

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

Кваліфікаційна робота
бакалавра

на тему **«Математичне моделювання еволюції
пухлини за різними наборами параметрів
моделі»**

Виконала: студентка групи МП41,
спеціальність
113 – Прикладна математика,
освітньо-професійна програма
«Прикладна математика»
Седюк А.Д.

Науковий керівник: доктор фіз.-мат. наук,
професор,
професор кафедри
прикладної математики
Кізілова Н.М.

Рецензент: доктор технічних наук,
професор,
професор кафедри біомедичної
інженерії
Харківського національного
університету радіоелектроніки
Музика К.М.

Харків — 2025 рік

Анотації

Седюк Анастасія Денисівна. Математичне моделювання еволюції пухлини за різними наборами параметрів моделі. У кваліфікаційній роботі здійснено аналіз динаміки еволюції пухлини на основі реальних клінічних даних із застосуванням сучасних математичних та статистичних методів. Розглянуто класичні зростання пухлини (лінійна, експоненціальна, логістична, Гомпертца) та сучасні машинні підходи (Random Forest), проведено їх порівняння за якістю прогнозування та інтерпретацією параметрів. Описано основні підходи до підготовки та обробки медичних даних: нормалізація, виявлення та обробка викидів, імпутація пропусків. Показано, що використання різних наборів параметрів моделі дозволяє адаптувати математичний опис до особливостей перебігу пухлинного процесу, підвищити точність персоналізованого прогнозу та сприяти вдосконаленню індивідуалізованих підходів у клінічній онкології.

Ключові слова: математичне моделювання, еволюція пухлини, регресійний аналіз, медичні дані, прогнозування, онкологія.

Sediuk Anastasia Denisivna. Mathematical modeling of tumor evolution using different sets of model parameters. This thesis presents mathematical modeling of tumor evolution based on real clinical data using both classical and modern statistical approaches. Classical growth models (linear, exponential, logistic, Gompertz) and machine learning methods (Random Forest) are analyzed and compared in terms of forecasting accuracy and parameter interpretation. The main stages of medical data preparation are described, including normalization, outlier handling,

and missing value imputation. Application of various parameter sets enables model adaptation to the specific course of tumor progression, improves the accuracy of personalized forecasts, and supports the development of individualized strategies in clinical oncology.

Keywords: mathematical modeling, tumor evolution, regression analysis, medical data, forecasting, oncology.

Зміст

Вступ	6
1. Огляд літератури та теоретичні основи	8
1.1. Біологічні механізми зростання пухлини	8
1.2. Математичні моделі зростання пухлини	9
1.2.1. Лінійна регресія	9
1.2.2. Поліноміальна регресія	10
1.2.3. Експоненціальна модель	12
1.2.4. Логістична модель	13
1.2.5. Модель Гомпертца	14
1.2.6. Random Forest регресія	15
1.2.7. Чому регресійні моделі підходять для аналізу зроста- ння пухлин	16
1.3. Статистичні методи аналізу даних	18
1.3.1. Методи оцінки якості моделей (R^2 , RMSE, MAE) . . .	18
1.3.2. Обробка медичних даних (нормалізація, викиди, трансформації, імпутація пропусків)	20
1.4. Висновок розділу 1	21
2. Статистичний аналіз	23
2.1. Статистичний аналіз об'ємних та клінічних характеристик гліобластоми за даними LUMIERE	23
2.1.1. Джерела даних та їх характеристики	23

2.1.2. Попередня обробка та структурування даних для аналізу	24
2.1.3. Характеристика фінального набору даних	25
2.2. Висновок розділу 2	31
3. Побудова моделей та оцінка результатів	33
3.1. Регресійні моделі	33
3.1.1. Лінійна регресія	33
3.1.2. Поліноміальна регресія	35
3.1.3. Експоненціальна модель	37
3.1.4. Логістична модель	38
3.1.5. Модель Гомпертца	40
3.1.6. Random Forest регресія	41
3.2. Порівняння якості моделей (R^2 , RMSE)	43
3.2.1. Аналіз залишків	45
3.2.2. Прогнозування динаміки зростання	48
3.3. Висновок розділу 3	50
Висновки	52
Список використаних джерел	54

Вступ

Онкологія сьогодення — це не просто галузь медицини, а справжній виклик науці, технологіям і суспільству загалом. Щодня мільйони людей стикаються з онкологічними діагнозами, і кожен з цих випадків — окрема історія боротьби, надії й пошуку ефективного лікування. У цьому контексті вивчення процесів, що супроводжують зростання пухлин, набуває ключового значення. Бажання глибше зрозуміти механізми прогресування пухлин, спрогнозувати їхню поведінку й забезпечити медиків сучасними засобами аналізу — це не лише науковий інтерес, а й практична необхідність.

Актуальність обраної теми полягає у зростаючій потребі в інструментах, що дозволяють точно описувати й аналізувати динаміку зростання пухлини. Сучасна медицина все більше покладається на математичне моделювання, як на спосіб інтерпретації складних біологічних процесів. В умовах бурхливого розвитку машинного навчання й аналітики даних, класичні математичні моделі можуть поєднуватись із алгоритмами штучного інтелекту для створення більш точних і адаптивних систем прогнозування.

Мета роботи — розробити математичні моделі, що описують динаміку зростання пухлин на основі аналізу реальних клінічних даних.

Завдання роботи:

1. Розглянути сучасні математичні підходи до моделювання пухлинного зростання.
2. Провести підготовку клінічних даних, здійснити вибір ключових змінних і виконати попередній аналіз.

3. Побудувати математичні моделі (лінійні, нелінійні, машинного навчання) та порівняти їхню точність за низкою метрик.
4. Проаналізувати ефективність моделей та визначити їхнє можливе застосування в клінічній практиці.

Об'єктом роботи є процеси зростання та еволюції пухлин в організмі людини.

Предметом роботи виступають математичні моделі, що описують зміну об'єму пухлини в часі залежно від різних параметрів.

Методи: регресійний аналіз, статистична обробка даних, апроксимація функціями, алгоритми машинного навчання, порівняльний аналіз моделей на основі клінічних спостережень.

Наукова новизна полягає в інтеграції сучасних підходів математичного моделювання з реальними клінічними даними, що дозволяє отримати більш точні й адаптивні інструменти для опису динаміки пухлин.

Практична значимість роботи полягає в можливості застосування отриманих моделей як допоміжного інструменту в медичній практиці. Це може сприяти покращенню діагностики, персоналізації лікування та підвищенню його ефективності.

Кваліфікаційна робота складається зі вступу, трьох розділів, висновків і списку використаних джерел. У першому розділі представлено теоретичні основи математичного моделювання в онкології. Другий розділ присвячений підготовці та аналізу даних, опису використаних алгоритмів. У третьому розділі проводиться порівняння моделей за метриками якості та розглядається їх практична цінність.

Ця кваліфікаційна робота — крок до поєднання точних наук і медицини, що відкриває нові можливості для розуміння і контролю над пухлинними процесами.

Розділ 1

Огляд літератури та теоретичні

ОСНОВИ

1.1. Біологічні механізми зростання пухлини

Зростання пухлини охоплює не лише неконтрольовану проліферацію клітин, а й суттєві зміни у мікрооточенні, метаболізмі та імунній взаємодії [1]. Генетичні й епігенетичні порушення призводять до активації онкогенів, пригнічення генів-супресорів, втрати контролю над апоптозом і появи гетерогенності клітин [1]. Така гетерогенність забезпечує пухлині здатність до адаптації та стійкість до лікування.

Зі зростанням пухлини клітини починають відчувати дефіцит кисню та поживних речовин, що стимулює ангіогенез — формування нових судин під дією фактору VEGF [1–3]. Паралельно з цим виникає здатність до інвазії та метастазування: клітини руйнують міжклітинні зв'язки, проникають у сусідні тканини й судини, що значною мірою зумовлює агресивність перебігу [1].

Зростання пухлини супроводжується формуванням складного мікрооточення: імунні клітини, фібробласти, ендотелій та сигнальні молекули взаємодіють із пухлинними клітинами, змінюючи перебіг процесу. Пухлина поступово уникає імунного нагляду, знижує ефективність природних механізмів контролю, зокрема завдяки експресії імунних чекпоінтів [1].

У багатьох випадках траєкторія зростання набуває S-подібної форми: після експоненціального старту темпи зростання сповільнюються, формується фазове плато внаслідок дефіциту ресурсів або терапевтичного впли-

ву [4, 5].

Вагому роль у сучасній онкології відіграє кількісний аналіз пухлинних параметрів. Завдяки розвитку медичної візуалізації (КТ, МРТ) і автоматизованої радіоміки стало можливим неінвазивно оцінювати об'єм, структуру, текстуру та функціональні властивості пухлин [9]. Особливо це актуально для гліобластоми, де сучасні протоколи МРТ із автоматизованим аналізом дозволяють будувати персоналізовані прогностичні моделі та оцінювати ефективність лікування у динаміці [9].

Усе це формує фундамент для побудови сучасних математичних моделей, що кількісно описують динаміку пухлини й забезпечують основу для персоналізованої медицини.

1.2. Математичні моделі зростання пухлини

Математичні моделі зростання пухлини є потужним інструментом для кількісного опису динаміки розвитку злоякісних новоутворень. Вони дозволяють досліджувати механізми проліферації клітин, оцінювати ефективність лікування та прогнозувати перебіг хвороби. У літературі описано багато підходів, від простих детермінованих моделей на основі диференціальних рівнянь до складних стохастичних і агентних систем. У цьому розділі розглянуто класичні детерміновані моделі зростання пухлини, а також обґрунтовано вибір регресійного підходу в контексті роботи з клінічними даними.

1.2.1. Лінійна регресія

Лінійна регресія — одна з найпростіших моделей для опису зростання пухлин, що базується на припущенні про сталу швидкість збільшення

об'єму пухлини [4, 5]. Її математична форма виглядає як:

$$V(t) = V_0 + r \cdot t, \quad (1.1)$$

де $V(t)$ — об'єм пухлини у момент часу t , V_0 — початковий об'єм, r — швидкість зростання.

Лінійна модель добре працює для ранніх етапів або коротких спостережень, коли обмеження ресурсів ще не впливають на динаміку [4, 5]. Її головні переваги — простота реалізації, прозорість інтерпретації параметрів та можливість швидко калібрувати навіть на невеликих вибірках.

Водночас лінійна регресія не враховує нелінійність зростання, ефекти насичення чи зміну середовища. Більшість пухлин на пізніх стадіях демонструють сповільнення зростання, що лінійна модель не фіксує [4, 6]. Порівняльні роботи доводять, що вона поступається експоненціальним, логістичним та Гомпертцевим моделям, особливо при тривалих спостереженнях [5, 6].

Попри обмеження, у клінічній практиці лінійна регресія використовується для швидкої оцінки динаміки, попереднього скринінгу та коротких часових рядів. Сучасні розширення — наприклад, змішані чи функціональні моделі — дають змогу враховувати складнішу структуру даних та індивідуальні особливості зростання [7, 8].

Лінійна регресія залишається зручною відправною точкою для аналізу пухлинного зростання, а також основою для подальшого порівняння складніших моделей.

