

Харківський національний університет імені В. Н. Каразіна
Факультет математики і інформатики
Кафедра прикладної математики

Кваліфікаційна робота

магістра

на тему

«Геометрична звивистість випадкових пористих структур»

Виконав: студент групи МП61, спеціальність 113 — Прикладна математика, освітньо-професійна програма «Прикладна математика»,
Татаренко В.В.

Науковий керівник: кандидат фіз.-мат. наук, доцент кафедри прикладної математики,
Пославський С.О.

Рецензент: доктор технічних наук, професор, професор кафедри комп'ютерно-інтегрованих технологій, автоматизації та робототехніки Харківського національного університету радіоелектроніки,
Ромашов Ю.В.

Анотації

Татаренко Віталій Віталійович. Геометрична звивистість випадкових пористих структур. У роботі досліджено геометричну звивистість двовимірних випадкових пористих середовищ, сформованих множиною прямокутних включень. Запропоновано стохастичну цифрову модель, що дозволяє варіювати морфологічні властивості структури та контролювати пористість. Для кожної реалізації обчислюються морфологічні інваріанти (функціонали Мінковського), топологічні характеристики та траєкторна метрика **zigzag**. Звивистість визначається за допомогою алгоритмів пошуку найкоротшого шляху (алгоритм Дейкстри / алгоритм A^*). На основі сформованого датасету побудовано моделі машинного навчання для прогнозування звивистості. Показано, що локальні особливості каналу та характеристика **zigzag** є ключовими чинниками, що визначають величину звивистості.

Ключові слова: пористе середовище, звивистість, функціонали Мінковського, характеристика Ейлера, машинне навчання, найкоротший шлях.

Tatarenko Vitalii. Geometric tortuosity of random porous structures. The thesis investigates the geometric tortuosity of two-dimensional random porous media composed of multiple rectangular solid inclusions. A stochastic digital model is developed to control morphological variability and porosity. For each realization, Minkowski morphological functionals, Euler characteristic, and the zigzag trajectory metric are computed. Tortuosity is evaluated using shortest-path algorithms (Dijkstra / A^*). Machine-learning regression models are trained to predict tortuosity from geometric descriptors. The study demonstrates that local channel features and the zigzag characteristic are the dominant factors influencing tortuosity.

Keywords: porous media, tortuosity, Minkowski functionals, Euler characteristic, machine learning, shortest path.

Зміст

Вступ	4
1 Теоретичні основи дослідження	5
1.1 Пористі середовища: властивості та моделі	5
1.2 Геометрична звивистість	6
1.3 Морфологічні характеристики середовища	7
1.4 Зв'язок звивистості з морфологічними характеристиками	9
2 Постановка задачі	10
2.1 Математичне формулювання	10
2.2 Морфологічні характеристики	12
2.3 Графова модель та пошук найкоротшого шляху	14
2.4 Вектор ознак і вихідна величина	14
3 Реалізація моделі	15
3.1 Загальна структура програмної реалізації	15
3.2 Генерація прямокутних структур	16
3.3 Обчислення морфологічних характеристик	18
3.4 Побудова графа прохідності та пошук шляху	19
3.5 Обчислення звивистості	20
3.6 Формування навчальних вибірок	21
4 Аналіз згенерованих даних	21
4.1 Розподіли пористості та звивистості	22
4.2 Залежність звивистості від пористості та периметра	23
4.3 Поведінка параметра zigzag	25
4.4 Проміжні висновки	26
5 Моделювання та результати прогнозування звивистості	26
5.1 Експеримент 1: випадкові пересічні включення	27
5.2 Експеримент 2: “цегляна” структура	29
5.3 Підсумкові висновки щодо моделей	32
Висновки	33
Література	35

Вступ

Пористі середовища є об'єктами підвищеної уваги в сучасних наукових дослідженнях, оскільки вони відіграють ключову роль у процесах фільтрації, дифузії, тепло- та масопереносу, а також у розробці нових матеріалів з контрольованими транспортними властивостями. Прикладами таких систем є ґрунти, нафтоносні пласти, пористі кераміки, каталізатори, біологічні тканини та мембрани [1, 2]. Однією з основних характеристик, яка визначає ефективність переносу речовини в подібних середовищах, є їхня *геометрична звивистість* (англ. *tortuosity*) — міра відхилення реального шляху потоку або частинки від прямолінійного [3, 4].

Звивистість тісно пов'язана з морфологією середовища: пористістю, топологічною зв'язністю та геометричною складністю мікроструктури. У низці робіт показано, що зміна пористості та топології порового простору призводить до суттєвих змін ефективної дифузії та проникності [2, 3, 4]. Однак значна частина існуючих моделей або надто спрощує структуру матеріалу, або потребує обчислювально затратних симуляцій потоків чи дифузійних процесів у тривимірних зразках. Тому актуальним завданням є розробка ефективних методів прогнозування звивистості на основі чисто геометричних характеристик пористого середовища.

У цій роботі розглядається стохастична цифрова модель пористого середовища, в якій тверда фаза представлена множиною випадково розташованих прямокутних включень у двовимірній області. На відміну від традиційних моделей із круглими порами або простими решітковими структурами [1, 2], такий підхід дозволяє врахувати орієнтацію, видовження та взаємне розташування включень, що істотно впливає на топологію порового простору та реальну довжину шляхів переносу. Для бар'єрних композитних плівок подібні схеми з орієнтованими пластинчастими або «цегляними» включеннями є стандартними при аналізі звивистості дифузійних шляхів.

Сучасні підходи до опису мікроструктур базуються на використанні морфологічних інваріантів — *функціоналів Мінковського*, які дозволяють кількісно оцінювати форму, площу поверхні та топологічну структуру об'єктів [1, 5, 8]. У двовимірному випадку така система включає площу, периметр та характеристику Ейлера–Пуанкаре, що утворює повний морфометричний опис бінарного зображення. Поєднання цих морфологічних параметрів з методами машинного навчання відкриває можливість побудови швидких емпіричних моделей, здатних передбачати звивистість без проведення повноцінних чисельних розрахунків потоків або дифузії у складних випадкових структурах [4].

Метою роботи є розробка математичної та обчислювальної моделі для прогнозування геометричної звивистості стохастичних пористих структур із прямокутними перешкодами на основі їхніх морфологічних характеристик із використанням методів машинного навчання.

Для досягнення поставленої мети необхідно вирішити такі основні **завдання**:

1. проаналізувати існуючі теоретичні підходи до оцінювання звивистості та морфології пористих середовищ [1, 2, 3, 4];
2. побудувати стохастичну модель пористої структури на основі випадкового розподілу прямокутних включень;
3. обчислити морфологічні інваріанти (пористість, функціонали Мінковського, характеристику Ейлера) для згенерованих структур [1, 5, 8];

4. визначити геометричну звивистість за допомогою алгоритмів пошуку найкоротшого шляху в поровому просторі;
5. сформувати набір даних і побудувати моделі машинного навчання для прогнозування звивистості;
6. порівняти результати моделей із відомими аналітичними та емпіричними залежностями для звивистості в пористих середовищах та композитних плівках [3, 4, 9].

Об'єктом дослідження є цифрові моделі випадкових пористих середовищ із прямокутними або паралелепіпедними включеннями. **Предметом дослідження** є залежність геометричної звивистості від морфологічних характеристик середовища та форми твердих включень.

У роботі використано методи стохастичного моделювання, обчислювальної геометрії, морфометричного аналізу та машинного навчання [1, 2, 5, 8]. Розроблена модель дозволяє встановити кількісні зв'язки між морфологічними інваріантами та ефективною звивистістю, а також оцінити вплив форми та орієнтації включень на транспортні властивості середовища.

1 Теоретичні основи дослідження

1.1 Пористі середовища: властивості та моделі

Пористим середовищем називають матеріал, який складається з *твердої матриці* та системи пор, каналів або пустот, заповнених рідиною чи газом [1, 2]. Геометрія та топологія цієї подвійної структури (тверда фаза + поровий простір) визначають такі ефективні характеристики, як проникність, ефективна дифузія, теплопровідність, електропровідність та інші транспортні властивості [1, 2].

Для опису пористих середовищ у літературі використовують низку основних макроскопічних характеристик [1, 2, 3]:

- **Пористість** $\phi = V_{\text{пор}}/V_{\text{заг}}$ — відношення об'єму порового простору до загального об'єму зразка. Це перший базовий параметр, який характеризує, скільки «вільного місця» доступно для руху рідини чи газу.
- **Проникність** — параметр, що описує здатність середовища пропускати потік під дією градієнта тиску (закон Дарсі). Вона залежить не лише від пористості, а й від розмірів, форми та зв'язності пор [2].
- **Топологічна зв'язність** — описує, чи утворюють пори та канали єдину нескінченну кластерну структуру, чи розбиті на ізольовані «острови». Ця властивість пов'язана з характеристикою Ейлера–Пуанкаре та перколяційними властивостями середовища [1, 5].
- **Геометрична звивистість** τ — безрозмірна величина, яка характеризує, наскільки реальний шлях потоку в поровому просторі довший за прямолінійний відрізок між тими ж точками [3, 4].

У теорії випадкових неоднорідних середовищ широко застосовуються **стохастичні цифрові моделі**, в яких середовище дискретизується на регулярну сітку (пікселі або вокселі), а тверда й порова фаза подаються як двійкова маска [5, 8]. Такий підхід зручний для чисельного обчислення морфологічних характеристик (функціоналів Мінковського) та для реалізації алгоритмів пошуку шляхів у поровому просторі [1, 5].

У даній роботі використовується двовимірний цифровий стохастичний модель, де тверда фаза моделюється множиною **прямокутних включень** («перешкод») випадкового розміру та положення. Ці прямокутники можуть частково перекриватися і формувати складну систему вузьких каналів і «розширень», аналогічну до системи пластинчастих наповнювачів у бар'єрних полімерних плівках [4, 9]. Такий опис прямо пов'язаний із класичними моделями дифузійних бар'єрів з орієнтованими пластинками наповнювача, де перешкоди змушують дифундуючі молекули йти довгими звивистими траєкторіями [9].

