonic structures by image processing using a combination of convolutional neural networks and recurrent neural networks,” *Microsyst Nanoeng*, vol. 5, no. 27, Jun. 2019, doi:10.1038/s41378-019-0069-y.

- [68] Yeung, Christopher & Pham, Benjamin & Zhang, Zihan & Fountaine, Katherine & Raman, Aaswath. (2022). Hybrid Supervised and Reinforcement Learning for the Design and Optimization of Nanophotonic Structures.
- [69] Ma, Wei & Cheng, Feng & Xu, Yihao & Wen, Qinlong & Liu, Yongmin. (2019). Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy. *Advanced Materials*. 31. 1901111. 10.1002/adma.201901111.
- [70] Liao, Xianglai & Gui, Lili & Zhenming, Yu & Zhang, Tian & Xu, Kun. (2022). Deep Learning for design of 3D ChiralPlasmonic Metasurfaces. *Optical Materials Express*. 12. 10.1364/OME.449152.
- [71] Liao, Xianglai & Gui, Lili & Zhenming, Yu & Zhang, Tian & Xu, Kun. (2022). Deep Learning for design of 3D ChiralPlasmonic Metasurfaces. *Optical Materials Express*. 12. 10.1364/OME.449152.
- [72] Soltani, Mahdi & Ghorbani, Fardin & Beyraghi, Sina & Azizi, Yousef & Soleimani, M.. (2022). Multifunctional Automatic Metasurface Design Using Deep Learning Approach. 10.36227/techrxiv.21507918.v1.
- [73] Dai, Minyi & Demirel, Mehmet & Liang, Yingyu & Hu, Jiamian. (2021). Graph neural networks for an accurate and interpretable prediction of the properties of polycrystalline materials. *npj Computational Materials*. 7. 103. 10.1038/s41524-021-00574-w.
- [74] Sedighi, Hamid M & Lallart, Mickael & Zhuang, Xiaoying. (2022). Metamaterials: Design, modelling, simulation and implementation. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*. 236. 2153-2156. 10.1177/14644207221137433.

- [75] Dorn, Charles & Kochmann, Dennis. (2022). Ray theory for elastic wave propagation in graded metamaterials. *Journal of the Mechanics and Physics of Solids*. 168. 105049. 10.1016/j.jmps.2022.105049.
- [76] Padilla, Willie & Averitt, Richard. (2017). Properties of dynamical electromagnetic metamaterials. *Journal of Optics*. 19. 10.1088/2040-8986/aa728d.
- [77] Nair, Raveendranath & Dutta, Maumita & P.S., Mohammed & Venu, Kola. (2018). EM Material Characterization Techniques for Metamaterials. 10.1007/978-981-10-6517-0.
- [78] X. Yu, J. Zhou, H. Liang, Z. Jiang and L. Wu, “Mechanical metamaterials associated with stiffness, rigidity and compressibility: A brief review,” *Progress in Materials Science*, vol. 94, pp. 114-173, May. 2018. doi:10.1016/j.pmatsci.2017.12.003.
- [79] Goyal, Archana & Gupta, Vishal & Kumar, Manish. (2023). A deep neural framework for named entity recognition with boosted word embeddings. *Multimedia Tools and Applications*. 10.1007/s11042-023-16176-1.
- [80] Xu, He & Zheng, Qunli & Zhu, Jingshu & Xie, Zuoling & Cheng, Haitao & Li, Peng & Ji, Yimu. (2022). A Deep Learning Model Incorporating Knowledge Representation Vectors and Its Application in Diabetes Prediction. *Disease Markers*. 2022. 1-17. 10.1155/2022/7593750.
- [81] Yang, Jie & Xu, Jinfan & Zhang, Xiaolei & Wu, Chiyu & Lin, Tao & Ying, Yibin. (2019). Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Analytica Chimica Acta*. 1081. 10.1016/j.aca.2019.06.012.
- [82] Imanian, Hanifeh & Shirkhani, Hamidreza & Mohammadian, Abdolmajid & Hiedra Cobo, Juan & Payeur, Pierre. (2023). Spatial Interpolation of Soil Temperature and Water Content in the Land-Water Interface Using Artificial Intelligence. *Water*. 15. 473. 10.3390/w15030473.