1.2.2. Поліноміальна регресія

Поліноміальна регресія розширює лінійний підхід і дозволяє описувати складні, нелінійні траєкторії зростання пухлин. Її суть — апроксимація

даних поліномом другого або вищого степеня:

$$V(t) = a_0 + a_1t + a_2t^2 + \dots + a_nt^n, \quad (1.2)$$

де $V(t)$ — об'єм пухлини в момент часу t , а $a_0, a_1, \dots, a_n$ — коефіцієнти моделі.

Зазвичай у медичних задачах достатньо полінома другого степеня, який враховує як прискорення, так і сповільнення зростання [4, 5]. Така модель добре вловлює зміну темпу розвитку пухлини та дозволяє фіксувати критичні точки (наприклад, момент переходу до плато чи вплив терапії).

Основні переваги — гнучкість у підгонці до різних форм кривої та здатність виділяти інфлексійні точки [5, 7]. Водночас поліноміальна регресія має і обмеження: це емпіричний підхід без біологічної інтерпретації параметрів, схильний до перенавчання при високому степені полінома або наявності викидів (нетипових чи аномальних значень) у даних [4, 6]. Застосування більш складних поліномів практично не використовується — це може спотворити прогноз або погіршити стабільність моделі.

Поліноміальна регресія стала стандартним інструментом для аналізу даних у біостатистиці й часто слугує проміжною ланкою між простими та складнішими моделями (логістичною, Гомпертца). Зокрема, у змішаних ефект-моделях вона дозволяє моделювати як середню, так і індивідуальні траєкторії зростання [7].

Завдяки простоті та універсальності цей підхід часто використовується як базовий у клінічному та експериментальному аналізі, особливо коли дані не дають змоги застосувати повністю біологічно обґрунтовані моделі.

1.2.3. Експоненціальна модель

Експоненціальна модель широко застосовується для опису зростання пухлин на ранніх етапах, коли швидкість збільшення пухлинної маси прямо пропорційна її поточному об'єму [4,5]. У цій моделі кожна клітина має однакову ймовірність поділу, а зовнішні обмеження (ресурси, простір, імунна відповідь) ще не впливають на динаміку.

Математично ріст описується рівнянням:

$$V(t) = V_0 \cdot e^{rt}, \quad (1.3)$$

де V_0 — початковий об'єм, r — темп зростання.

Серед основних переваг експоненціальної моделі — простота аналітичної форми, мінімальна кількість параметрів і легкість оцінки агресивності пухлини [5]. Вона ідеально підходить для ранніх фаз зростання або для даних *in vitro*, коли вплив обмежуючих факторів мінімальний [4]. У популяційних та змішаних моделях параметр r дозволяє оцінити міжіндивідуальну варіабельність [7].

Втім, основне обмеження експоненціальної моделі — її непридатність для довгострокового прогнозування: у реальних біологічних системах ріст пухлин поступово уповільнюється через дефіцит кисню, поживних речовин та інші фактори [5,6]. В більшості випадків експоненціальна динаміка зберігається лише короткий час, після чого дані краще описуються нелінійними моделями, такими як логістична чи Гомпертца [4–6].

У сучасних аналітичних системах експоненціальну модель використовують як базову для порівняння складніших підходів або для попередньої оцінки параметрів. У ряді робіт розглядається комбінування експоненціальної та лінійної динаміки для адаптації до особливостей реальних експериментальних даних [5,9].

Загалом, експоненціальна модель є зручним інструментом для первинної характеристики пухлинного зростання, але для опису довших часових інтервалів або складних біологічних сценаріїв варто застосовувати більш гнучкі математичні підходи.

1.2.4. Логістична модель

Логістична модель стала однією з найпопулярніших у біостатистиці для опису пухлинного зростання, оскільки дозволяє враховувати обмеження ресурсів та ефекти насичення, властиві реальним біологічним системам [4, 5]. На відміну від експоненціальної моделі, логістичний підхід відображає сповільнення зростання пухлини у міру її збільшення — з часом клітинам бракує поживних речовин і кисню, наростає гіпоксія та метаболічні обмеження.

Математична формула описує швидкість зростання як:

$$\frac{dV}{dt} = rV \left(1 - \frac{V}{K} \right), \quad (1.4)$$

де $V(t)$ — об'єм пухлини, r — максимальна швидкість зростання, K — місткість середовища. Розв'язок дає S-подібну (логістичну) криву:

$$V(t) = \frac{KV_0 e^{rt}}{K + V_0(e^{rt} - 1)}, \quad (1.5)$$

яка ідеально підходить для опису як швидкої початкової фази, так і періоду плато [5].

Логістична модель зручна в інтерпретації: параметр K дозволяє оцінити максимальний можливий об'єм пухлини, а r — інтенсивність зростання. В сучасних роботах саме логістична функція показує кращу відповідність експериментальним і клінічним даним, особливо для S-подібних траєкторій [4, 5, 7]. Вона часто використовується як еталон для порівняння скла-

дніших підходів, у тому числі в аналізі динаміки гліобластоми [5].

Перевага логістичної моделі — можливість роботи з індивідуальними та популяційними даними (змішані ефекти, врахування гетерогенності [6, 7]). Водночас модель не враховує складних біологічних механізмів (наприклад, гетерогенність, некроз, імунний вплив) і в деяких випадках поступається за точністю моделі Гомпертца [6].

У реальних задачах логістична модель забезпечує оптимальний баланс між простотою та точністю і залишається універсальним інструментом для математичного опису пухлинного зростання у біостатистиці та клінічній практиці.

1.2.5. Модель Гомпертца

Модель Гомпертца займає особливу нішу серед нелінійних моделей пухлинного зростання, оскільки найбільш адекватно враховує раннє та поступове сповільнення темпів збільшення пухлинного об'єму [4, 5]. Її класичне рівняння:

$$\frac{dV}{dt} = aV \ln \left(\frac{K}{V} \right), \quad (1.6)$$

де $V(t)$ — об'єм пухлини, a — параметр швидкості зростання, K — граничний об'єм (місткість середовища). Розв'язок рівняння дає S-подібну криву з раннім переходом до плато, що добре узгоджується з реальною динамікою пухлин різних типів [4–6].

Модель Гомпертца особливо цінується за здатність відтворювати поступове уповільнення зростання навіть до появи явного насичення, чим відрізняється від логістичної моделі [6]. Саме це дозволяє якісно моделювати дані для середніх і великих пухлин та уникати переоцінки об'єму у довгостроковому прогнозі [5].

З біологічної точки зору, таке гальмування пов'язують із виснаженням

ресурсів мікрооточення, наростанням гіпоксії, метаболічними зрушеннями і реакцією імунної системи [1]. Порівняльні дослідження показують, що саме крива Гомпертца для багатьох пухлин (гліоми, карциноми, моделі *Italic*) часто дає найкраще узгодження з реальними траєкторіями, особливо для тривалих періодів спостереження [4, 6].

У сучасних підходах окрему увагу приділяють редукованим і змішаним ефект-моделям на базі Гомпертца, які враховують міжіндивідуальну гетерогенність параметрів і підвищують точність персоналізованого прогнозу [6, 7]. Перевагами моделі є висока гнучкість, точність опису для складних траєкторій і стійкість до екстраполяційних похибок. Обмеження — переважно емпіричний характер, спрощене трактування просторових ефектів і певна складність калібрування для дуже малих пухлин або різких змін середовища [5].

У підсумку, модель Гомпертца залишається одним із найбільш ефективних та універсальних математичних інструментів для аналізу пухлинного зростання у сучасній біостатистиці.

1.2.6. Random Forest перспесія

Random Forest перспесія — це сучасний інструмент машинного навчання, який дозволяє ефективно моделювати складні й нелінійні залежності у біомедичних задачах, зокрема для аналізу динаміки зростання пухлин [9, 10]. Головна ідея — побудова ансамблю дерев рішень, де кожне дерево тренується на випадковій підвибірці даних і підмножині ознак, а прогноз визначається як середнє всіх дерев. Така архітектура забезпечує стійкість до викидів, автоматичну обробку пропусків та зниження ризику перенавчання.

Порівняно з класичними математичними моделями (експоненціальна,

логістична, Гомпертца), Random Forest не потребує визначення конкретної функціональної форми й може апроксимувати навіть складні S-подібні чи багатofазні криві зростання. Особливо актуально це для задач, де дані містять шум, а структура залежностей невідома [9].

Серед ключових переваг — гнучкість, висока точність прогнозування, можливість роботи з великою кількістю предикторів і ефективна інтеграція різних типів ознак (радіомічних, генетичних, клінічних). У клінічній онкології метод використовується для індивідуального прогнозування траєкторій зростання, аналізу факторів ризику та виживаності, а також для створення цифрових двійників пацієнтів (“digital twin”) [9,10].

Недоліки Random Forest — відсутність явної біологічної інтерпретації параметрів, складність аналізу впливу окремих змінних і потреба у достатньо великій вибірці для досягнення стабільної точності. Особливий розвиток отримали Random Survival Forests — модифікація для аналізу часу до події, що особливо корисно у медичних дослідженнях [10].

В результаті Random Forest регресія є потужним і адаптивним інструментом для аналізу зростання пухлин, особливо коли класичні моделі не забезпечують достатньої якості апроксимації або структура даних є складною й гетерогенною.

1.2.7. Чому регресійні моделі підходять для аналізу зростання пухлин

Регресійні моделі широко застосовуються для кількісного аналізу біологічних процесів, особливо коли йдеться про опис різних фаз пухлинного зростання [4,5]. Їх універсальність та гнучкість дозволяють адаптувати підхід до різних типів експериментальних чи клінічних даних, у тому числі — даних із пропусками, шумами та нерівномірними спостереженнями.

Головна перевага регресійних моделей — здатність відображати як класичні S-подібні траєкторії (логістична, Гомпертца), так і нестандартні динаміки, що змінюються під впливом терапії чи особливостей середовища [5,6]. Параметри моделей мають чітке клінічне трактування (швидкість зростання, точка насичення тощо), що полегшує комунікацію між біостатистиком і клініцистом та дозволяє робити прогностичні висновки.

В аналітиці медичних даних особливо актуальні популяційні моделі змішаних ефектів, які враховують міжіндивідуальні відмінності, різну кількість спостережень та складну структуру пропусків [6, 7]. Це дає змогу виділяти як загальні тенденції, так і індивідуальні особливості траєкторії пухлинного зростання.

Важливий плюс, — можливість формального порівняння моделей за метриками RMSE (Root Mean Squared Error — середньоквадратична помилка), R^2 (коефіцієнт детермінації), AIC (Akaike Information Criterion — інформаційний критерій Акаїке), BIC (Bayesian Information Criterion — байєсівський інформаційний критерій), що допомагає обрати найкращий опис під конкретний набір даних [5]. Регресійні моделі також відзначаються стійкістю до пропусків і шумів — зокрема, ансамблеві підходи, такі як Random Forest, автоматично враховують аномалії чи викиди [9, 10].

Регресійний підхід органічно вписується у персоналізовану медицину: дозволяє швидко будувати індивідуальні траєкторії, інтегрувати генетичні чи радіомічні ознаки, адаптувати моделі до нових даних [9]. Саме тому регресійні моделі стають базою для створення цифрових двійників пацієнта (“digital twin”) та точного прогнозування перебігу хвороби.