1.2 Геометрична звивистість

2.2.1. Визначення звивистості

У більшості робіт із переносу в пористих середовищах *геометричну звивистість* визначають як відношення середньої довжини реального шляху через поровий простір до відстані між вхідною та вихідною границями [3, 4]:

$$\tau = \frac{L_{\text{реальний}}}{L_{\text{прямий}}}, \quad \tau \geq 1. \quad (1)$$

Якщо $\tau = 1$, то поровий простір еквівалентний прямому каналу без перешкод; при $\tau > 1$ траєкторії потоку подовжуються за рахунок обходу твердих включень.

У контексті дифузії в бар'єрних плівках звивистість часто вводять через ефективний коефіцієнт дифузії або проникність [4, 9]:

$$D_{\text{еф}} = \frac{D_0}{\tau}, \quad P_{\text{відн}} = \frac{P}{P_0} = \frac{\phi_p}{\tau}, \quad (2)$$

де D_0 і P_0 — дифузія та проникність чистого полімеру, ϕ_p — об'ємна частка полімерної матриці, τ — фактор звивистості, пов'язаний з подовженням траєкторій у присутності непроникних частинок [9].

2.2.2. Чисельне обчислення звивистості в роботі

У нашій моделі звивистість обчислюється суто геометрично. Розглядається прямокутна область, на верхній та нижній границях якої обираються дві точки A та B з однаковою горизонтальною координатою (однаковий x). Це задає «номінальний» напрямок транспорту (зверху вниз).

Тверда фаза (прямокутні перешкоди) задається двійковою маскою, а поровий простір складається з усіх комірок, де немає твердого матеріалу. На цій дискретній сітці застосовується алгоритм пошуку найкоротшого шляху (модифікація алгоритму Дейкстри або A^*) по порових комірках з вагами, що відповідають довжині кроку: 1 — для вертикальних та горизонтальних переходів, $\sqrt{2}$ — для діагональних.

У результаті отримуємо довжину найкоротшого порового шляху L між A та B , а також довжину прямої відстані

$$L_0 = |y_B - y_A|.$$

Далі геометрична звивистість визначається як

$$\tau = \frac{L}{L_0}. \quad (3)$$

Саме це значення τ записується в набір даних і надалі використовується як цільова змінна для моделей машинного навчання.

1.3 Морфологічні характеристики середовища

Щоб описати, *чому* той чи інший зразок має більшу чи меншу звивистість, недостатньо знати лише пористість. Необхідні більш тонкі морфометричні характеристики, які враховують форму, розвиненість поверхні та топологію фаз. Для цього в теорії випадкових середовищ активно використовують **функціонали Мінковського** [1, 5, 8]. У двовимірному випадку вони утворюють повну систему геометричних інваріантів: площу, периметр і характеристику Ейлера.

У даній роботі чисельно обчислюються такі величини:

$$\phi, \quad V_0, \quad V_1, \quad V_2 = \chi, \quad \tau, \quad \text{zigzag}.$$

Далі розглянемо кожен з них докладніше.

2.3.1. Пористість ϕ та площа порового простору V_0

Пористість у дискретній моделі визначається як

$$\phi = \frac{N_{\text{пор}}}{N_{\text{усі}}},$$

де $N_{\text{пор}}$ — кількість «порових» пікселів (комірок сітки, які не належать твердій фазі), а $N_{\text{усі}}$ — загальна кількість пікселів. Ця величина відповідає класичному визначенню $\phi = V_{\text{пор}}/V_{\text{заг}}$ [1, 2].

Функціонал Мінковського V_0 у двовимірному випадку збігається з площею вибраної фази [1, 5]. У цифровій масці порового простору це:

$$V_0 = |\Omega_p| = N_{\text{пор}},$$

тобто V_0 пропорційний площі доступного для переносу каналу. Таким чином, пара величин (ϕ, V_0) фіксує, скільки «вільного місця» є для потоку в цілому.

2.3.2. Периметр межі фаз V_1

Другий функціонал Мінковського V_1 у двовимірному випадку пов'язаний із довжиною межі між твердою та поровою фазами [1, 5]:

$$V_1 \propto P(\partial\Omega_s),$$

де $P(\partial\Omega_s)$ — сумарний периметр твердих включень. У нашій реалізації V_1 обчислюється як

$$V_1 = \frac{1}{4} P_{\text{цифр.}},$$

де $P_{\text{цифр.}}$ — периметр, обчислений для двійкової маски за стандартними алгоритмами цифрової морфології [5, 8].

Фізично V_1 характеризує розвиненість поверхні твердої фази: чим більше дрібних включень, кутів і «зубців» на межі, тим більшим буде V_1 при фіксованій пористості. Для транспортних процесів це означає збільшення площі контакту між потоками в порах і твердою матрицею, а також появу додаткових локальних «обходів», що природно пов'язано зі зростанням звивистості [1, 2].

2.3.3. Характеристика Ейлера–Пуанкаре $V_2 = \chi$

Третій функціонал Мінковського V_2 у 2D є нормованою характеристикою Ейлера–Пуанкаре χ [5]:

$$\chi = N_{\text{компонент}} - N_{\text{отворів}},$$

тобто кількість зв'язних компонент фази мінус число «петель» (отворів) у ній. Для складних стохастичних структур величина χ є важливим топологічним інваріантом, чутливим до перколяційного переходу та утворення «лабіринтів» у поровому просторі [1, 5].

У нашій моделі V_2 обчислюється окремо для твердої фази. Від'ємні значення χ означають домінування «петель» та складних замкнених контурів у структурі включень, що корелює з утворенням численних обхідних шляхів у поровому просторі й, відповідно, зі зростанням звивистості.

2.3.4. Геометрична звивистість τ як макропараметр

Звивистість τ уже визначена вище як відношення L/L_0 . На відміну від ϕ та V_0 , це величина, яка *нелінійно* залежить від усієї геометрії та топології структури: на однаковій пористості різні конфігурації можуть мати дуже різні значення τ [2, 3].

У багатьох аналітичних моделях для композитних плівок звивистість записують у вигляді простих функцій від об'ємної частки наповнювача, її аспектного відношення і параметрів «звужень» між пластинками [4, 9]. Наприклад, моделі, що розвивають підходи Нільсена та подальші модифікації, виражають τ через товщину плівки, довжину й товщину пластинки, об'ємну частку та «wiggles» (зигзаги) траєкторій [4, 9]. У нашій роботі замість явної аналітичної формули використовується *чисельне* визначення τ через пошук найкоротшого шляху в дискретному поровому просторі.

2.3.5. Показник zigzag як міра поперечних відхилень шляху

Окремим внеском, специфічним для даної роботи, є введення **локального показника звивистості zigzag**, який вимірює поперечні відхилення траєкторії від «номінального» напрямку потоку.

Нехай шлях, знайдений алгоритмом найкоротшого маршруту, представлено у вигляді послідовності пікселів

$$(y_0, x_0), (y_1, x_1), \dots, (y_K, x_K),$$

де координата y змінюється від верхнього до нижнього краю, а координата x описує поперечне зміщення. Тоді показник **zigzag** визначається як

$$\text{zigzag} = \frac{1}{K} \sum_{i=1}^K |x_i - x_{i-1}|. \quad (4)$$

Іншими словами, це середня величина *горизонтального кроку* між послідовними точками на шляху. Якщо шлях йде майже строго в одному напрямку, $x_i \approx x_{i-1}$, і **zigzag** близький до нуля. Якщо ж траєкторія змушена часто «петляти» вліво–вправо через розташування перешкод, **zigzag** стає більшим.

Фізично **zigzag** можна інтерпретувати як дискретний аналог *wiggles* (зигзагів траєкторії), які виділяються в аналітичних моделях бар’єрних плівок із пластинчастими наповнювачами [4, 9]. У роботах з моделювання дифузії через композити з орієнтованими пластинками показано, що внесок у звивистість можна розкласти на кілька ефектів: (i) власне подовження траєкторії навколо пластинок, (ii) звуження перетину («slits»), (iii) зигзаги внаслідок неоднорідної орієнтації та розподілу пластинок [4, 9]. Наш показник **zigzag** саме й намагається *відокремлено* кількісно описати третій ефект — поперечні відхилення шляху, які не обов’язково сильно змінюють загальну довжину L , але є важливими для структури потоку.

Таким чином, пара параметрів (τ, zigzag) дає більш багату характеристику шляху:

- τ описує **загальне подовження** траєкторії порівняно з прямою;
- **zigzag** описує **ступінь латерального «петляння»** траєкторії в поперечному напрямку.

Дві структури можуть мати однакову τ , але різний профіль **zigzag**: одна — з плавно вигнутим каналом, інша — з численними локальними «поворотами», що є важливим для механізмів переносу в реальних композитних плівках [4, 9].

1.4 Зв’язок звивистості з морфологічними характеристиками

Класичні підходи намагаються виразити звивистість як функцію кількох глобальних параметрів, наприклад [1, 2, 3]:

$$\tau \approx f(\phi, V_1, \chi). \quad (5)$$

Зменшення пористості ϕ та збільшення периметра V_1 зазвичай ведуть до зростання τ , оскільки канал стає вужчим і більш «ламким». Топологічна характеристика χ відображає, скільки «петель» і розгалужень має структура; складні лабіринтні конфігурації з великою кількістю отворів можуть суттєво підвищувати звивистість [1, 2, 5].

Однак, як показано в роботах з морфологічного аналізу неупорядкованих середовищ, одних лише глобальних функціоналів Мінковського не завжди достатньо для точного відтворення транспортних властивостей [1, 5]. Суттєву роль відіграють *локальні* особливості каналу: наявність «пляшкових горлечок», вузьких шийок, локальних «коридорів» та областей із високою локальною щільністю перешкод. Саме такі ефекти у випадку композитних плівок пов’язують із підвищенням звивистості та зменшенням ефективної дифузії [4, 9]. У нашій постановці ці явища описуються введеним показником **zigzag** та іншими локальними ознаками (у розширеній моделі)

— локальною відкритою площею й числом зв’язних компонент у центральній частині області.

У загальному випадку точний вигляд залежності

$$\tau = f(\phi, V_0, V_1, V_2, \text{zigzag}, \dots)$$

невідомий і є суттєво нелінійним. Аналітичні моделі, запропоновані для композитних бар’єрних плівок, добре працюють для сильно спрощених геометрій та обмеженого діапазону параметрів [4, 9], але їх важко прямо застосувати до випадкових структур із перекривними прямокутними перешкодами.