- [83] Rodríguez, Pau & Bautista, Miguel & González, Jordi & Escalera, Sergio. (2018). Beyond One-hot Encoding: lower dimensional target embedding. *Image and Vision Computing*. 75. 10.1016/j.imavis.2018.04.004.
- [84] Farhan, Muhammad & Dwifabri Purbolaksono, Mahendra & Astuti, Widi. (2023). Sentiment Analysis of Practo App Reviews using KNN and Word2Vec. *Building of Informatics, Technology and Science (BITS)*. 5. 10.47065/bits.v5i1.3598.
- [85] Huisman, Mike & Moerland, Thomas & Plaat, Aske & van Rijn, Jan. (2023). Are LSTMs good few-shot learners?. *Machine Learning*. 10.1007/s10994-023-06394-x.
- [86] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010.
- [87] Myklebust, Erik & Jiménez-Ruiz, Ernesto & Chen, Jiaoyan & Wolf, Raoul & Tollefsen, Knut Erik. (2022). Prediction of adverse biological effects of chemicals using knowledge graph embeddings. *Semantic Web*. 13. 1-40. 10.3233/SW-222804.
- [88] C. Huang, “Efficient and broadband polarization conversion with the coupled metasurfaces,” *Optics Express*, vol. 23, no. 25, pp. 32015-32024, Dec. 2015, doi:10.1364/OE.23.032015.
- [89] A. Shaltout, J. Liu, A. Kildishev and V. Shalaev, “Photonic spin Hall effect in gap-plasmon metasurfaces for on-chip chiroptical spectroscopy,” *Optica*, vol. 2, no. 10, pp. 860-863, Sep. 2015, doi:10.1364/OPTICA.2.000860.
- [90] N. Grady, J. Heyes, C. Roy, Z. Yong, R. Matthew, A. Abul, T. Antoinette and H. Chen, “Terahertz Metamaterials for Linear Polarization Conversion and Anomalous Refraction,” *Science*, vol. 340, no. 6138, pp. 1304-1307, May 2013, doi:10.1126/science.1235399.

- [91] H. W. Herwanto, A. N. Handayani, A. P. Wibawa, K. L. Chandrika and K. Arai, "Comparison of Min-Max, Z-Score and Decimal Scaling Normalization for Zoning Feature Extraction on Javanese Character Recognition," 2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Malang, Indonesia, 2021, pp. 1-3, doi: 10.1109/ICEEIE52663.2021.9616665.
- [92] Salazar, Jose & Garland, Lean & Ochoa, Jesus & Pyrcz, Michael. (2021). Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. *Journal of Petroleum Science and Engineering*. 209. 109885. 10.1016/j.petrol.2021.109885.
- [93] Rastegari, Mohammad & Ordonez, Vicente & Redmon, Joseph & Farhadi, Ali. (2016). XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks. 9908. 525-542. 10.1007/978-3-319-46493-0_32.
- [94] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*. 25. 10.1145/3065386.
- [95] Satapathy, Ashutosh & Livingston, Jenila. (2023). A lightweight convolutional neural network built on inception-residual and reduction modules for deep facial recognition in realistic conditions. *The Imaging Science Journal*. 71. 1-19. 10.1080/13682199.2023.2176735.
- [96] Xu, Fei & Wang, Peng & Xu, Huimin. (2023). Deep pyramidal residual networks with inception sub-structure in image classification. *Journal of Intelligent & Fuzzy Systems*. 45. 1-22. 10.3233/JIFS-230569.
- [97] Hütten, Nils & Meyes, Richard & Meisen, Tobias. (2022). Vision Transformer in Industrial Visual Inspection. *Applied Sciences*. 12. 11981. 10.3390/app122311981.