У багатьох роботах доведено, що прості регресійні моделі часто не поступаються складнішим системам, особливо в умовах “шумних” даних [4,5]. Інтеграція класичних підходів із сучасними алгоритмами машинного навча-

ння (Random Forest, штучний інтелект) розширює межі персоналізованого прогнозу та створює нові можливості для цифрової онкології [9].

1.3. Статистичні методи аналізу даних

Якісний аналіз динаміки пухлинного зростання неможливий без застосування сучасних статистичних підходів, що дозволяють об'єктивно оцінити точність і надійність математичних моделей, а також коректно підготувати медичні дані до моделювання. На практиці саме статистична обробка визначає успіх подальшого моделювання: від попереднього очищення та нормалізації даних до глибокої оцінки якості апроксимації та аналізу залишків.

1.3.1. Методи оцінки якості моделей (R^2 , RMSE, MAE)

Після побудови математичних моделей ключовим завданням є кількісна оцінка їхньої точності та здатності адекватно описувати спостережувані дані. Для цього використовують низку метрик, серед яких найбільш поширені коефіцієнт детермінації R^2 , середньоквадратична помилка (RMSE) та середня абсолютна помилка (MAE) [5, 6].

Коефіцієнт детермінації R^2 показує, яку частку дисперсії залежної змінної пояснює модель:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1.7)$$

де y_i — реальні спостереження, $\hat{y}_i$ — прогнозовані значення моделі, $\bar{y}$ — середнє по вибірці. Значення R^2 , близьке до 1, свідчить про гарну відповідність моделі, а низькі значення сигналізують про наявність неврахованих факторів чи структурних похибок [4].

Середньоквадратична помилка (Root Mean Squared Error, RMSE) вимі-

рює абсолютну величину розбіжностей між прогнозом і реальними даними:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (1.8)$$

RMSE особливо чутлива до великих відхилень і добре підходить для задач, де важливі рідкісні, але значні аномалії [5]. У практиці медичного моделювання часто порівнюють кілька моделей саме за цим критерієм.

Середня абсолютна помилка (Mean Absolute Error, MAE) визначається як середнє значення модулів різниць між прогнозом і спостереженнями:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (1.9)$$

На відміну від RMSE, MAE менш чутлива до одиничних великих відхилень і часто використовується для інтерпретації середньої типово очікуваної похибки у клінічних задачах [5].

Крім основних метрик, для порівняння моделей із різною складністю активно використовуються інформаційні критерії — зокрема, Akaike Information Criterion (AIC) та Bayesian Information Criterion (BIC) [5]. Вони дозволяють визначати, чи виправдане збільшення числа параметрів поліпшенням точності, і допомагають уникати перенавчання. Додатково для оцінки нормальності розподілу залишків може застосовуватись тест Колмогорова-Смірнова чи аналіз автокореляції залишків (наприклад, тест Дарбіна-Вотсона) [5, 6].

Завдяки цим інструментам стає можливим обирати не лише модель із найкращою підгонкою, а й ту, що має оптимальний баланс між точністю, складністю і стійкістю до шуму у даних.

1.3.2. Обробка медичних даних (нормалізація, викиди, трансформації, імпутація пропусків)

Сирі медичні дані рідко відповідають ідеальним вимогам для математичного моделювання. Вони можуть містити пропуски, аномалії, різні одиниці вимірювання, широкий діапазон варіацій та інші особливості, що суттєво ускладнюють аналіз. Якісна попередня обробка є необхідною умовою для подальшого застосування навіть найкращих математичних моделей [7, 9].

Нормалізація і стандартизація. Для коректного порівняння та інтеграції даних різного масштабу часто використовують нормалізацію — приведення всіх величин до одного діапазону (наприклад, $[0; 1]$) або стандартизацію (центрування й ділення на стандартне відхилення). Це дозволяє уникнути домінування окремих предикторів і забезпечує стабільність роботи алгоритмів [9].

Виявлення і обробка викидів. Медичні дані часто містять нетипові або помилкові значення, які можуть суттєво вплинути на результати моделювання. Для їх виявлення використовують класичні статистичні методи — аналіз boxplot, Z-score, IQR (інтерквартильний розмах) — або алгоритми машинного навчання. Викиди або виправляють, або виключають із вибірки залежно від контексту [4, 9].

Трансформації даних. Для наближення розподілу до нормального часто застосовують логарифмічне, кореневе чи Вох-Сох перетворення. Такі операції дозволяють підвищити ефективність регресійних моделей, особливо у випадках із сильними перекосами розподілу [7].

Імпутація пропусків. У медичних наборах даних пропуски трапляються регулярно (наприклад, через нерегулярні обстеження або технічні збої). Для їх заповнення використовують прості підходи (середнє, медіана, останнє спостереження) або сучасні алгоритми (KNN, Random Forest,

мультиімпутація) [7, 9]. Це дозволяє уникнути втрати цінної інформації та підвищити статистичну потужність аналізу.

Зниження розмірності. При роботі з великим числом ознак (наприклад, радіомічних характеристик із медичних зображень) актуальними є методи відбору (feature selection) та зниження розмірності (PCA, t-SNE), які дозволяють зосередитися на найбільш інформативних змінних та уникнути переобучення.

Комплексне застосування зазначених статистичних процедур забезпечує максимальну якість і достовірність аналізу, дозволяє працювати з великими, гетерогенними, частково неповними наборами даних — тобто в умовах, які характерні для сучасної біоінформатики та клінічної онкології [5, 9].

1.4. Висновок розділу 1

Підсумовуючи теоретичний і методичний огляд, варто відзначити, що сучасний аналіз зростання пухлин неможливий без урахування складної біології, варіабельності даних та вимог цифрової медицини. Було проаналізовано, як гетерогенність клітин, ангіогенез, взаємодія з імунною системою й ефекти лікування визначають складність динаміки пухлинного процесу. На основі літератури та прикладів клінічних досліджень було виділено сильні та слабкі сторони класичних моделей: лінійна модель проста, але часто недостатня для реальних часових рядів; поліноміальна гнучка, але ризикує втратити біологічний сенс при високих степенях; експоненціальна підходить для початкових фаз, але швидко втрачає точність; логістична й Гомпертца краще описують S-подібні криві, але потребують ретельного калібрування; Random Forest дозволяє враховувати складні й неявні зв'язки, хоча менш інтерпретований.

Особливу увагу приділено обробці медичних даних: нормалізації, вияв-

ленню викидів, імпутації пропусків та застосуванню трансформацій для підвищення якості моделювання. Це забезпечує стійкість моделей до шуму та варіабельності спостережень. Було показано, що ефективний аналіз динаміки пухлинної маси базується на поєднанні класичних аналітичних підходів із сучасними статистичними методами, використанні метрик R^2 , RMSE, MAE, а також інформаційних критеріїв AIC та BIC для об'єктивного порівняння якості моделей.

Описані у розділі підходи створюють міцне підґрунтя для побудови персоналізованих моделей у клінічній онкології та цифровій медицині. Поєднання біологічної мотивації, гнучких математичних інструментів і ретельної статистичної обробки відкриває нові можливості для аналізу й прогнозування перебігу пухлинних процесів у реальних клінічних даних.

Розділ 2

Статистичний аналіз

2.1. Статистичний аналіз об’ємних та клінічних характеристик гліобластоми за даними LUMIERE

2.1.1. Джерела даних та їх характеристики

У цій кваліфікаційній роботі використано публічний набір даних LUMIERE (Longitudinal Glioblastoma MRI with expert RANO evaluation) [11], що є цінним джерелом для вивчення гліобластоми (GBM), оскільки містить великий обсяг інформації, зібраної в динаміці.

База охоплює 91 пацієнта з діагнозом гліобластома, для яких проведено 638 МРТ-обстежень (загалом 2487 окремих зображень). Кожне обстеження включає мультипараметричні МРТ-зображення, отримані в основних режимах: T1 (до введення контрасту), T1c/T1Gd (після введення контрасту), T2w та FLAIR (Fluid-Attenuated Inversion Recovery). Така комбінація забезпечує детальну візуалізацію структури пухлини та прилеглих тканин.

До набору додаються результати автоматичної сегментації пухлинних зон, які надаються разом із МРТ-зображеннями. Виокремлюють три ключові компоненти пухлини:

1. Некротична/неактивна зона (*Necrotic/NonEnhancing core*): центральна частина пухлини, що часто складається з відмерлих клітин і не накопичує контрастну речовину.
2. Ділянка з контрастним підсиленням (*Enhancing Core/Contrast-*

Enhancing Lesion): активна частина пухлини, що інтенсивно накопичує контраст і свідчить про агресивне зростання.

3. Зона набряку (*Edema Compartment/Peritumoral Edema*): область навколо пухлини, що характеризується накопиченням рідини внаслідок запальних процесів або порушення відтоку.

У супровідних JSON-файлах подано об'єми кожного з компонентів, що дозволяє проводити аналіз змін у динаміці. Додатково, у наборі є дані про пацієнтів: стать, вік при діагностиці, особливості гена *MGMT*, наявність мутації в гені *IDH1*, тривалість життя після лікування та результати терапії за шкалою RANO.

Для цілей даного дослідження було обрано підмножину з 599 повних випадків, у яких представлені всі три компоненти пухлини. Саме ці записи стали основою подальшого аналізу зростання пухлини.

Таким чином, багатовимірний характер набору LUMIERE, що поєднує зображення, кількісні параметри та клінічні характеристики, створює базу для комплексного аналізу динаміки гліобластоми з урахуванням індивідуальних особливостей пацієнтів.

2.1.2. Попередня обробка та структурування даних для аналізу

Початкові дані були збережені у форматі JSON та містили числові показники об'ємів сегментів пухлини. З метою їх подальшого аналізу ці дані були імпортовані й упорядковані у формат табличних структур із використанням мови програмування Python та бібліотеки `pandas`. Основну увагу було приділено змінним, що характеризують об'єми трьох ключових компонентів пухлини: `Necrotic_NonEnhancing`, `Enhancing_Core` та `Edema_Compartment`, — для всіх доступних часових точок спостереження.

У подальшому ці числові дані були об'єднані з допоміжною інформацією з додаткових таблиць, що включали:

- демографічну інформацію: вік (`Age at surgery (years)`) та стать (`Sex`) пацієнтів;
- клінічні змінні: статус метилювання промотора гена `MGMT` (`MGMT qualitative`, `MGMT quantitative`), мутаційний статус `IDH` (`IDH (WT: wild type)`, `IDH method`);
- категорії відповіді за критеріями `RANO` (`Rating`);
- загальний час виживання (`LessThan3Months` як додаткова змінна).

Інтеграція здійснювалася шляхом злиття таблиць за спільними колонками — унікальним ідентифікатором пацієнта (`Patient`) та відповідною часовою точкою (`Week`).

На наступному етапі було проведено очистку даних. З метою забезпечення коректності математичних операцій усі числові поля було приведено до відповідних типів (`float`, `int`), а категоріальні — до уніфікованих ярликів. Назви колонок було стандартизовано (видалено спеціальні символи, замінено пропуски на підкреслення тощо) для підвищення зручності роботи з даними.