Тому в даній роботі для відновлення залежності між морфологічними характеристиками та звивистістю τ використовується **підхід машинного навчання**. Спочатку за допомогою стохастичної моделі та алгоритму найкоротшого шляху генерується великий набір $(\phi, V_0, V_1, V_2, \text{zigzag}, \tau)$. Потім на цьому наборі навчаються регресійні моделі (Random Forest, Gradient Boosting, XGBoost, MLP), які апроксимують невідому нелінійну функцію f . Такий підхід узгоджується із сучасними тенденціями в галузі, де класичні морфологічні інваріанти поєднують із методами штучного інтелекту для опису складних пористих структур [1, 4, 5].

2 Постановка задачі

У цьому розділі формулюється математична постановка задачі моделювання двовимірного пористого середовища з прямокутними перешкодами, а також визначаються основні характеристики, які використовуються далі як вхідні ознаки для моделі машинного навчання: пористість, функціонали Мінковського V_0, V_1, V_2 , геометрична звивистість τ та додаткова характеристика “змійкоподібності” траєкторії (*zigzag*). Отримана постановка є спільною для обох чисельних моделей, що розглядаються в роботі (модель з перетинаннями прямокутників та модель “цегляної кладки”).

2.1 Математичне формулювання

1. Геометрична область і фазовий розподіл

Розглядається прямокутна область

$$\Omega = [0, W] \times [0, H] \subset \mathbb{R}^2,$$

яка складається з двох фаз:

$$\Omega = \Omega_p \cup \Omega_s, \quad \Omega_p \cap \Omega_s = \emptyset,$$

де Ω_p — поровий простір (область, доступна для руху частинок), а Ω_s — тверда фаза, що утворює перешкоди.

Тверда фаза моделюється як об’єднання прямокутних включень:

$$\Omega_s = \bigcup_{i=1}^N R_i.$$

Кожне включення R_i задається параметрами центра (x_i, y_i) та півдовжин сторін $a_i/2, b_i/2$:

$$R_i = \left\{ (x, y) \in \mathbb{R}^2 : |x - x_i| \leq \frac{a_i}{2}, |y - y_i| \leq \frac{b_i}{2} \right\}.$$

У першій моделі центри й розміри прямокутників задаються як незалежні випадкові величини з рівномірним розподілом у заданих інтервалах, і перекривання включень дозволяється. Така схема відповідає класичним стохастичним моделям випадкових перешкод у композитах і пористих плівках [4].

У другій моделі використовується напіврегулярна “цегляна” структура: усі прямокутники мають однакову висоту, розташовані рядами, між якими залишаються зазори фіксованої товщини, а сусідні ряди зсунуті на пів довжини типового прямокутника. Така конфігурація дає змогу контролювати ширину каналів і має меншу варіативність, ніж повністю випадкова структура, що корисно для порівняння моделей.

2. Пористість

Пористість ϕ — це частка площі порового простору в загальній області:

$$\phi = \frac{|\Omega_p|}{|\Omega|} = 1 - \frac{|\Omega_s|}{|\Omega|}. \quad (6)$$

Якщо площа кожного прямокутника дорівнює $|R_i| = a_i b_i$ і перекривання не враховуються, то формально

$$|\Omega_s| \approx \sum_{i=1}^N a_i b_i,$$

але в чисельній реалізації площа твердої фази визначається по бінарній сітці (див. нижче), де кожна заповнена комірка відноситься до Ω_s .

Із фізичної точки зору, при великих значеннях пористості поровий простір добре перколює, а шляхи мають невелику звивистість ($\tau \approx 1$). При зменшенні ϕ поровий простір стає більш звуженим і звивистим, що відображається у зростанні τ ; подібні закономірності відмічено в експериментальних та чисельних дослідженнях пористих матеріалів [4].

3. Дискретизація та побудова бінарної маски

Для чисельного аналізу область Ω дискретизується на регулярну сітку розміру $H \times W$ (у роботі $H = W = 128$). Кожній комірці з центром (x_j, y_k) відповідає значення бінарної маски

$$M(j, k) = \begin{cases} 1, & \text{якщо комірка належить твердій фазі } \Omega_s, \\ 0, & \text{якщо комірка належить поровому простору } \Omega_p. \end{cases}$$

У коді така маска зберігається у вигляді логічного масиву `solid[y,x]`. Поровий простір описується масивом `pore = ~solid`, де логічне заперечення означає доступні для руху комірки.

Дискретна маска є основою як для обчислення морфологічних функціоналів, так і для побудови графа прохідності, по якому шукається найкоротший шлях.

2.2 Морфологічні характеристики

1. Функціонали Мінковського

У двовимірному випадку повний набір адитивних, інваріантних відносно рухів та безперервних (у сенсі додавання компактних множин) морфологічних інваріантів складається з трьох функціоналів Мінковського [5, 8]:

- **Площа порового простору V_0 .** У дискретній моделі

$$V_0 = |\Omega_p| = \#\{(j, k) : M(j, k) = 0\},$$

тобто кількість порових пікселів. Нормоване значення $V_0/|\Omega|$ збігається з пористістю ϕ .

- **Периметр фазової межі V_1 .** У безперервному випадку V_1 пропорційний довжині межі між поровою та твердою фазами. У цифровому зображенні його наближено обчислюють як

$$V_1 = \frac{1}{4} P(\partial\Omega_s), \quad (7)$$

де $P(\partial\Omega_s)$ — оцінений периметр на основі метрик, адаптованих до квадратної сітки. В роботі для цього використовується реалізація з бібліотеки `scikit-image`, що фактично враховує горизонтальні, вертикальні та діагональні переходи між пікселями.

- **Характеристика Ейлера–Пуанкаре $V_2 = \chi$.** Вона визначається як

$$\chi = C - H, \quad (8)$$

де C — кількість зв’язних компонент твердої фази, а H — кількість “дірок” (петель) всередині них [5]. Додатні значення χ відповідають простим структурам з невеликою кількістю замкнених пор, тоді як від’ємні — складним лабіринтним конфігураціям з великою кількістю порожнин.

Функціонали Мінковського дають глобальний морфометричний опис пористого середовища, який узгоджується з класичною теорією стохастичної геометрії [5, 8] і широко застосовується для аналізу композитів та випадкових пористих структур [5].

2. Геометрична звивистість

Геометрична звивистість τ описує подовження реального шляху частинки в поровому просторі порівняно з найкоротшою прямою лінією між вхідною та вихідною точками. У загальному вигляді [4, 9]

$$\tau = \frac{L_{\text{реальний}}}{L_{\text{прямий}}}, \quad \tau \geq 1. \quad (9)$$

У нашій дискретній моделі вибираються дві точки на протилежних границях області — “вхід” та “вихід”. Для моделі з довільними перетинаннями прямокутників точки розташовуються на верхній та нижній границі з однаковою випадковою горизонтальною координатою x , але тільки в тих місцях, де відповідні пікселі належать поровому простору. Це гарантує, що пряма вертикальна лінія з’єднує обидві границі.

Реальний шлях L визначається як довжина найкоротшої ламаної, що проходить через порові пікселі й не перетинає тверду фазу. Його знаходять за допомогою алгоритму A^* або Дейкстри на графі прохідності (див. нижче). Пряма відстань L_0 для таких вибраних точок дорівнює різниці координат по y :

$$L_0 = |y_{\text{top}} - y_{\text{bottom}}| = H - 1.$$

Тоді звивистість у чисельній моделі обчислюється як

$$\tau = \frac{L}{L_0}. \quad (10)$$

У класичних моделях переносу маси чи тепла звивистість входить у ефективні коефіцієнти дифузії й теплопровідності у вигляді

$$D_{\text{еф}} \approx \frac{\phi}{\tau^2} D_0,$$

що узгоджується з експериментальними даними для пористих плівок та композитів [4, 9].

3. Характеристика “zigzag” траєкторії

Функціонали Мінковського є глобальними характеристиками структури і не враховують безпосередньо, як саме “вигинається” конкретний шлях крізь середовище. Щоб кількісно оцінити поперечні відхилення траєкторії від прямої, вводиться додаткова скалярна характеристика, яку в роботі позначено як *zigzag*.

Після застосування алгоритму пошуку найкоротшого шляху маємо множину дискретних точок

$$\gamma = \{(y_k, x_k)\}_{k=0}^K,$$

де k — індекс кроку уздовж шляху, x_k — горизонтальна координата, y_k — вертикальна. Для послідовних комірок шляху вводимо поперечні зміщення

$$\Delta x_k = x_k - x_{k-1}, \quad k = 1, \dots, K.$$

Тоді міра “змійкоподібності” (*zigzag*) визначається як

$$\text{zigzag} = \frac{1}{K} \sum_{k=1}^K |\Delta x_k|. \quad (11)$$

Інтуїтивно:

- якщо шлях майже вертикальний, зміщення $|\Delta x_k|$ малі, і *zigzag* близький до нуля;
- якщо траєкторія робить багато поперечних “зигзагів” між перешкодами, середнє $|\Delta x_k|$ збільшується;
- для сильно звивистих каналів *zigzag* набуває великих значень і корелює зі збільшенням τ .

Таким чином, *zigzag* можна розглядати як нормовану оцінку “поперечної варіації” траєкторії, яка доповнює суто геометричну характеристику τ . На відміну від функціоналів Мінковського, що описують структуру середовища, *zigzag* є характеристикою конкретного шляху в даній структурі і тому містить інформацію про локальне розташування перешкод відносно напрямку руху. Запровадження такої ознаки є одним із внесків цієї роботи.

2.3 Графова модель та пошук найкоротшого шляху

1. Побудова графа

На базі бінарної маски `pore` будується орієнтований або неорієнтований зважений граф $G = (V, E)$:

- кожній поровій коміріці (y, x) ставиться у відповідність вершина $v_{yx} \in V$;
- між вершинами проводяться ребра до всіх сусідніх порових комірок (4- або 8-сусідство залежно від варіанту моделі);
- вага ребра дорівнює

$$w = \begin{cases} 1, & \text{для горизонтальних/вертикальних переходів,} \\ \sqrt{2}, & \text{для діагональних переходів.} \end{cases}$$

Точки входу A та виходу B задаються як окремі вершини на протилежних границях області. У моделі з вирівняними по горизонталі точками A і B (верх/низ) це гарантує, що пряма відстань L_0 відома наперед і не залежить від конкретної конфігурації перешкод.