- [98] Papoutsis, Ioannis & Bountos, Nikolaos & Zavras, Angelos & Michail, Dimitrios & Tryfonopoulos, Christos. (2023). Benchmarking and scaling of deep learning models for land cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*. 195. 250-268. 10.1016/j.isprsjprs.2022.11.012.
- [99] Lee, Sang-Hwa. (2023). Text Classification of Mixed Model Based on Deep Learning. *Tehnički glasnik*. 17. 367-374. 10.31803/tg-20221228180808.
- [100] Seabe, Phumudzo & Moutsinga, Claude & Pindza, Edson. (2023). Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach. *Fractal and Fractional*. 7. 203. 10.3390/fractalfract7020203.
- [101] Wang, Yunfeng & Zhou, Wengang & Zhang, Qilin & Li, Houqiang. (2018). Enhanced Action Recognition With Visual Attribute-Augmented 3D Convolutional Neural Network. 10.1109/ICMEW.2018.8551536.
- [102] Peng, Dandan & Liu, Chenyu & Desmet, Wim & Gryllias, Konstantinos. (2022). Condition Monitoring of Wind Turbines Based on Anomaly Detection Using Deep Support Vector Data Description. 10.1115/GT2022-82624.
- [103] Rasulzade, Jalal & Rustamov, Samir & Akhmetov, B. & Maksum, Yelaman & Nogaibayeva, Makpal. (2022). Computational Acceleration of Topology Optimization Using Deep Learning. *Applied Sciences*. 13. 479. 10.3390/app13010479.
- [104] Shikalgar, Sajeeda & Yadav, Rakesh & Mahalle, Parikshit. (2023). Lightweight MobileNet Model for Image Tempering Detection. *International Journal on Recent and Innovation Trends in Computing and Communication*. 11. 55-69. 10.17762/ijritec.v11i5.6524.
- [105] Listio, Syntia. (2022). Performance of Deep Learning Inception Model and MobileNet Model on Gender Prediction Through Eye Image. *Sinkron*. 7. 2593-2601. 10.33395/sinkron.v7i4.11887.

- [106] Jiang, Lingzi & Zou, Beiji & Liu, Shu & Yang, Wenjun & Wang, Min & Huang, Enquan. (2022). Recognition of abnormal human behavior in dual-channel convolutional 3D construction site based on deep learning. *Neural Computing and Applications*. 35. 1-13. 10.1007/s00521-022-07881-3.
- [107] Ioffe, S. & Szegedy, C.. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proceedings of the 32nd International Conference on Machine Learning*, in *Proceedings of Machine Learning Research* 37:448-456 Available from <https://proceedings.mlr.press/v37/ioffe15.html>.
- [108] Chen, Liangyu & Lu, Xin & Zhang, Jie & Chu, Xiaojie & Chen, Chengpeng. (2021). HINet: Half Instance Normalization Network for Image Restoration. 182-192. 10.1109/CVPRW53098.2021.00027.
- [109] Wu, Yuxin & He, Kaiming. (2020). Group Normalization. *International Journal of Computer Vision*. 128. 10.1007/s11263-019-01198-w.
- [110] Rodrawangpai, Ben & Daungjaiboon, Witawat. (2022). Improving text classification with transformers and layer normalization. *Machine Learning with Applications*. 10. 100403. 10.1016/j.mlwa.2022.100403.
- [111] Mohamed, Aya & Hemeida, Ashraf & Hassan, Moatamad. (2022). Image classification based deep learning: A Review. *Aswan University Journal of Sciences and Technology*. 2. 10.21608/aujst.2022.259887.
- [112] Salehin, Imrus & Kang, Dae-Ki. (2023). A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain. *Electronics*. 12. 3106. 10.3390/electronics12143106.

- [113] Muhammad, Afif & Utomo, Hamzah & Hidayatullah, Priyanto & Syakrani, N.. (2022). Early Stopping Effectiveness for YOLOv4. *Journal of Information Systems Engineering and Business Intelligence*. 8. 11-20. 10.20473/jisebi.8.1.11-20.
- [114] Zhao, Hui & An, Jing & yu, mengjie & Lv, Diankai & kuang, kaida & zhang, tianqi. (2021). Nesterov accelerated adaptive momentumestimation-based wavefront distortioncorrection algorithm. *Applied Optics*. 60. 10.1364/AO.428465.
- [115] Bharadiya, Jasmin. (2023). Transfer Learning in Natural Language Processing (NLP). *European Journal of Technology*. 7. 10.47672/ejt.1490.
- [116] Akhare, Rakhi. (2023). Hybrid approach between Supervised and Unsupervised Learning: Self Supervised Learning. *Journal of Xidian University*. 17. 317-324. 10.37896/jxu17.8/026.
- [117] Hussain, Shehzeen & Nguyen, Van & Zhang, Shuhua & Visser, Erik. (2022). Multi-Task Voice Activated Framework Using Self-Supervised Learning. 10.1109/ICASSP43922.2022.9746409.
- [118] Fatehi, Kavan & Torres, Mercedes & Kucukyilmaz, Ayse. (2022). ScoutWav: Two-Step Fine-Tuning on Self-Supervised Automatic Speech Recognition for Low-Resource Environments. 3523-3527. 10.21437/Interspeech.2022-10270.
- [119] Guo, Dongyue & Zhang, Zichen & Yang, Bo & Zhang, Jianwei & Lin, Yi. (2023). Boosting Low-Resource Speech Recognition in Air Traffic Communication via Pretrained Feature Aggregation and Multi-Task Learning. *IEEE Transactions on Circuits and Systems II: Express Briefs*. PP. 1-1. 10.1109/TCSII.2023.3269051.
- [120] Lahoud, Jean & Cao, Jiale & Khan, Fahad & Cholakkal, Hisham & Anwer, Rao & Khan, Salman & Yang, Ming-Hsuan. (2022). 3D Vision with Transformers: A Survey. 10.48550/arXiv.2208.04309.