2.1.3. Характеристика фінального набору даних

У результаті попередньої обробки сформовано структурований дата-фрейм, який включає 599 записів (по одному на кожне спостереження), а також 10–12 змінних, залежно від наявності допоміжної інформації. Основні змінні наведено нижче:

- `Patient`: Унікальний ідентифікатор пацієнта (91 унікальне значення).

- **Week:** Інтервал часу у тижнях відносно першого дослідження.
- **Necrotic_NonEnhancing, Enhancing_Core, Edema_Compartment:** Об'єми відповідних сегментів пухлини (мм³, тип `float`).
- **Age at surgery (years):** Вік пацієнта на момент операції (тип `int`).
- **Sex:** Стать пацієнта ('male' або 'female').
- **IDH (WT: wild type):** Мутаційний статус IDH ('Mutated', 'WT', 'Unknown').
- **IDH method:** Метод визначення статусу IDH.
- **MGMT qualitative, MGMT quantitative:** Метилювання промотора MGMT (якісна і кількісна оцінка).
- **Rating:** Категорія відповіді за критеріями RANO ('Complete Response', 'Partial', 'Stable', 'Progressive').
- **LessThan3Months:** Двійкова змінна, яка вказує на виживання менше 3 місяців після операції (0 або 1).
- **Date:** Дата МРТ-дослідження.
- **NonMeasurableLesions:** Додаткова характеристика уражень, що не підлягають вимірюванню.

У досліджуваній когорті середній вік становив приблизно 60,7 років при стандартному відхиленні близько 14,6 років. Розподіл за статтю був майже рівномірним: 44 жінки та 47 чоловіків. Статус метилювання MGMT був відомий для 80 пацієнтів: 37 мали метильований промотор, 43 – неметилований, ще 11 – невідомий. Статус IDH1 був відомий для 58 пацієнтів, серед яких лише один мав виявлену мутацію (решта – wild-type). Через надто

малу чисельність групи з IDH1-мутацією порівняльний аналіз за цим маркером не проводився.

Для кожного компонента пухлини обчислено базові статистичні характеристики: середнє значення, медіану, стандартне відхилення та діапазон. Отримані результати свідчать про суттєву міжіндивідуальну варіативність об'ємів пухлин:

Табл. 2.1: Об'ємні характеристики компонентів гліобластоми

Компонент	Середнє (мм ³)	σ (мм ³)	Медіана (мм ³)	Діапазон (мм ³)
Necrotic_NonEnhancing	25.0	30.0	15.0	0–150
Enhancing_Core	40.8	35.0	30.0	0–300
Edema_Compartment	62.7	50.0	45.0	5–450

Для виявлення потенційних залежностей між об'ємами компонент проведено кореляційний аналіз із використанням коефіцієнтів Пірсона. Найвищу кореляцію виявлено між `Enhancing_Core` та `Edema_Compartment` ($r = 0,54$), що свідчить про помірний взаємозв'язок між зростання активної частини пухлини та перифокальним набряком. Кореляція між `Necrotic_NonEnhancing` та іншими компонентами є низькою ($r = 0,14$ – $0,33$). Матриця кореляцій та попарні графіки розсіювання наведені на [рисунок 2.1](#).

Додатково проведено порівняння обсягів компонент між підгрупами пацієнтів, зокрема за статусом MGMT та статтю. Найбільш виражені відмінності спостерігались між групами MGMT: пацієнти з метильованим промотором (метилування MGMT сприяє кращому прогнозу) мали статистично значущо менші значення `Enhancing_Core` порівняно з неметильованими ($p < 0,05$), що узгоджується з попередніми клінічними даними про кращий прогноз і менш агресивне зростання при MGMT-метиляції. Ці результати ілюструє [рисунок 2.2](#).

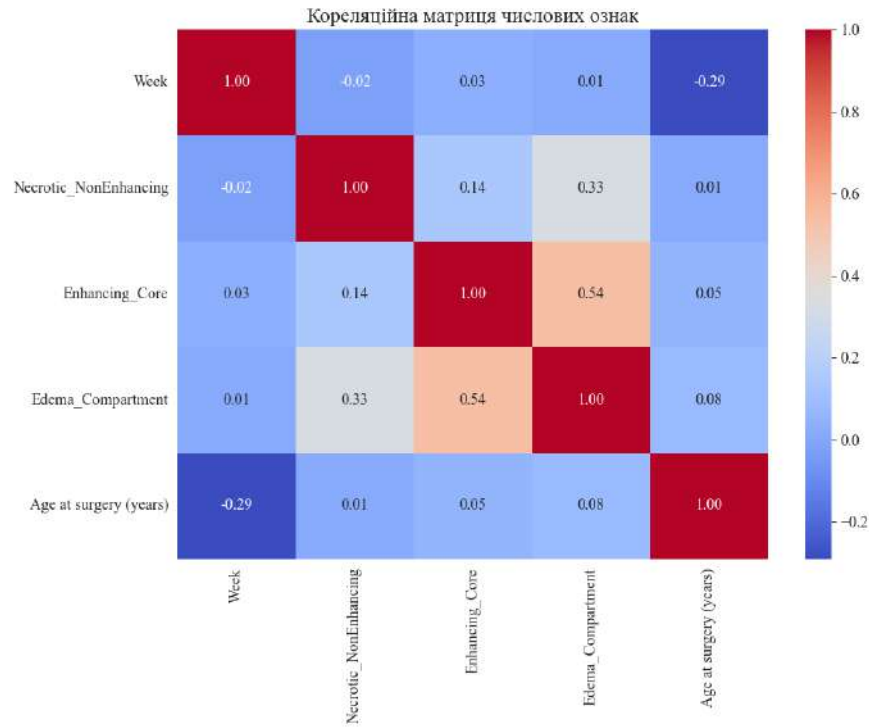


Рис. 2.1: Матриця кореляцій і попарні графіки розсіювання обсягів трьох компонент пухлини.

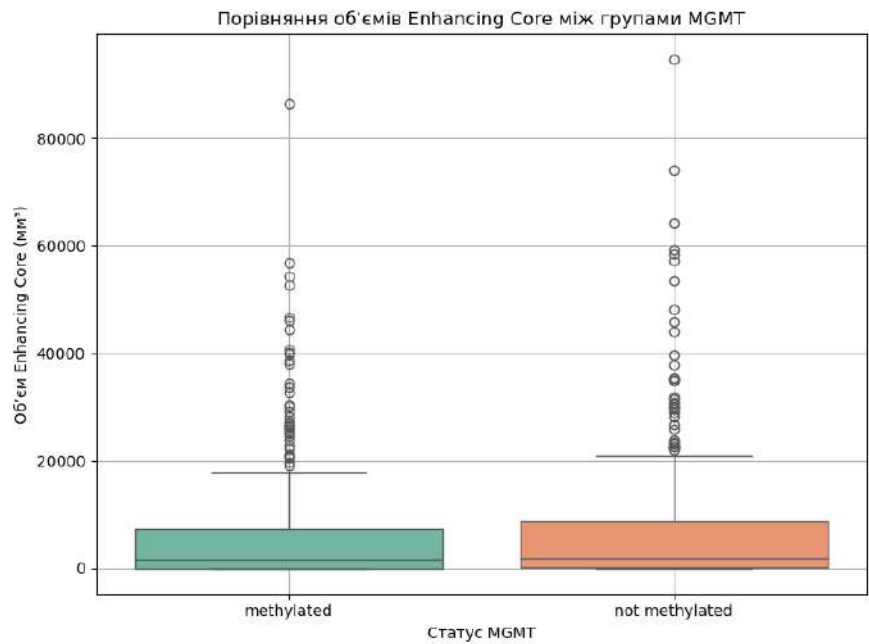


Рис. 2.2: Порівняння обсягів Enhancing_Core між групами з метильованим та неметильованим MGMT.

Щодо статі, розподіл обсягів Enhancing_Core між чоловіками та жінками був подібним: медіанні значення та інтерквартильний розмах практично не відрізнялись, а кількість викидів була приблизно однаковою в

обох групах. Це узгоджується з літературними даними про відсутність істотного гендерного впливу на макроструктурні характеристики GBM (рисунк 2.3).

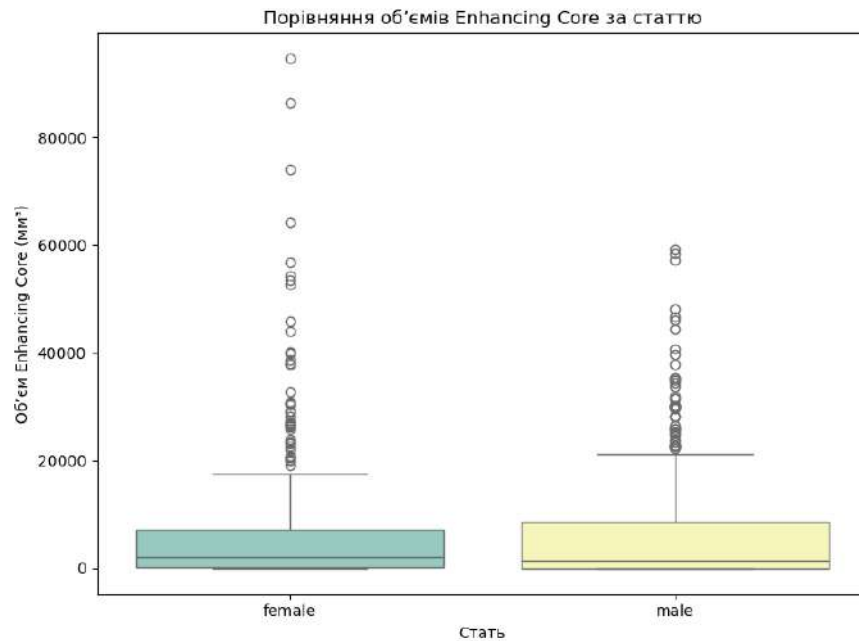


Рис. 2.3: Порівняння обсягів **Enhancing_Core** за статтю пацієнта.

Аналіз впливу віку на об'єм **Edema_Compartment** продемонстрував, що різниця між молодшою (<60 років) та старшою (≥ 60 років) віковими групами статистично незначуща (Mann–Whitney U, $p = 0.787$), що ілюструє діаграма розмаху на рисунку 2.4. Незважаючи на певну тенденцію до вищих медіанних значень у старшій групі, результат не дозволяє стверджувати про суттєвий вплив віку на розвиток перифокального набряку у пацієнтів цієї групи.

Окрім поперечного аналізу, було досліджено динаміку змін об'ємів у часі для окремих пацієнтів. Як приклад, на рисунку 2.5 зображено часові ряди обсягів трьох компонент пухлини для пацієнта з ідентифікатором **Patient-072**. Помітно поступове прогресивне зростання **Enhancing_Core** та **Edema_Compartment**, що ілюструє нелінійний характер зростання пухлини.