2. Алгоритм пошуку найкоротшого шляху

Для знаходження мінімального шляху між A та B застосовується класичний алгоритм Дейкстри, а в моделі з 8-сусідством — його евристичне розширення A^* , де у якості евристичної функції використовується евклідова відстань до цільової точки. Обидва алгоритми гарантують знаходження шляху мінімальної довжини у зваженому графі з невід’ємними вагами.

Результатом роботи алгоритму є:

- довжина шляху L (сума ваг уздовж траєкторії),
- послідовність пікселів шляху $\gamma = \{(y_k, x_k)\}_{k=0}^K$.

Якщо вершина B недосяжна з A (відсутній перколюючий канал), така реалізація відкидається, оскільки звивистість у цьому випадку формально нескінченна.

2.4 Вектор ознак і вихідна величина

1. Вхідні ознаки

Для кожної згенерованої структури зберігаються значення:

$$\phi, \quad V_0, \quad V_1, \quad V_2, \quad \text{zigzag},$$

які утворюють вектор ознак

$$\mathbf{x}_i = (\phi_i, V_{0,i}, V_{1,i}, V_{2,i}, \text{zigzag}_i). \quad (12)$$

Таким чином, у моделі машинного навчання використовуються чотири стандартні морфологічні характеристики (пористість та три функціонали Мінковського) і одна динамічна характеристика траєкторії — *zigzag*.

Глобальні величини (ϕ, V_0, V_1, V_2) описують, наскільки сильно середовище “заповнене” твердою фазою і наскільки складною є його топологія, тоді як *zigzag* фіксує реальну геометрію проходу крізь цю структуру.

2. Вихідна величина

Вихідною змінною в задачі регресії є геометрична звивистість τ :

$$y_i = \tau_i = \frac{L_i}{L_{0,i}}.$$

Після проведення великої кількості симуляцій формується датасет

$$\mathcal{D} = \{(\mathbf{x}_i, \tau_i)\}_{i=1}^M,$$

на основі якого в подальших розділах будуються та аналізуються моделі машинного навчання (Random Forest, Gradient Boosting, XGBoost, нейронна мережа), що апроксимують невідому нелінійну залежність

$$\tau \approx f(\phi, V_0, V_1, V_2, \text{zigzag}).$$

3 Реалізація моделі

У цьому розділі наведено програмну реалізацію двовимірної моделі пористого середовища з прямокутними перешкодами, а також методику чисельного обчислення геометричної звивистості та додаткової характеристики змійкоподібності шляху (**zigzag**). Модель побудована як дискретизована (піксельна) область, у якій тверда фаза описується множиною прямокутних включень, а поровий простір утворює сітковий граф, по якому можна визначити найкоротший шлях між заданими точками. Кінцевою метою є формування навчальних вибірок $(\mathbf{x}_i, \tau_i)$ для побудови моделей машинного навчання.

У роботі розглядаються дві близькі за ідеєю, але різні за геометрією чисельні моделі:

- **Модель 1** — випадкові прямокутні включення у квадратній області з контролем пористості та вирівнюванням вхідної і вихідної точок по одній вертикалі;
- **Модель 2** — напіврегулярна “цегляна” структура, в якій прямокутники формують ряди з проміжками, а вхід і вихід розташовані на верхній та нижній границях області.

В обох випадках для кожної реалізації обчислюються однакові морфологічні характеристики (ϕ, V_0, V_1, V_2) , геометрична звивистість τ та характеристика **zigzag**.

3.1 Загальна структура програмної реалізації

Реалізація моделі виконана мовою Python із використанням наступних бібліотек:

- NumPy — робота з багатовимірними масивами та генерація випадкових структур;
- scikit-image — обчислення морфологічних характеристик, зокрема периметра та характеристики Ейлера на дискретній сітці [5, 8];
- Matplotlib — побудова зображень середовища та знайдених шляхів;

- `heapq` — реалізація пріоритетної черги для алгоритмів Дейкстри та A^* ;
- `pandas` — формування та збереження датасетів у форматі CSV.

Код розбитий на окремі функціональні блоки:

1. генерація прямокутних включень (дві різні схеми для Моделі 1 і Моделі 2);
2. побудова бінарної матриці прохідності `solid/pore`;
3. обчислення глобальних морфологічних інваріантів (ϕ, V_0, V_1, V_2) ;
4. побудова зваженого графа прохідності;
5. пошук найкоротшого шляху (алгоритм Дейкстри або A^*);
6. обчислення геометричної звивистості τ ;
7. обчислення характеристики змійкоподібності шляху `zigzag`;
8. формування одного запису вибірки;
9. повторення симуляцій для побудови датасетів з тисячі прикладів

Такий модульний підхід дозволяє легко змінювати параметри моделювання, додавати нові ознаки або варіювати алгоритми пошуку шляху.

3.2 Генерація прямокутних структур

1. Модель 1: випадкові прямокутники з перекриванням

Область моделювання визначається як квадрат

$$\Omega = [0, W] \times [0, H], \quad W = H = 128.$$

Для кожної реалізації випадковим чином задається кількість прямокутників N у межах $[N_{\min}, N_{\max}]$, після чого послідовно генеруються включення із випадковими розмірами:

$$h_i \in [h_{\min}, h_{\max}], \quad w_i \in [w_{\min}(h_i), w_{\max}],$$

де мінімальна ширина $w_{\min}(h_i)$ вибирається так, щоб прямокутники були видовженими (у коді застосовується умова $w_{\min} \geq 2h_i$). Розташування верхнього лівого кута (x_i, y_i) береться з рівномірних інтервалів, що гарантують повне входження прямокутника в межі області.

На відміну від класичних пакувальних моделей, у Моделі 1 **перекривання прямокутників дозволяється**. Кожне нове включення просто “накладається” на вже наявні, і в масиві `solid` усі пікселі, що потрапляють у зону $[x_i, x_i + w_i] \times [y_i, y_i + h_i]$, позначаються як належні твердій фазі. Якщо декілька прямокутників перекриваються, відповідні пікселі залишаються заповненими (`solid = True`), тобто формується об’єднання всіх включень.

У підсумку утворюється стохастична структура з видовженими прямокутними перешкодами, які можуть злипатися та формувати складні кластери, що наближує модель до реалістичних композитних середовищ із високим наповненням.



Рис. 1: Приклад згенерованого пористого середовища для Моделі 1 (випадкові прямокутники з перекриванням) із позначеним найкоротшим шляхом.

2. Модель 2: “цегляна” структура

У другій моделі прямокутники формують ряди висотою h , що чергуються із вертикальними проміжками фіксованої товщини. У кожному рядку відрізки заповненої твердої фази мають випадкову довжину в межах

$$w_i \in [w_{\min}, w_{\max}],$$

але розташовані послідовно, без перекривання всередині рядка, з горизонтальними зазорами між ними. Сусідні ряди зміщені відносно один одного приблизно на половину середньої довжини прямокутників, утворюючи структуру, подібну до цегляної кладки, що призводить до характерної мережі зміщених каналів.

Такий варіант дозволяє дослідити, як змінюється звивистість у більш регулярних, квазіперіодичних структурах при тих самих глобальних морфологічних параметрах.

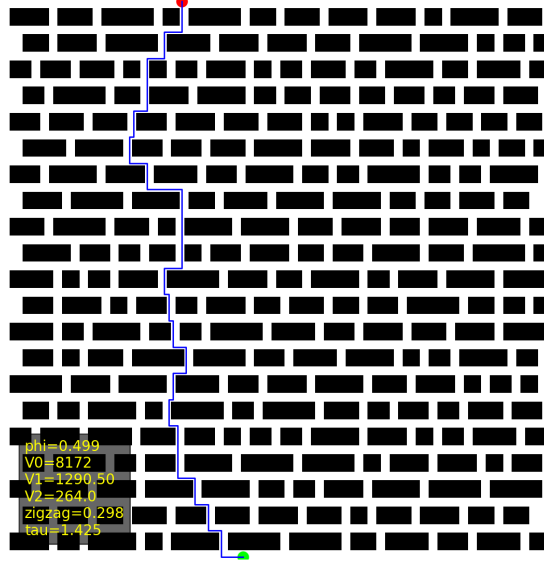


Рис. 2: Приклад згенерованого пористого середовища для Моделі 2 (“цегляна” структура) із позначеним транспортним каналом.

3. Бінарна маска фаз

На основі параметрів включень будується логічний масив `solid[y,x]`, де значення `True` відповідає твердій фазі Ω_s , а `False` — поровому простору Ω_p . Порову фазу зручно представляти як `pore = ~solid`. Цей масив використовується як для обчислення морфологічних функціоналів, так і для побудови графа прохідності.

3.3 Обчислення морфологічних характеристик

1. Пористість і площа порового простору

Пористість чисельно обчислюється як

$$\phi = 1 - \frac{1}{HW} \sum_{y,x} \text{solid}[y,x],$$

тобто як частка порових пікселів. Площа порового простору (функціонал Мінковського V_0) дорівнює кількості елементів `False`:

$$V_0 = \#\{(y, x) : \text{solid}[y, x] = 0\}.$$

У нормованому вигляді $V_0/(HW) = \phi$.

2. Периметр V_1

Функціонал Мінковського V_1 пов’язаний з довжиною межі між поровою та твердою фазами. На дискретній сітці периметр $\partial\Omega_s$ оцінюється за допомогою функції

`measure.perimeter` з бібліотеки `scikit-image`, яка враховує сумарну довжину переходів між заповненими й порожніми пікселями [8]. У нашій постановці значення V_1 визначається відповідно до формули (7), де множник $1/4$ узгоджує одиниці вимірювання з розмірністю клітинки.

Велике значення V_1 відповідає сильно розвиненій поверхні твердих включень, що, як правило, збільшує складність порових каналів.