- [121] Gao, Mingyu & Chen, Jianfeng & Mu, Hongbo & Qi, Dawei. (2021). A Transfer Residual Neural Network Based on ResNet-34 for Detection of Wood Knot Defects. *Forests*. 12. 212. 10.3390/f12020212.
- [122] Rombach, Robin & Blattmann, Andreas & Lorenz, Dominik & Esser, Patrick & Ommer, Bjorn. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. 10674-10685. 10.1109/CVPR52688.2022.01042.

ДОДАТКИ

ДОДАТОК А. Список публікацій здобувача за темою дисертації:

Праці, в яких опубліковано основні результати дисертації:

[1] M. O. Zozyuk, D. V. Koroliouk, P. I. Krysenko, A. I. Yurikov and Y. I. Yakymenko, “Prediction of characteristics using a convolutional neural network based on experimental data on the structure and composition of metamaterials,” *STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING*, vol. 11, no. 3, pp. 777–787, Jun. 2023, doi:10.19139/soic-2310-5070-1707. *(Стаття у закордонному виданні, проіндексованому у базі даних Web of Science Core Collection та Scopus та віднесеному до 3 квартилю (Q3) відповідно до класифікації SCImago Journal and Country Rank).*

Особистий внесок здобувача: здобувачем відредактовано текст статті, розроблено та втілено алгоритм прогнозування коефіцієнту пропускання метаматеріалів на основі структури та складу метаматеріалів, проведено аналіз результатів прогнозування коефіцієнту пропускання метаматеріалів на основі структури та їх складу та проведено обговорення результатів із співавторами

[2] P. I. Krysenko, M. O. Zoziuk, O. I. Yurikov, D. V. Koroliuk and Yu. I. Yakymenko, “Chladni Figures Simulation on a Rectangular Plate,” *Microsystems, Electronics and Acoustics*, vol. 26, no. 1, 2021, pp. 241698.1-241698.6, Dec. 2021. doi:10.20535/2523-4455.me.241698. *(Стаття у періодичному науковому фаховому виданні категорії Б).*

Особистий внесок здобувача: здобувачем відредактовано текст статті, підготовлено математичну модель для моделювання фігур хладні, підготовлений код для моделювання фігур Хладні та графіків залежності коефіцієнту зміщення від частоти для хвиль.

[3] P. I. Krysenko, M. O. Zoziuk “Using Information about Experimental Conditions to Predict Properties of Metamaterials,” *Microsystems, Electronics and Acoustics*, vol. 28, no. 3, 2023, pp. 287808.1-287808.7, Dec. 2023. doi: 10.20535/2523-4455.me.287808. **(Стаття у періодичному науковому фаховому виданні категорії Б).**

Особистий внесок здобувача: здобувачем підготовлений текст статті, запропонований та розроблений метод закріплення інформації о умовах експериментальних досліджень, проведено навчання штучної нейронної мережі, були отримані та проаналізовані результати

Праці, які засвідчують апробацію матеріалів дисертації:

[1] М.О. Зозюк, Д.В. Королюк, П.І. Крисенко, О.І. Юріков та Ю.І. Якименко, “Використання нейронних мереж для прогнозування властивостей метаматеріалів,” представлено на *IX Міжнародній науковій конференції імені І. І. Ляшка «Обчислювальна та прикладна математика»*, Київ, Україна, 2022.

Особистий внесок здобувача: здобувачем відредактовано текст виступу, проведено моделювання характеристик метаматеріалів на основі їх структури, проведено аналіз результатів моделювання та проведено обговорення отриманих результатів із співавторами.