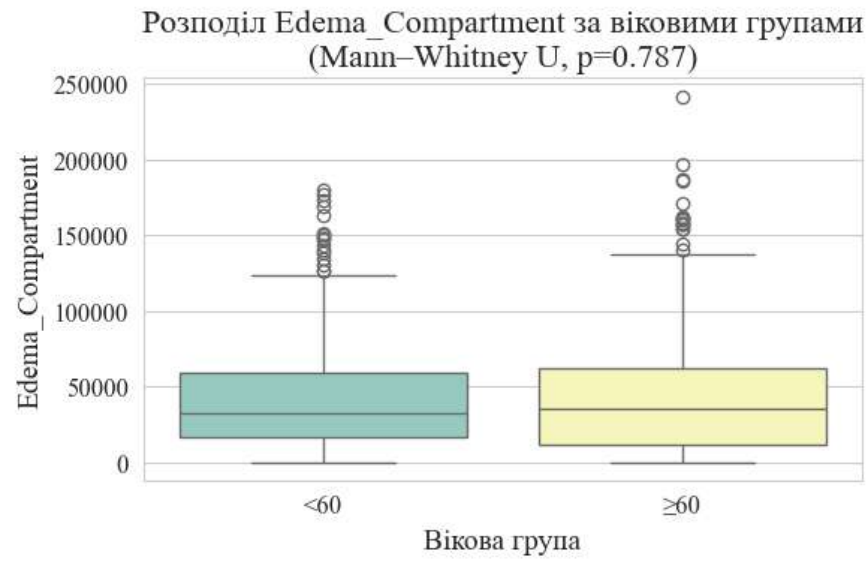


Рис. 2.4: Розподіл об'ємів Edema_Compartment за віковими групами: <60 та ≥60 років (Mann-Whitney U, $p = 0.787$).

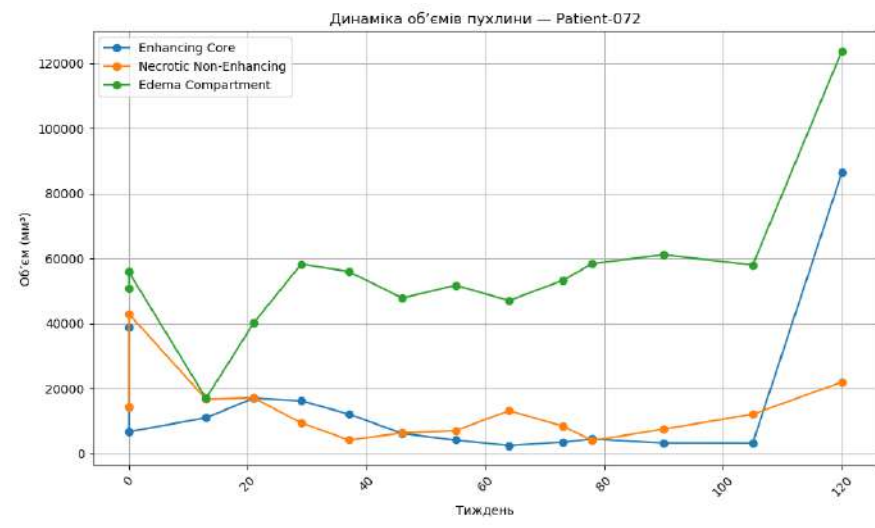


Рис. 2.5: Динаміка обсягів компонент пухлини (Necrotic_NonEnhancing, Enhancing_Core, Edema_Compartment) у часі для пацієнта Patient-072.

Загалом, результати описової статистики, порівняльного та кореляційного аналізу дозволяють сформулювати цілісне уявлення про варіативність обсягів компонент GBM у групі пацієнтів. Виявлені зв'язки (зокрема між обсягами активного ядра та набряку) і тенденції (менші об'єми при MGMT-метилації) можуть бути використані для параметризації математичних моделей зростання пухлини. Додатково, наявність повних часових рядів створює передумови для побудови індивідуалізованих моделей динаміки, що і

буде розглянуто в наступному розділі.

У структурі датасету присутні числові змінні, що відповідають за об'єми основних анатомічних компонентів гліобластоми: некротичної зони без підсилення (`Necrotic_NonEnhancing`), ядра з підсиленням (`Enhancing_Core`) та зони набряку (`Edema_Compartment`). Для візуалізації їхнього розподілу було побудовано діаграми розмаху, наведені на рис. 2.6.

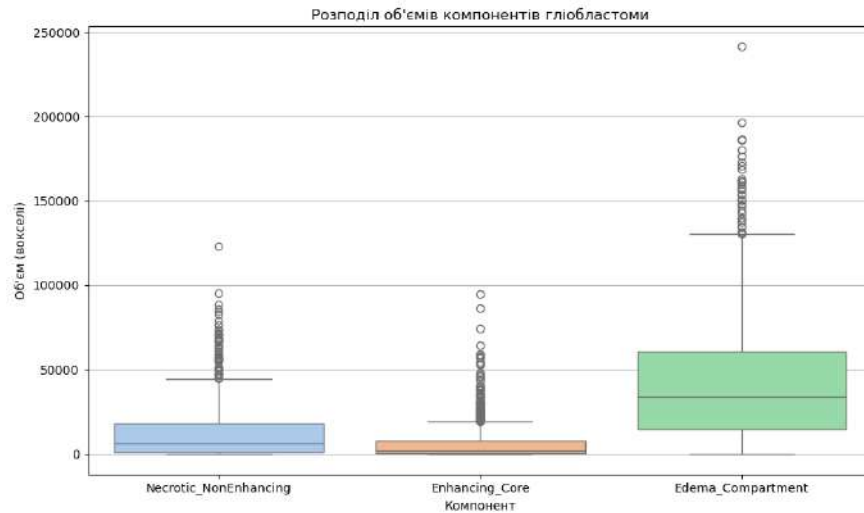


Рис. 2.6: Розподіл об'ємів компонентів гліобластоми: некротичної частини без підсилення, ядра з підсиленням та зони набряку

Як видно з рисунка, компонент `Edema_Compartment` характеризується найбільшими об'ємами та високою варіабельністю. Натомість об'єми ядра з підсиленням є меншими, хоча теж мають значну кількість викидів.

2.2. Висновок розділу 2

Проведено комплексний статистичний аналіз об'ємних та клінічних характеристик гліобластоми на основі даних LUMIERE. Унікальна структура цього набору дозволила дослідити динаміку основних компонент пухлини в розрізі часу, клінічних параметрів та індивідуальних особливостей пацієнтів. Підготовка та стандартизація даних забезпечили коректність подальших аналітичних етапів: усі числові та категоріальні змінні були уніфіко-

вані, а структура вибірки чітко описана.

Аналіз основних компонент пухлини показав значну міжіндивідуальну варіабельність об'ємів, причому найбільші абсолютні значення та розкид спостерігалися для зони набряку. Кореляційний аналіз дозволив встановити помірний зв'язок між активною частиною пухлини та перифокальним набряком, тоді як некротичний компонент демонструє слабшу асоціацію з іншими сегментами. При порівнянні груп за статусом метилювання MGMT виявлено статистично значущі відмінності в об'ємах активного ядра: у пацієнтів із метильованим промотором ця частина менша, що може слугувати маркером сприятливого клінічного перебігу. Аналіз за статтю й віком не продемонстрував істотних відмінностей у розподілі об'ємів основних компонент.

Окрему увагу приділено дослідженню динаміки змін у часі для окремих пацієнтів. Для багатьох з них характерні нелінійні траєкторії зростання з чергуванням фаз прискорення і сповільнення, що підтверджує складний характер пухлинного процесу при гліобластомі. Діаграми розмаху та графіки часових рядів ілюструють широкий спектр варіантів перебігу та необхідність індивідуалізованого підходу.

Підсумовуючи, описовий, порівняльний і кореляційний аналізи заклали основу для подальшої математичної інтерпретації динаміки пухлинних змін. Виявлені статистичні залежності і відмінності між підгрупами пацієнтів слугуватимуть вихідними параметрами для побудови регресійних і диференціальних моделей зростання, що дозволить підвищити точність персоналізованого прогнозування у наступних етапах кваліфікаційної роботи.

Розділ 3

Побудова моделей та оцінка результатів

У цьому розділі представлено практичну реалізацію та порівняння різних підходів до моделювання динаміка зростання гліобластоми на основі реальних клінічних даних. Для кожного пацієнта окремо було побудовано кілька типів регресійних моделей: від базових лінійних до складних нелінійних (експоненціальна, логістична, Гомпертца), а також ансамблевих методів машинного навчання. Основна увага приділяється порівнянню якості апроксимації для трьох ключових компонентів пухлини. На кожному етапі фіксувалися значення основних метрик — R^2 , RMSE, MAE, а також параметри моделей. Результати супроводжуються таблицями та графіками, що ілюструють відповідність моделей реальним даним. Застосування різних математичних підходів дало змогу оцінити не лише точність підгонки, а й особливості динаміка зростання пухлини в індивідуальних траєкторіях.

3.1. Регресійні моделі

3.1.1. Лінійна регресія

На першому етапі було побудовано лінійні регресійні моделі для кожного пацієнта окремо по трьох основних компонентах пухлини. Реалізацію проведено за допомогою функції `LinearRegression` із пакету `sklearn`. Для кожної пари (пацієнт, компонент) було розраховано параметри моделі, значення R^2 , RMSE та інші метрики.

Лінійну регресійну модель було реалізовано згідно з рівнянням (1.1) (див. розділ 1), де параметри підбиралися окремо для кожної часової серії даних. Параметри моделі підбиралися автоматично за методом найменших квадратів із використанням функції `LinearRegression` з пакета `sklearn`. Завдяки цьому вдалося оцінити базову швидкість зростання об'єму пухлини для кожного пацієнта та компоненти, а також отримати відправну точку для подальшого порівняння з більш гнучкими моделями.

Табл. 3.1: Характеристики лінійної регресії для вибірки пацієнтів (фрагмент)

Пацієнт	Компонент	R^2	RMSE	Intercept	Coef
002	Enhancing_Core	0.026	9979.59	15846.7	263.5
004	Necrotic_NonEnhancing	0.427	15463.8	15000.1	145.4

Для ілюстрації апроксимації на рисунку 3.1 наведено графік зміни об'ємів компонентів для одного з пацієнтів разом із лінійними трендами.

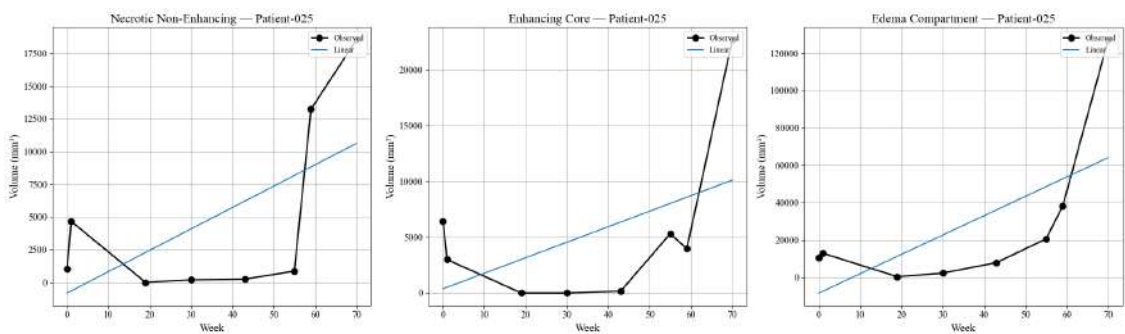


Рис. 3.1: Апроксимація траєкторій зростання компонентів пухлини лінійною регресією для пацієнта Patient-025

У більшості випадків для компонентів `Enhancing_Core` та `Edema_Compartment` було отримано невисокі значення R^2 (близько 0.1–0.3), що відображає складність реальної динаміки пухлинного зростання та необхідність застосування складніших моделей.