3. Характеристика Ейлера V_2

Третій функціонал Мінковського V_2 у двовимірному випадку співпадає з характеристикою Ейлера–Пуанкаре χ , яка визначає топологічну складність твердої фази як різницю між кількістю зв'язних компонент і кількістю «отворів» [5]. У нашій роботі значення V_2 обчислюється відповідно до формули (8). У програмній реалізації характеристику χ отримують за допомогою функції `measure.euler_number` з параметром `connectivity=2`, що відповідає 8-зв'язності на квадратній сітці.

Додатні значення χ відповідають простішим конфігураціям, тоді як від'ємні свідчать про наявність численних перфорацій та лабіринтної структури твердої фази [5].

4. Характеристика змійкоподібності шляху `zigzag`

Після знаходження найкоротшого шляху зберігається послідовність пікселів траєкторії

$$\gamma = \{(y_k, x_k)\}_{k=0}^K.$$

Щоб кількісно оцінити поперечні відхилення шляху від «ідеального» вертикального руху, вводиться характеристика змійкоподібності шляху. У нашій роботі показник `zigzag` визначається відповідно до формули (11), яка відображає середню величину горизонтального зміщення між сусідніми точками траєкторії. Якщо шлях проходить майже по прямій (мінімум обхідних маневрів), значення `zigzag` близьке до нуля. Для структур, у яких частинка змушена багаторазово зміщуватися вправо-вліво, ця величина суттєво зростає. Таким чином, `zigzag` доповнює глобальні морфологічні інваріанти, безпосередньо характеризуючи геометрію конкретного шляху в даному середовищі.

3.4 Побудова графа прохідності та пошук шляху

1. Вершини та ребра графа

На основі масиву `pore` будується граф $G = (V, E)$: кожній поровій клітинці (y, x) відповідає вершина v_{yx} ; між вершинами проводяться ребра до сусідніх порових комірок.

У Моделі 1 використовується 8-сусідність:

$$(dy, dx) \in \{(\pm 1, 0), (0, \pm 1), (\pm 1, \pm 1)\},$$

що дозволяє діагональні переходи. Вага ребра дорівнює

$$w = \begin{cases} 1, & \text{для горизонтальних і вертикальних переходів,} \\ \sqrt{2}, & \text{для діагональних.} \end{cases}$$

У Моделі 2 застосовується 4-сусідність $(\pm 1, 0), (0, \pm 1)$, тобто рух дозволено лише по вертикалі та горизонталі, що робить топологію графа простішою.

2. Вибір стартової та кінцевої точок

Для Моделі 1 точки входу та виходу беруться на верхній та нижній границях області. Спочатку на обох границях шукаються стовпчики x , у яких верхній і нижній пікселі належать поровому простору; з множини допустимих x випадково вибирається один, після чого стартова та кінцева точки мають координати

$$A = (y_{\text{top}}, x), \quad B = (y_{\text{bottom}}, x).$$

Це гарантує, що пряма відстань L_0 є сталою (рівною висоті області) для всіх реалізацій.

У Моделі 2 старт і фініш також розташовуються на верхній та нижній границях, але їх горизонтальні координати відбираються незалежно в межах центральної частини області, що відображає іншу геометрію каналів.

3. Алгоритми пошуку найкоротшого шляху

У Моделі 1 для пошуку шляху використовується алгоритм A^* з евклідовою евристикою:

$$h((y, x)) = \sqrt{(y - y_B)^2 + (x - x_B)^2}.$$

Алгоритм A^* узагальнює Дейкстру та дозволяє прискорити обчислення за рахунок спрямованого “просування” фронту у напрямку до цілі.

У Моделі 2 застосовується класичний алгоритм Дейкстри на графі з невід’ємними вагами. В обох випадках результатом є:

- мінімальна довжина шляху L між A та B ;
- послідовність вершин шляху γ , яка потім використовується для обчислення zigzag.

Якщо цільова вершина B недосяжна (відсутня перколююча компонента порового простору), реалізація відкидається і не потрапляє до датасету.

3.5 Обчислення звивистості

Пряма відстань між A і B у побудованих конфігураціях дорівнює різниці координат по вертикалі:

$$L_0 = |y_{\text{bottom}} - y_{\text{top}}|.$$

Мінімальна довжина шляху L вже отримана після роботи алгоритму A^* або Дейкстри як сума ваг уздовж траєкторії.

Геометрична звивистість обчислюється як

$$\tau = \frac{L}{L_0}, \quad \tau \geq 1. \quad (13)$$

Це відповідає класичному визначенню звивистості в теорії ефективного переносу в пористих середовищах [4, 9].

Якщо шлях проходить практично по прямій (мінімум обхідних маневрів), одержуємо $\tau \approx 1$; сильна невпорядкованість та звуження каналів приводять до значень $\tau > 1.5$.

3.6 Формування навчальних вибірок

Для кожної успішної симуляції (тобто для кожної структури, в якій існує прохідний канал між A та B) обчислюється набір величин

$$\phi, \quad V_0, \quad V_1, \quad V_2, \quad \text{zigzag}, \quad \tau.$$

Ці значення зберігаються у вигляді одного рядка таблиці `pandas`, що інтерпретується як вектор ознак та цільова змінна:

$$\mathbf{x}_i = (\phi_i, V_{0,i}, V_{1,i}, V_{2,i}, \text{zigzag}_i), \quad y_i = \tau_i.$$

Повторення процедури моделювання M разів дає датасет

$$\mathcal{D} = \{(\mathbf{x}_i, \tau_i)\}_{i=1}^M,$$

який експортується у формат `CSV`. Для Моделі 1 формується файл `dataset.csv`, для Моделі 2 — файл `reduced_minimal_dataset.csv`. У кожному випадку кількість елементів вибірки становить приблизно $M \approx 10^3$, що забезпечує достатню статистику для подальшого навчання та оцінки моделей машинного навчання.

4 Аналіз згенерованих даних

У цьому розділі проаналізовано статистичні властивості двох наборів даних, згенерованих за допомогою описаних у попередньому розділі моделей пористого середовища:

- **Модель 1** — випадкові прямокутні перешкоди, що можуть перетинатися; набір даних `dataset.csv`;
- **Модель 2** — “цегляна” структура з вирівняними горизонтальними рядами прямокутників; набір даних `reduced_minimal_dataset.csv`.

В обох випадках для кожної реалізації середовища обчислюється один і той самий набір морфологічних характеристик:

$$\phi, \quad V_0, \quad V_1, \quad V_2, \quad \text{zigzag}, \quad \tau,$$

де ϕ — пористість, V_0, V_1, V_2 — функціонали Мінковського, τ — геометрична звивистість, а `zigzag` — середнє абсолютне горизонтальне зміщення на один крок вздовж знайденого найкоротшого шляху.

Кожен датасет містить 1000 реалізацій, які надалі використовуються для побудови моделей машинного навчання.

Параметри генерації структур

Перед аналізом розподілів зафіксуємо параметри стохастичної генерації пористого середовища для обох моделей.

Табл. 1: Параметри генерації середовища для Моделі 1 (випадкові прямокутники).

Розмір області	$W = H = 128$ пікселів
Кількість прямокутників	$N \in [330, 480]$
Ширина прямокутників	$w \in [1, 3]$
Висота прямокутників	$h \in [2, 6]$
Цільова пористість	$\phi \in [0.50, 0.9]$

Табл. 2: Параметри генерації середовища для Моделі 2 (“цегляна” структура).

Розмір області	$W = H = 128$ пікселів
Кількість прямокутників	$N \in [250, 430]$
Ширина прямокутників	$w \in [4, 12]$
Цільова пористість	$\phi \in [0.30, 0.9]$

4.1 Розподіли пористості та звивистості

На рис. 3 показано розподіл звивистості для першої моделі.

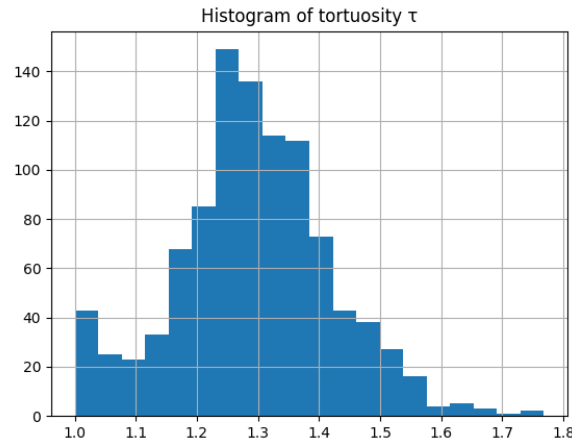


Рис. 3: Модель 1. Розподіл геометричної звивистості τ .

Розподіл звивистості для Моделі 1 має максимум у діапазоні 1.2–1.4, при цьому крайні значення наближаються до $\tau \approx 1.0$ (майже прямий шлях) та $\tau \approx 1.8$ (дуже звивисті траєкторії). Такий широкий розкид зумовлений тим, що прямокутники можуть перекриватися, утворюючи як порівняно просту, так і дуже лабіринтну геометрію.

Для другої, “цегляної” моделі гістограму наведено на рис. 4.

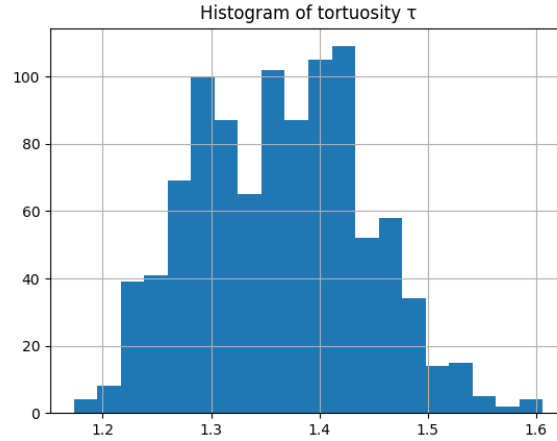


Рис. 4: Модель 2. Розподіл геометричної звивистості τ .

Розподіл звивистості для другої моделі охоплює інтервал $1.18 \lesssim \tau \lesssim 1.62$, причому більшість значень зосереджена поблизу центру інтервалу ($\tau \approx 1.35\text{--}1.45$). Порівняно з першою моделлю розкид менший, що відображає більш регулярну структуру каналів.

4.2 Залежність звивистості від пористості та периметра

На рис. 5–6 подано розсіювання пар (ϕ, τ) та (V_1, τ) для першої моделі.