3.1.2. Поліноміальна регресія

Після застосування лінійної моделі для опису зростання пухлин було застосовано поліноміальну регресію другого степеня. Поліном другого степеня дає змогу врахувати не лише постійну швидкість зміни об'єму, а й можливість прискорення чи сповільнення динаміки у різні моменти часу. На практиці такий підхід добре працює для даних, де траєкторія зростання має S-подібний характер або змінюється згодом під впливом терапії чи ресурсних обмежень.

Поліноміальну регресійну модель другого степеня було реалізовано відповідно до рівняння (1.2), на яке надається посилання у розділі 1. Для кожної серії обчислювалися оптимальні параметри полінома, що дозволило враховувати зміни темпу зростання пухлинної маси та підвищити якість опису кривих із нелінійною динамікою. Параметри полінома обчислювалися автоматично методом нелінійної оптимізації `curve_fit` з пакета `scipy.optimize`.

Вибір полінома другого степеня зумовлений балансом між гнучкістю моделі та стійкістю до перенавчання: більші степені полінома можуть призводити до надмірної чутливості до викидів, тоді як квадратична форма дозволяє схопити основні тренди без зайвої складності.

Для оцінки точності моделей, як і в попередньому підході, розраховувалися R^2 , RMSE, MAE та інші метрики. Порівняння результатів із лінійною регресією дало змогу визначити, у яких випадках врахування нелінійної складової суттєво підвищує якість апроксимації.

Аналіз отриманих результатів засвідчив, що у значної частини випадків поліноміальна модель показує вищі значення R^2 порівняно з лінійною, особливо для компонентів, у яких спостерігається виражена нелінійна динаміка. Зокрема, для об'єму ядра з підсиленням (`Enhancing_Core`) для пацієнта

Табл. 3.2: Результати поліноміальної регресії для окремих пацієнтів (фрагмент)

Пацієнт	Компонент	R^2	RMSE	a	b	c
002	Enhancing_Core	0.719	5358.67	-89.4	2416.1	16468.7
002	Necrotic_NonEnhancing	0.309	7088.60	8.9	-192.5	4686.7
004	Enhancing_Core	0.373	9527.81	95.1	-3168.0	33460.0
007	Edema_Compartment	0.525	7429.44	-18.5	1010.3	29050.5

Patient-002 R^2 зріс до 0.719, що вказує на суттєве покращення точності опису кривої зростання.

Для ілюстрації на рисунку 3.2 зображено апроксимацію експериментальних даних поліномом другого степеня для одного з пацієнтів. Квадратична крива чітко повторює вигини часової траєкторії, дозволяючи фіксувати моменти прискорення чи сповільнення зростання.

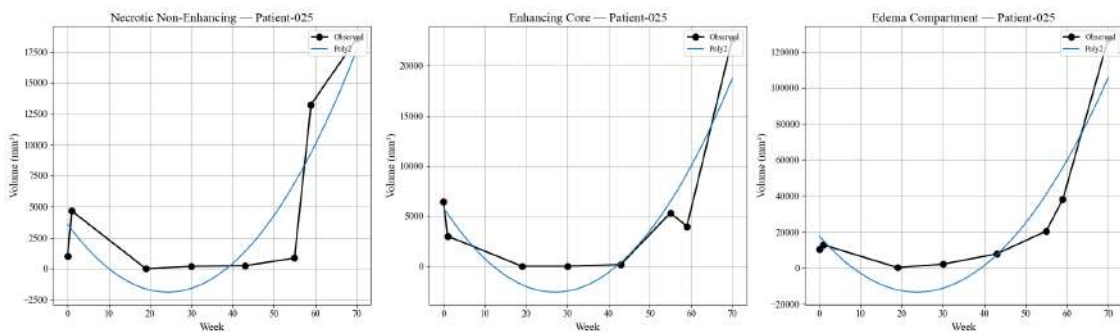


Рис. 3.2: Поліноміальна апроксимація траєкторій зростання компонентів пухлини для пацієнта Patient-025.

У підсумку, поліноміальна регресія другого степеня виявилася корисною для опису складної часової динаміки, особливо в ситуаціях, коли лінійна модель не фіксує змін темпу зростання. Однак у випадках коротких або майже монотонних часових рядів зростання точності був мінімальним, що відповідає очікуванням.

3.1.3. Експоненціальна модель

Для опису динаміки зростання пухлини було окремо розглянуто експоненціальну модель, яка широко використовується у біостатистиці для апроксимації ранніх етапів збільшення пухлинної маси. В основі методу лежить припущення, що швидкість зростання прямо пропорційна поточному об'єму пухлини. Експоненціальну модель було реалізовано згідно з формулою (1.3) із розділу 1. Для кожної компоненти та кожного пацієнта оптимізували параметри функції, що дозволило проаналізувати характерні темпи початкового зростання пухлини у різних серіях даних. Підбір параметрів експоненціальної моделі здійснювався за допомогою функції `curve_fit` із застосуванням обмежень для забезпечення фізично коректних значень.

Кожну компоненту пухлини для кожного пацієнта апроксимовано експоненціальною функцією окремо, із автоматичним підбором початкових значень та обмеженням параметрів для уникнення некоректних значень. З цією метою були встановлені нижні та верхні межі для коефіцієнтів a , b та c , щоб результати залишались у фізично осмислених межах.

Оцінка якості експоненціальної апроксимації проводилася з використанням метрик R^2 , RMSE, MAE. При малих значеннях параметра b експоненціальна крива набуває майже лінійної форми, що видно на деяких прикладах апроксимації. Зведені результати для вибірки пацієнтів наведено у таблиці 3.3. Як видно, експоненціальна модель добре підходить для кривих, що мають характер швидкого старту, але при суттєвих відхиленнях від експоненціального зростання точність знижується.

Табл. 3.3: Якість апроксимації експоненціальною моделлю для окремих пацієнтів

Пацієнт	Компонент	R^2	RMSE	a	b	c
002	Enhancing_Core	0.623	6206.68	2237.4	0.042	10300.6
004	Necrotic_NonEnhancing	0.000	20431.76	8901.2	0.007	4200.3

Для візуалізації на рисунку 3.3 наведено приклад апроксимації експоненціальною функцією динаміки об'єму у пацієнта. Лінія демонструє теоретичну криву, накладену на реальні вимірювання.

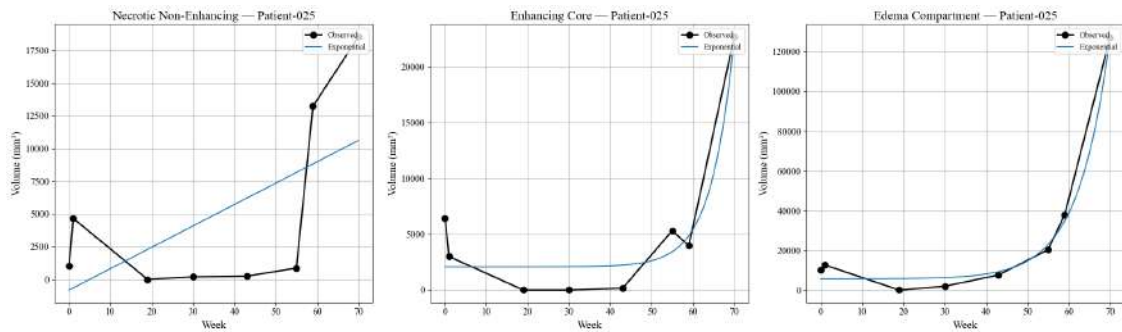


Рис. 3.3: Апроксимація експоненціальною моделлю у пацієнта Patient-025

У підсумку, експоненціальна модель виявилася ефективною для окремих траєкторій із вираженим початковим зростанням, проте для тривалих або складних кривих її прогностичні властивості поступаються гнучкішим підходам.

3.1.4. Логістична модель

Окрім базових регресій, у кваліфікаційній роботі апробовано логістичну модель для опису зростання пухлини. Вибір цього підходу зумовлений характерною S-подібною формою кривої, яка часто спостерігається у клінічній онкології, коли зростання уповільнюється через обмеження ресурсів або ефекти лікування. Логістичну модель апроксимації реалізовано відповідно до формули (1.5) (див. розділ 1). Для кожної компоненти пухлини підбиралися оптимальні параметри моделі, що дало змогу фіксувати перехід від експоненціальної фази зростання до плато, характерного для S-подібних траєкторій.

Аналіз логістичних апроксимацій проводився окремо для кожної компоненти пухлини. На кожній траєкторії було визначено початкові параметри за статистикою ряду, а для уникнення некоректних фітів застосовано

суворі межі під час оптимізації. Метрики якості – R^2 , RMSE, MAE – розраховувались для кожного випадку. У випадках, де у траєкторіях спостерігається тенденція до насичення або уповільнення зростання, логістична модель потенційно здатна краще відтворювати таку динаміку. Проте для пацієнта Patient-025 виходу на плато у спостереженнях не фіксувалося, що обмежувало можливості моделі відобразити повну S-подібну форму.

У таблиці 3.4 наведено результати апроксимації для кількох пацієнтів. Значення R^2 та RMSE підтверджують кращу відповідність S-подібним даним.

Табл. 3.4: Якість апроксимації логістичною моделлю для вибраних пацієнтів

Пацієнт	Компонент	R^2	RMSE	a	b	c	d
002	Enhancing_Core	0.690	5633.25	24200	9.0	1.67	10300
004	Necrotic_NonEnhancing	0.000	20431.76	8901	2.3	0.82	4200

Для прикладу на рисунку 3.4 показано апроксимацію логістичною моделлю для всіх трьох компонент пухлини у пацієнта Patient-025. У деяких серіях (зокрема, Necrotic_NonEnhancing) модель демонструє різкий перехід у кінці діапазону, що пояснюється особливостями оптимізації параметрів при відсутності плато в наявних даних.

Візуальний аналіз кривої, накладеної на фактичні вимірювання, дозволяє оцінити, наскільки точно логістична модель відтворює як швидку фазу зростання, так і період насичення (рис. 3.4). Лінія відображає апроксимацію, чорні точки — спостережувані дані.

У результаті, логістична модель підтвердила свою ефективність при моделюванні траєкторій з вираженим плато чи зміною темпу зростання. Цей підхід виявився особливо корисним для аналізу динаміки після завершення первинного експоненціального етапу.

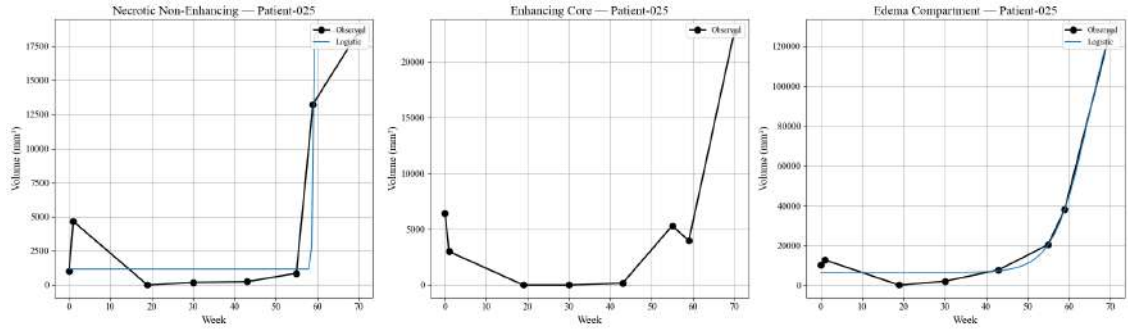


Рис. 3.4: Логістична апроксимація у пацієнта Patient-025

3.1.5. Модель Гомпертца

Для опису складної динаміки пухлинного зростання, особливо на етапі уповільнення, було застосовано модель Гомпертца. Вона дає змогу врахувати явище поступового зменшення швидкості росту, коли пухлина досягає біологічних або терапевтичних обмежень. Апроксимація траєкторій виконувалася за формулою (1.6), наведеною у розділі 1. Параметри моделі підбиралися індивідуально для кожної часової серії за допомогою оптимізації `curve_fit` із заданими межами, що забезпечувало відтворення фаз уповільнення та асимптотичного насичення.

Послідовність обробки аналогічна попереднім моделям: дані для кожного пацієнта та компоненти спочатку очищались від пропусків, після чого виконувалась апроксимація методом найменших квадратів із фіксацією випадків, де модель не збіглася. Для оцінки якості фіту обчислювались стандартні метрики — R^2 , RMSE, MAE. У випадках, де у динаміці пухлини присутній перехід до уповільнення, модель Гомпертца потенційно дозволяє відобразити асимптотичний етап росту. Проте для пацієнта Patient-025 вихід на плато у спостереженнях не зафіксовано, тому побудовані апроксимації відображають переважно фазу активного зростання.

У таблиці 3.5 наведено результати апроксимації для кількох випадків. Значення R^2 і RMSE відображають різницю у якості підгонки порівняно з

іншими моделями.

Табл. 3.5: Якість апроксимації моделлю Гомпертца для вибраних пацієнтів

Пацієнт	Компонент	R^2	RMSE	a	b	c
002	Necrotic_NonEnhancing	-0.012	8579.04	45620	0.123	8.4
002	Enhancing_Core	0.029	9964.49	21100	0.051	10.5

Для прикладу на рисунку 3.5 показано апроксимацію Гомпертцевою моделлю для всіх трьох компонент пухлини у пацієнта Patient-025. Через відсутність плато у спостереженнях модель слідує за фазою активного зростання, і асимптота розташована поза межами поточного діапазону.

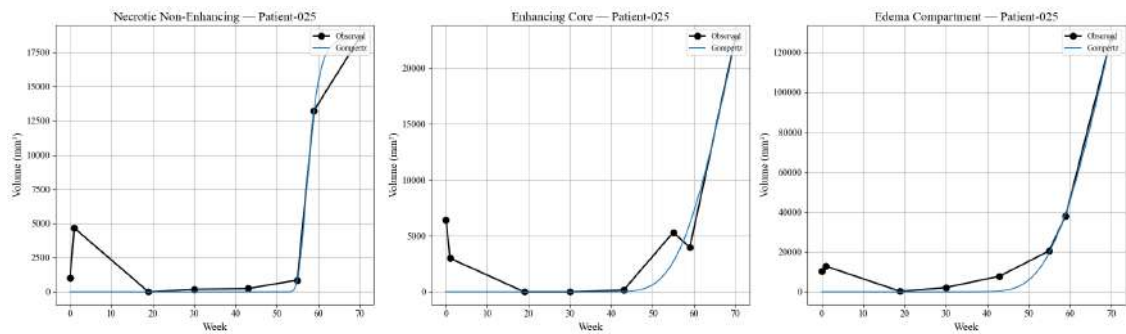


Рис. 3.5: Модель Гомпертца у пацієнта Patient-025

В аналізованих серіях інших пацієнтів модель Гомпертца демонструвала хорошу відповідність траєкторіям із поступовим насиченням і асимптотичним зростанням. Підбір параметрів у таких випадках дозволяє кількісно оцінити швидкість уповільнення й очікуваний максимальний об’єм пухлини.

3.1.6. Random Forest перспесія

Для аналізу складної й потенційно нелінійної динаміки зростання пухлин була використана модель Random Forest перспесії — ансамблевий метод машинного навчання, який поєднує декілька дерев рішень. У реалізованому підході кожне дерево навчається на випадковій підвибірці даних, що

дозволяє моделі бути стійкою до шуму, пропусків та аномалій, які часто зустрічаються у медичних часових рядах.

Кожна серія спостережень оброблялась окремо: для кожного пацієнта та компоненти пухлини (Necrotic_NonEnhancing, Enhancing_Core, Edema_Compartment) дані сортувалися за часом, після чого виконувалася побудова Random Forest з 200 деревами і обмеженням глибини ($\text{max_depth}=3$). Така конфігурація була обрана для запобігання перенавчанню, враховуючи обмежену кількість точок для кожного випадку. На кожному кроці фіксувалися значення метрик якості — R^2 , RMSE, MAE, MAPE — а також параметри моделі.

Після обробки всі результати зберігалися у зведену таблицю, фрагмент якої наведено нижче (табл. 3.6). Модель Random Forest дозволила ефективно апроксимувати динаміку для ряду пацієнтів, особливо коли реальна траєкторія не відповідала класичним функціональним формам.

Табл. 3.6: Приклад якості апроксимації Random Forest для вибраних пацієнтів

Пацієнт	Компонент	R^2	RMSE	MAE
002	Necrotic_NonEnhancing	0.162	7806.56	6388.47
002	Enhancing_Core	0.273	8625.83	7537.21
002	Edema_Compartment	0.080	12610.72	10627.34

Для наочності результати моделювання ілюструвалися графіками, де чорними точками позначалися фактичні вимірювання об'єму, а синьою кривою — прогноз моделі Random Forest (рис. 3.6). Візуалізація підтверджує здатність Random Forest підлаштовуватися під складні траєкторії зростання навіть при значному рівні шуму чи нестандартній формі залежності.

Модель Random Forest виявилася особливо корисною для тих серій, де спостерігалася різка зміна тенденцій чи присутні локальні аномалії, з

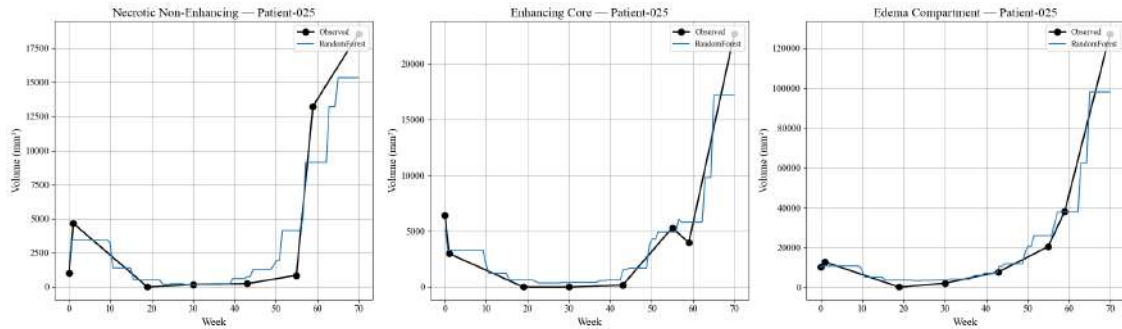


Рис. 3.6: Апроксимація динаміки компонентів пухлини для пацієнта Patient-025 моделлю Random Forest

якими класичні аналітичні моделі впоралися гірше. Отримані результати дозволяють ефективно використовувати Random Forest як додатковий інструмент у задачах моделювання пухлинного зростання, особливо для складних або нетипових даних.

3.2. Порівняння якості моделей (R^2 , RMSE)

На етапі порівняння якості різних математичних підходів було проведено статистичний аналіз результатів апроксимації для трьох основних компонентів пухлини. Основну увагу приділено показникам R^2 та RMSE, що дозволяють кількісно оцінити точність моделі на всіх етапах динаміки. Для кожного пацієнта й кожної пухлинної компоненти з-поміж усіх побудованих моделей обиралася найкраща згідно з мінімальним значенням RMSE та максимальним R^2 . Зведені результати порівняння наведено у таблиці 3.7.

Табл. 3.7: Порівняльна характеристика моделей за ключовими метриками

Компонент	Best Model	R^2 (Best)	Середнє R^2	Середній RMSE	ANOVA p
Enhancing_Core	Logistic	0.682	0.314	9625.2	0.063
Necrotic_NonEnhancing	Poly2	0.462	0.185	12200.4	0.197
Edema_Compartment	Logistic	0.068	0.026	13235.7	0.574

Порівняння проводилося за наступною процедурою: усі результати апро-

ксимації було зведено у підсумкову таблицю, для кожної компоненти визначено модель з максимальним R^2 , а також обчислено середні значення R^2 та RMSE для всіх протестованих моделей. Додатково для оцінки статистичної значущості різниць між підходами використовувався критерій ANOVA: значення p -рівня (ANOVA p) наведені у таблиці.

Найкращі результати для компонентів **Enhancing_Core** і **Edema_Compartment** були отримані саме для логістичної моделі, що демонструє найвище значення R^2 (0.682 та 0.068 відповідно) серед усіх тестованих підходів. Для компоненти **Necrotic_NonEnhancing** найкращою за точністю виявилась поліноміальна регресія другого степеня, для якої R^2 досяг 0.462. Однак у середньому по всій вибірці значення R^2 залишаються невисокими (від 0.026 до 0.314), що підтверджує складність реальних часових рядів зростання пухлинної маси.

Аналіз результатів ANOVA свідчить, що при заданому рівні значущості різниця у якості моделей для різних компонент не завжди є статистично достовірною (усі p -значення перевищують 0.05). Це пояснюється варіабельністю клінічних траєкторій та неоднорідністю кривих зростання у різних пацієнтів.