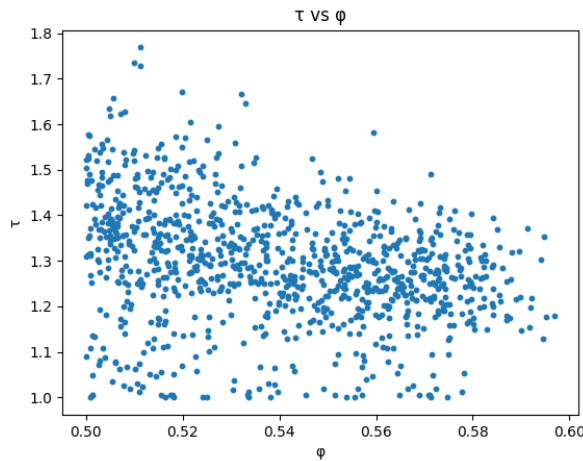


Рис. 5: Модель 1. Залежність звивистості від пористості.

У першому датасеті видно слабо виражений спадний тренд: за збільшення ϕ звивистість у середньому зменшується, однак для кожного фіксованого рівня пористості існує великий розкид значень τ . Це означає, що *самой пористості недостатньо для однозначного передбачення довжини шляху*.

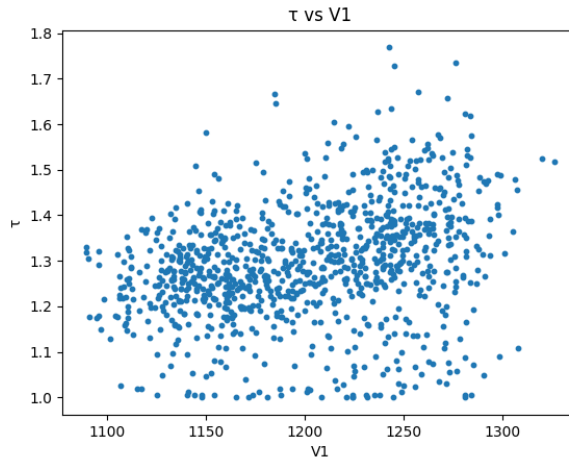


Рис. 6: Модель 1. Залежність звивистості від периметра твердої фази V_1 .

Аналогічно, зі збільшенням V_1 (розвиненості межі твердої фази) значення τ в середньому зростають, проте залежність залишається доволі “розмитою”.

Для другої моделі (рис. 7–8) картина інша.

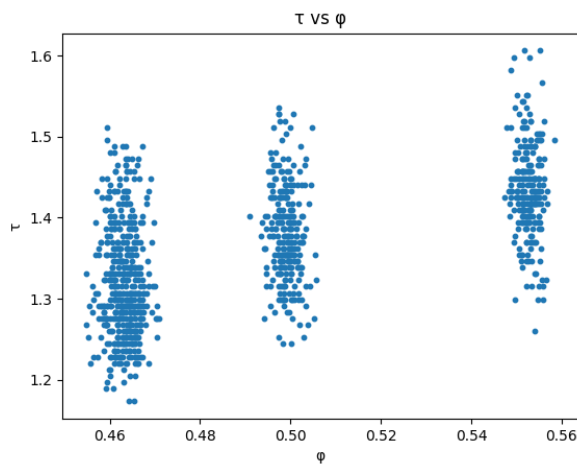


Рис. 7: Модель 2. Залежність звивистості від пористості.

На графіку τ – ϕ видно три вузькі вертикальні “хмари”, що відповідають трьом кластерам пористості. Усередині кожного кластера значення τ змінюються в досить вузькому діапазоні, тобто пористість тут вже тісніше пов’язана з довжиною шляху, ніж у випадку повністю випадкових перешкод.

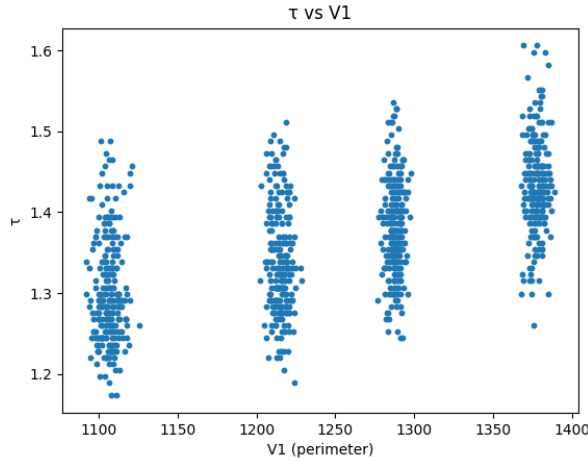


Рис. 8: Модель 2. Залежність звивистості від V_1 .

Залежність τ від периметра V_1 для “цегляної” геометрії також має кластерну структуру; однак і тут помітно, що збільшення V_1 (більш “розірвані” та довгі перешкоди) веде до деякого зростання звивистості.

4.3 Поведінка параметра zigzag

Найяскравіші залежності виявляються для введеного нами параметра **zigzag**, який описує середню горизонтальну зміну координати шляху на один крок. У нашій роботі це значення обчислюється відповідно до формули (11), що кількісно відображає ступінь латеральних відхилень траєкторії.

По суті, **zigzag** вимірює *ступінь бокового “блукання” шляху*: для майже прямолінійного проходу він близький до нуля, а для сильно зигзагоподібної траєкторії зростає.

На рис. 9 показано залежність τ від **zigzag** для першого датасету.

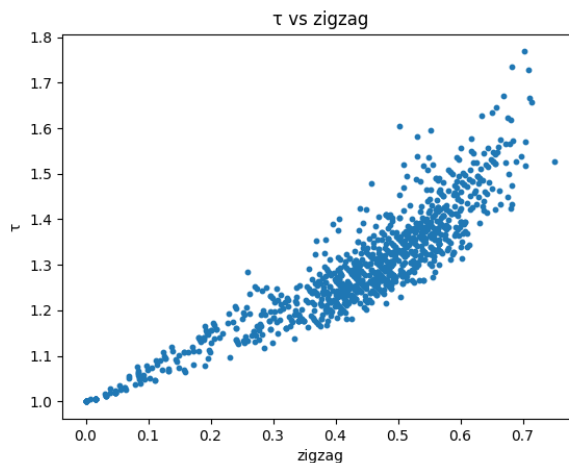


Рис. 9: Модель 1. Звивистість як функція параметра **zigzag**.

Точки лежать вздовж чіткої зростаючої кривої: зі збільшенням зигзагоподібності шляху значення τ монотонно зростає. Розкид навколо гладкої залежності невеликий, що свідчить про те, що **zigzag** вже майже однозначно визначає звивистість.

У другій моделі ця залежність стає ще більш вираженою (рис. 10).

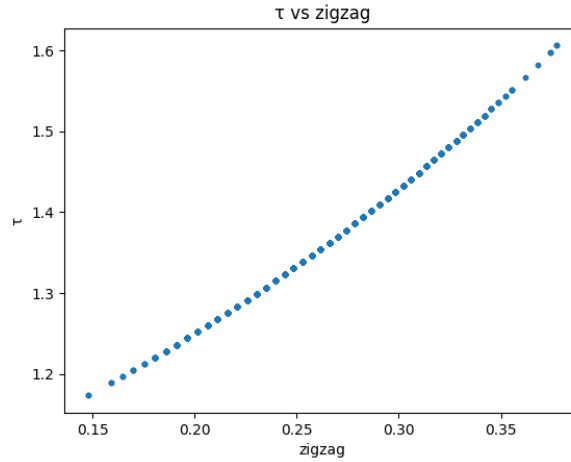


Рис. 10: Модель 2. Залежність τ від zigzag.

Графік практично утворює гладку, майже одновимірну криву без помітного розкиду. Це означає, що в “цегляній” конфігурації задані нами глобальні морфологічні параметри та спосіб вибору старту/фінішу роблять zigzag фактично *єдиним* суттєвим предиктором звивистості.

4.4 Проміжні висновки

Отримані результати можна підсумувати так:

- пористість ϕ та периметр V_1 демонструють коректні фізичні тренди (менше ϕ та більше V_1 — більша τ), але поодиночі не дозволяють точно передбачити звивистість;
- для випадкової моделі з пересічними прямокутниками розподіли ϕ і τ досить широкі, що відображає високу варіативність структури;
- у “цегляній” моделі пористість та звивистість мають дискретнішу, кластерну структуру через регулярну геометрію;
- введений нами параметр zigzag виявляється надзвичайно інформативним: у другій моделі він майже повністю визначає значення τ .

Ці спостереження пояснюють, чому в наступному розділі моделі машинного навчання, що використовують набір ознак $(\phi, V_0, V_1, V_2, \text{zigzag})$, демонструють дуже високу точність, особливо коли структура має регулярний “цегляний” характер.

5 Моделювання та результати прогнозування звивистості

У цьому розділі наведено результати побудови моделей машинного навчання для прогнозування геометричної звивистості τ на основі морфологічних характеристик згенерованих середовищ. Окремо аналізуються дві серії експериментів:

- **Експеримент 1** (Модель 1): випадкові прямокутники, що можуть перетинатися; набір `dataset.csv`;
- **Експеримент 2** (Модель 2): “цегляна” структура; набір `reduced_minimal_dataset.csv`.

В обох серіях використовується той самий вектор ознак

$$\mathbf{x} = (\phi, V_0, V_1, V_2, \text{zigzag}),$$

а цільова змінна — звивистість τ .

Для навчання застосовано чотири регресійні моделі:

- Random Forest Regressor;
- Gradient Boosting Regressor;
- XGBoost Regressor;
- багатошаровий персептрон (MLP) на базі Keras.

Дані поділяються на навчальну й тестову частини у пропорції 80/20, а ознаки попередньо стандартизуються.

5.1 Експеримент 1: випадкові пересічні включення

1. Точність моделей

Для першого датасету (`dataset.csv`) були отримані такі значення метрик (MAE, RMSE, R^2):

Табл. 3: Експеримент 1. Метрики якості моделей

Модель	MAE	RMSE	R^2
Random Forest	0.0396	0.0557	0.810
Gradient Boosting	0.0375	0.0541	0.821
XGBoost	0.0383	0.0541	0.820
MLP	0.0446	0.0598	0.781

Усі ансамблеві моделі дерев (Random Forest, Gradient Boosting, XGBoost) демонструють дуже близькі результати з коефіцієнтом детермінації $R^2 \approx 0.81$ – 0.82 . Багатошаровий персептрон дещо поступається, але все одно надає задовільну точність ($R^2 \approx 0.78$).

Графіки “істинне проти прогнозованого” для цього експерименту показані на рис. 11–14.

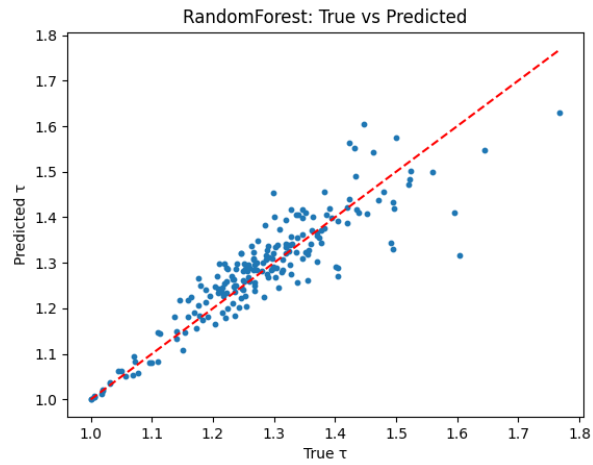


Рис. 11: Експеримент 1. Random Forest: істинні та прогнозовані значення τ .

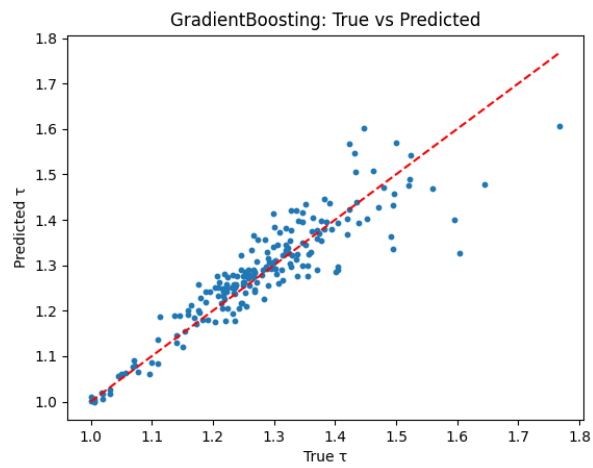


Рис. 12: Експеримент 1. Gradient Boosting: істинні та прогнозовані значення τ .

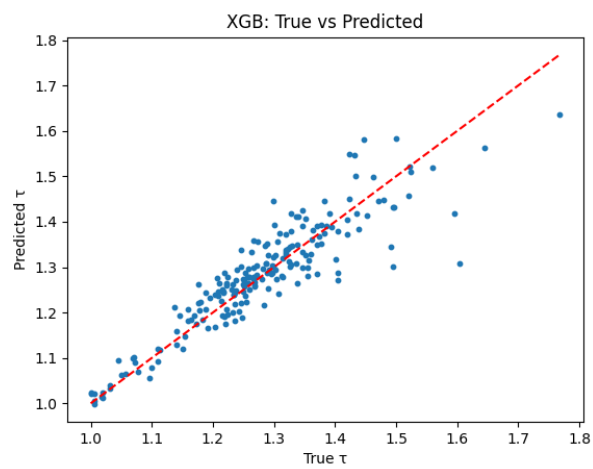


Рис. 13: Експеримент 1. XGBoost: істинні та прогнозовані значення τ .

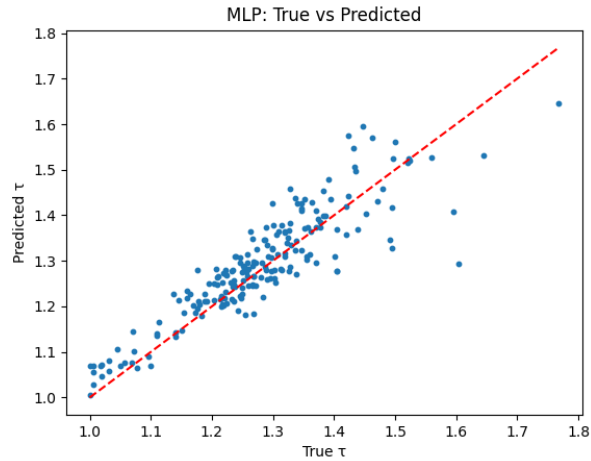


Рис. 14: Експеримент 1. MLP: істинні та прогнозовані значення τ .

У всіх випадках точки тісно згруповані навколо діагоналі $y = x$, що свідчить про хорошу якість відновлення τ .

2. Важливість ознак

Для Random Forest було проаналізовано важливості ознак (рис. 15).

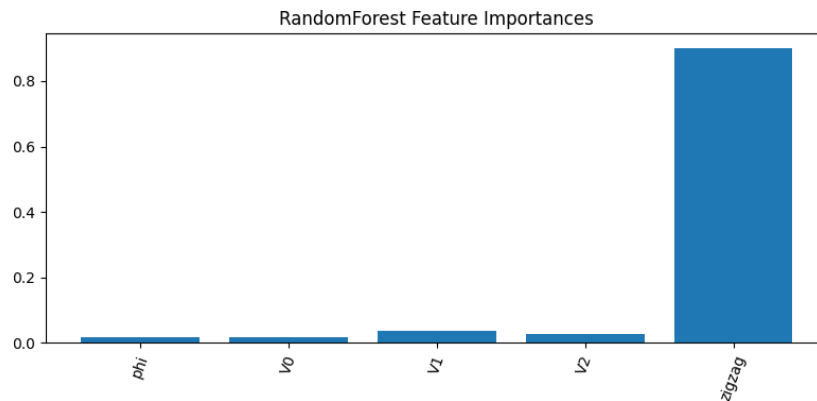


Рис. 15: Експеримент 1. Важливість ознак у моделі Random Forest.

Бачимо, що майже вся вага припадає на параметр **zigzag** (понад 90% сумарної важливості), тоді як ϕ , V_0 , V_1 , V_2 відіграють лише допоміжну роль. Це узгоджується з попереднім аналізом: саме середня бокова “зигзагоподібність” шляху практично однозначно визначає геометричну звивистість, а глобальні морфологічні інваріанти лише трохи коригують результат.

5.2 Експеримент 2: “цегляна” структура

1. Точність моделей

Для другого датасету (`reduced_minimal_dataset.csv`) отримано ще вищі показники якості:

Табл. 4: Експеримент 2. Метрики якості моделей

Модель	MAE	RMSE	R^2
Random Forest	0.00018	0.00101	0.99984
Gradient Boosting	0.00008	0.00111	0.99980
XGBoost	0.00100	0.00163	0.99958
MLP	0.00419	0.00560	0.99503

Ансамблеві моделі майже ідеально відтворюють значення τ ($R^2 > 0.9995$), а навіть MLP досягає $R^2 \approx 0.995$. Такі результати означають, що в “цегляній” геометрії обрана система ознак практично повністю описує звивистість.

Це добре видно на графіках “істинне проти прогнозованого” (рис. 16–19): всі точки лежать майже точно на діагоналі.

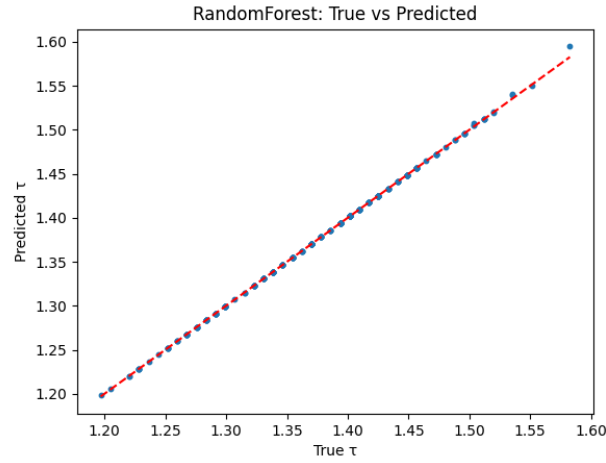


Рис. 16: Експеримент 2. Random Forest: істинні та прогнозовані значення τ .

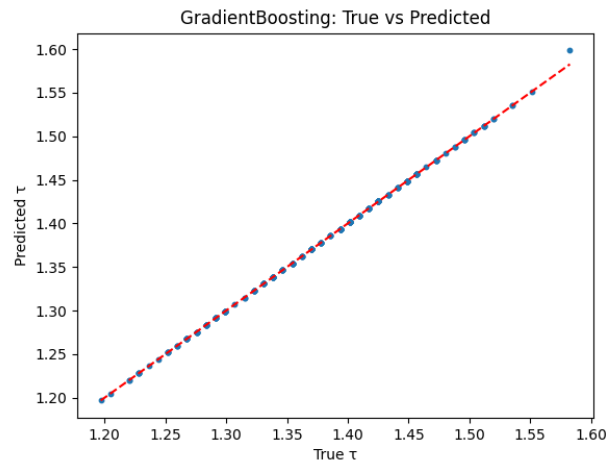


Рис. 17: Експеримент 2. Gradient Boosting: істинні та прогнозовані значення τ .

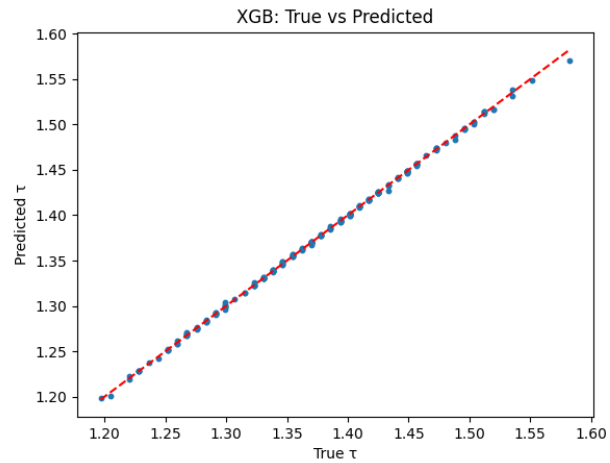


Рис. 18: Експеримент 2. XGBoost: істинні та прогнозовані значення τ .

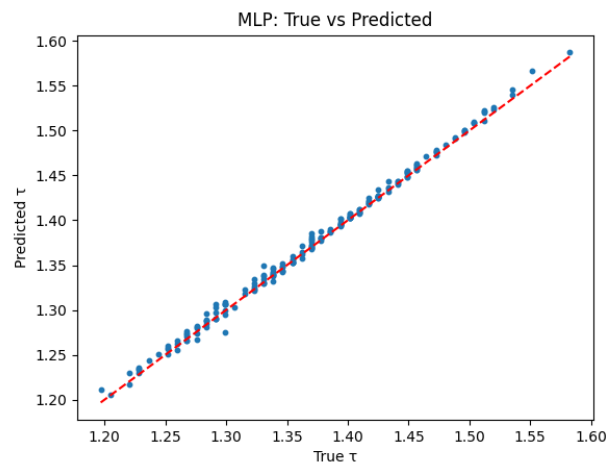


Рис. 19: Експеримент 2. MLP: істинні та прогнозовані значення τ .

2. Важливість ознак

Для другої моделі розподіл важливостей у Random Forest ще більш “екстремальний” (рис. 20).

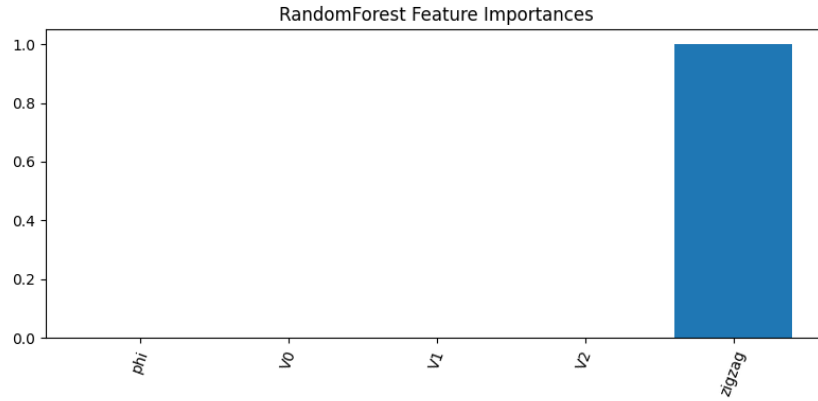


Рис. 20: Експеримент 2. Важливість ознак у моделі Random Forest.

Усі Minkowski-функціонали (ϕ , V_0 , V_1 , V_2) мають майже нульову важливість, тоді як параметр **zigzag** фактично один пояснює зміну τ . Це природний наслідок того, що у “цегляній” геометрії глобальні морфологічні параметри жорстко пов’язані між собою, і **zigzag** вже відображає їхній сукупний вплив на форму шляху.

5.3 Підсумкові висновки щодо моделей

- Для більш “хаотичної” моделі з пересічними прямокутниками обрана система ознак дозволяє відтворити звивистість із досить хорошою точністю ($R^2 \approx 0.8$).
- Для більш регулярної “цегляної” структури ті ж самі ознаки дають майже ідеальне передбачення ($R^2 > 0.999$), що вказує на майже детермінований зв’язок між **zigzag** та τ .
- У всіх моделях ключовим предиктором є введений нами параметр **zigzag**, тоді як Minkowski-функціонали відіграють роль невеликих коригувальних чинників.
- Таким чином, запропонована комбінація *узагальнених* морфологічних інваріантів та *траекторної* характеристики **zigzag** виявляється достатньою для швидкого й точного прогнозування геометричної звивистості в широкому класі штучних пористих середовищ.

Висновки

У роботі розроблено та досліджено комплексну математичну й обчислювальну модель двовимірних пористих середовищ з прямокутними включеннями. Метою було кількісно описати залежність геометричної звивистості τ від морфологічних характеристик середовища та перевірити можливість її точного прогнозування за допомогою методів машинного навчання на основі геометричних ознак.

Було послідовно виконано такі основні кроки:

1. Побудовано дві стохастичні моделі пористого середовища з прямокутними перешкодами:
 - модель з випадково розташованими та такими, що можуть перетинатися, прямокутниками;
 - “цегляну” модель з регулярними рядами непротинаючих прямокутних включень.
2. Реалізовано дискретизацію області, побудову графа прохідності та пошук найкоротшого шляху в поровому просторі (алгоритм A^* з 8-зв’язністю для першої моделі та чотиризв’язний алгоритм Дейкстри для “цегляної” структури).
3. Для кожної конфігурації обчислено морфологічні характеристики, зокрема пористість ϕ , функціонали Мінковського V_0 , V_1 та характеристику Ейлера–Пуанкаре $V_2 = \chi$ [1, 5].
4. Запропоновано й реалізовано нову геометричну ознаку *zigzag* — середнє горизонтальне зміщення траєкторії вздовж шляху, що кількісно описує ступінь поперечних відхилень частинки від напрямку транспорту.
5. Згенеровано два набори даних:
 - перший — для випадкових прямокутників, що перетинаються;
 - другий — для “цегляної” структури.

Кожен датасет містить понад 1000 реалізацій з повним вектором ознак $(\phi, V_0, V_1, V_2, \text{zigzag})$ та відповідною звивистістю τ .

6. На основі отриманих даних побудовано та порівняно моделі машинного навчання: Random Forest, Gradient Boosting, XGBoost та багатопаровий перцептрон (MLP).

Отримані чисельні результати дозволяють сформулювати такі наукові висновки.

- У обох розглянутих геометріях звивистість демонструє фізично природну поведінку: τ зростає зі зменшенням пористості та збільшенням периметра твердої фази V_1 , що узгоджується з класичними уявленнями про вплив мікроструктури на перенесення в пористих середовищах [1, 3].
- Функціонали Мінковського (V_0, V_1, V_2) є коректними глобальними дескрипторами структури, але їхньої інформації недостатньо для точного відновлення τ . Розкид значень звивистості при близьких значеннях ϕ та V_1 показує, що *локальна* форма каналу та характер вигинів траєкторії відіграють вирішальну роль.

- Запропонована ознака *zigzag* виявилася домінуючим фактором: у моделях RF та XGBoost її важливість наближається до 90–100%, тоді як внесок ϕ , V_0 , V_1 та V_2 є порівняно малим. Графіки залежності τ від *zigzag* для обох датасетів демонструють майже одновимірний нелінійний тренд.
- Для моделі з випадковими прямокутниками, що перетинаються отримано точність прогнозування $MAE \approx 0.04\text{--}0.06$ та $R^2 \approx 0.80\text{--}0.82$ для ансамблевих методів (Random Forest, Gradient Boosting, XGBoost), а також $R^2 \approx 0.78$ для нейронної мережі MLP. Це означає, що побудована модель адекватно відтворює нелінійну залежність $\tau = f(\phi, V_0, V_1, V_2, \text{zigzag})$ у випадку сильно стохастичної геометрії.
- Для “цегляної” структури отримано практично детермінований зв’язок між ознаками та звивистістю: дерева рішень (Random Forest, Gradient Boosting, XGBoost) демонструють значення $R^2 > 0.999$, $MAE \sim 10^{-4}$ при дуже щільному розташуванні точок уздовж прямої $y = x$ на графіках “істинне проти прогнозованого”. Нейронна мережа також показує високу точність ($R^2 \approx 0.995$).
- Порівняння двох геометрій показує, що *структурна регулярність* (“цегляний” малюнок) різко підвищує передбачуваність звивистості за геометричними ознаками. У випадкових конфігураціях з перекриттям прямокутників зберігається істотна стохастична компонента, але навіть у цьому випадку використання ознаки *zigzag* дозволяє досягти високих значень R^2 .
- Розроблені моделі дозволяють прогнозувати геометричну звивистість без виконання повного розрахунку найкоротшого шляху, що відкриває можливість прискореної оцінки ефективних транспортних коефіцієнтів (дифузії, теплопровідності тощо) через відомі співвідношення виду $D \sim D_0/\tau^2$ [9].

Отже, поєднання морфологічного аналізу на основі функціоналів Мінковського [1, 5] із сучасними методами машинного навчання *дозволяє побудувати швидкі та досить точні емпіричні моделі звивистості* для класу прямокутних пористих структур. Запропонований підхід є гнучким і може бути розширений у кількох напрямках:

- поширення на тривимірні середовища з довільними включеннями;
- поєднання з прямим чисельним моделюванням переносу (методи скінченних різниць, Lattice Boltzmann) для оцінки ефективної проникності й дифузії [3];
- вивчення узагальнених ознак траєкторій, подібних до *zigzag*, для інших типів мікроструктур.

Таким чином, результати роботи підтверджують практичну доцільність використання геометричних дескрипторів разом із методами машинного навчання для швидкого прогнозування транспортних властивостей пористих матеріалів.

Література

- [1] Torquato, S. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. Springer, New York, 2002.
- [2] Sahimi, M. *Flow and Transport in Porous Media and Fractured Rock: From Classical Methods to Modern Approaches*. Second Edition, Wiley-VCH Verlag GmbH & Co. KGaA, 2011.
- [3] Coelho, D., Shapiro, M., Thovert, J.-F., & Adler, P. M. *Electroosmotic Phenomena in Porous Media*. Journal of Colloid and Interface Science, 181(1), 169–190, 1996.
- [4] Idris, A., Muntean, A., & Mesić, B. *A Review on Predictive Tortuosity Models for Composite Films in Gas Barrier Applications*. Journal of Coatings Technology and Research, 19(3), 699–716, 2022.
- [5] Arns, C. H., Knackstedt, M. A., Pinczewski, W. V., & Mecke, K. R. *Euler–Poincaré Characteristics of Classes of Disordered Media*. Physical Review E, 63, 031112, 2001.
- [6] Van Rossum, G., & Drake, F. L., Jr. *Python 3 Reference Manual*. CreateSpace Independent Publishing Platform, 2009.
- [7] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. Nature Methods, 17(3), 261–272, 2020.
- [8] Serra, J. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982. (Класична монографія з математичної морфології)
- [9] Cussler, E. L. *Diffusion Barriers*. Diffusion Fundamentals, 6, 72.1–72.12, 2007.