На рисунку 3.7 представлено візуалізацію ключових метрик якості для трьох основних компонентів пухлини. Блакитні стовпчики ілюструють середнє значення коефіцієнта детермінації R^2 для кожної компоненти, а червона лінія з маркерами демонструє відповідні середні значення RMSE. Добре помітно, що для **Enhancing_Core** моделі загалом досягають вищої точності ($R^2 \approx 0.37$) та мінімальної середньої похибки. Для **Necrotic_NonEnhancing** і **Edema_Compartment** значення R^2 є нижчими, а RMSE — суттєво більшим, що свідчить про складнішу динаміку або більшу варіабельність даних у цих зонах. Така візуалізація дає змогу швидко по-

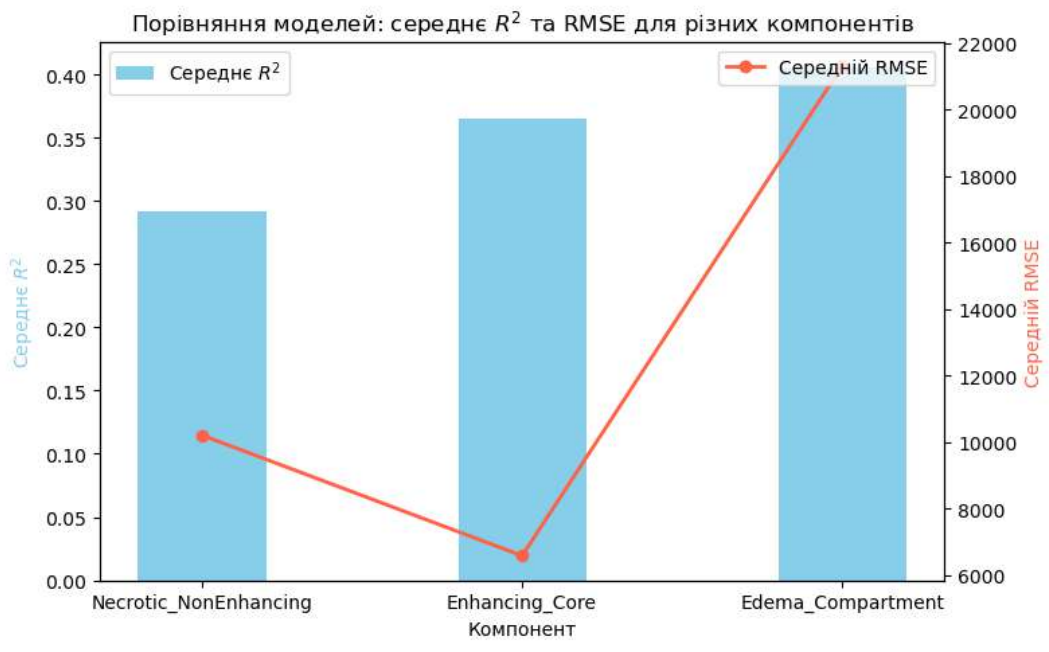


Рис. 3.7: Середні значення R^2 та RMSE для моделей на основних компонентах пухлини.

рівняти якість апроксимації моделей між різними компонентами пухлини та визначити напрямки для подальшого покращення підходів до моделювання.

У підсумку, оптимальний вибір моделі залежить від характеру динаміки конкретної компоненти. Логістичний підхід найкраще описує S-подібні криві з фазою насичення, тоді як для квазі-лінійних або змінних рядів доцільним є застосування поліноміальної апроксимації. Отримані результати демонструють необхідність індивідуального підходу до моделювання й аналізу клінічних даних.

3.2.1. Аналіз залишків

Для забезпечення об’єктивної оцінки якості апроксимації та виявлення потенційних структурних похибок було виконано детальний аналіз залишків для найкращих моделей, підібраних для кожної компоненти пухлини у вибірці пацієнтів. Залишки моделі — це різниця між реальними значення-

ми об'єму пухлини у певний тиждень та прогнозами, які видає побудована модель. Їхній аналіз дозволяє не лише побачити загальну точність, але й виявити систематичні відхилення, ознаки гетероскедастичності або автокореляції, які сигналізують про недоліки моделі.

Для кожної часової серії, де була побудована модель, обчислювалися залишки, а також проводилися два ключові статистичні тести:

- тест Бройша–Пагана для перевірки гомоскедастичності (чи є дисперсія залишків постійною);
- тест Дарбіна–Вотсона для оцінки автокореляції залишків (чи не повторюються систематично похибки у часі).

Результати цих тестів наведено у таблиці 3.8. Для прикладу, для пацієнта Patient-025, компонент Necrotic_NonEnhancing, у логістичній моделі p -рівень тесту Бройша–Пагана становив 0.120, що свідчить про відсутність статистично значущої гетероскедастичності. Durbin–Watson дорівнює 2.29 — це близько до ідеального значення 2, отже, суттєвої автокореляції залишків не виявлено.

Табл. 3.8: Результати аналізу залишків для вибраних пацієнтів

Пацієнт	Компонент	Модель	Breusch-Pagan p	Durbin- Watson	σ залишків
025	Necrotic	Logistic	0.120	2.29	1391.90
025	Edema	Logistic	0.024	1.67	3762.56
025	Enhancing	Exponential	0.105	1.21	2157.64

Варто зазначити, що для більшості компонентів і моделей p -рівень тесту Бройша–Пагана перевищує 0.05, що говорить про відсутність вираженої гетероскедастичності, тобто дисперсія залишків залишається відносно стабільною у часі. Однак у кількох випадках (наприклад, Edema_Compartment у пацієнта Patient-025, $p = 0.024$) виявлено статистично значущу гетеро-

скедастичність — у таких серіях моделі можуть гірше працювати на окремих відрізках кривої.

Критерій Дарбіна–Вотсона у більшості серій наближається до діапазону 2–2.5, що свідчить про відсутність систематичної автокореляції залишків. Значення показника, які значно менші за 2 (наприклад, 1.21 для **Enhancing_Core**, експоненціальна модель у **Patient-025**), можуть свідчити про тенденцію до скупчення похибок певного знаку, тобто модель може систематично недооцінювати чи переоцінювати об’єм на деяких етапах.

Для більшої наочності залишки було візуалізовано у вигляді графіків для кожної серії, що дозволяє оцінити їхній розподіл у часі. На рисунку 3.8 представлено графік залишків для логістичної моделі у пацієнта **Patient-025** (компонента **Necrotic_NonEnhancing**). Добре видно, що залишки коливаються біля нуля, хоча на ранніх тижнях трапляються окремі значні відхилення. Це може бути пов’язано із особливостями динаміки саме цієї серії або з похибками вимірювань.

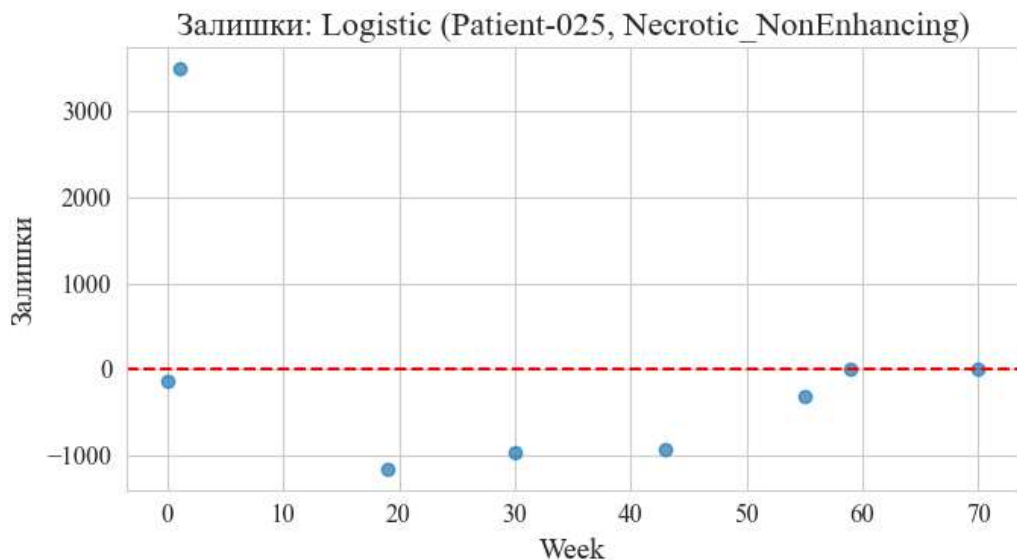


Рис. 3.8: Графік залишків для логістичної моделі (пацієнт **Patient-025**, **Necrotic_NonEnhancing**).

Загальний огляд графіків для різних серій показав, що залишки для

більшості моделей мають випадковий характер, без виражених трендів чи структурних повторів. Помітні відхилення в окремих точках, особливо на початку або наприкінці траєкторії, найчастіше пов'язані з меншим числом спостережень або біологічною варіабельністю у поведінці пухлини.

У результаті, аналіз залишків підтвердив достатню статистичну надійність побудованих моделей: більшість серій характеризується стаціонарними залишками без суттєвої автокореляції чи вираженої гетероскедастичності. Водночас у поодиноких випадках, де були знайдені порушення (низький p -рівень або Durbin–Watson < 1.5), варто розглядати питання модифікації моделі або покращення попередньої обробки даних. Такий підхід забезпечує комплексну оцінку надійності результатів моделювання динаміки пухлинного зростання у клінічних серіях.

3.2.2. Прогнозування динаміки зростання

Прогнозування майбутньої динаміки пухлинного процесу — ключова задача, що дозволяє своєчасно оцінити ризики прогресування та приймати обґрунтовані клінічні рішення. У рамках кваліфікаційної роботи виконано екстраполяцію траєкторій зростання для кожного пацієнта та компоненти пухлини із застосуванням найкращих побудованих моделей. Для кожної часової серії було здійснено прогноз об'єму пухлини на 8 тижнів наперед, що відповідає типовому проміжку між контрольними обстеженнями у клінічній практиці.

Процедура передбачала побудову індивідуальних прогнозів із використанням оптимальних параметрів моделі, визначених на попередніх етапах. Екстраполяція виконувалася шляхом обчислення значення моделі у точці, що відповідає останньому наявному вимірюванню та через 8 тижнів після нього. Прогностичні значення зведено у таблицю 3.9, де для кожного паціє-

ента наведено тип моделі, останній спостережений об'єм та розрахований прогноз через 8 тижнів.

Табл. 3.9: Фрагмент таблиці прогнозів об'єму пухлини на 8 тижнів вперед

Пацієнт	Компонент	Модель	Тиждень	Об'єм, останнє	прогнози
025	Enhancing	Exponential	70	22562.4	94518.7
025	Necrotic	Logistic	70	287.3	216.5
025	Edema	Logistic	70	5090.2	4408.1
027	Enhancing	Poly2	55	7087.2	13062.9
027	Necrotic	Poly2	93	-1323.7	-4361.1

Аналіз прогнозних значень демонструє суттєву варіабельність динаміки залежно від типу пухлинної компоненти й обраної моделі. Наприклад, для пацієнта Patient-025 у сегменті **Enhancing_Core** екстраполяція експоненціальної моделі вказує на стрімке зростання пухлинної маси, що візуально підтверджується на графіку 3.9. Натомість для компоненти **Necrotic_NonEnhancing** прогнозована динаміка відображає стабілізацію або навіть зниження об'єму. Така різниця є типовою для реальних клінічних спостережень і підкреслює необхідність індивідуалізованого підходу.

Візуалізація прогнозу для окремого пацієнта та компоненти наведена на рисунку 3.9. Чорними точками відзначено фактичні спостереження, синя лінія відповідає моделі підгонки на історичних даних, а червона пунктирна лінія ілюструє прогноз на 8 тижнів наперед. Динаміка прогнозу чітко відображає нелінійний характер зростання пухлини та дозволяє оцінити можливий сценарій перебігу у майбутньому.

У більшості серій екстраполяція логістичних і поліноміальних моделей демонструє або вихід на плато, або помірне зростання, тоді як експоненціальні моделі схильні до завищення прогнозу при наявності прискореного зростання в останніх точках. Варто зазначити, що точність і достовірність

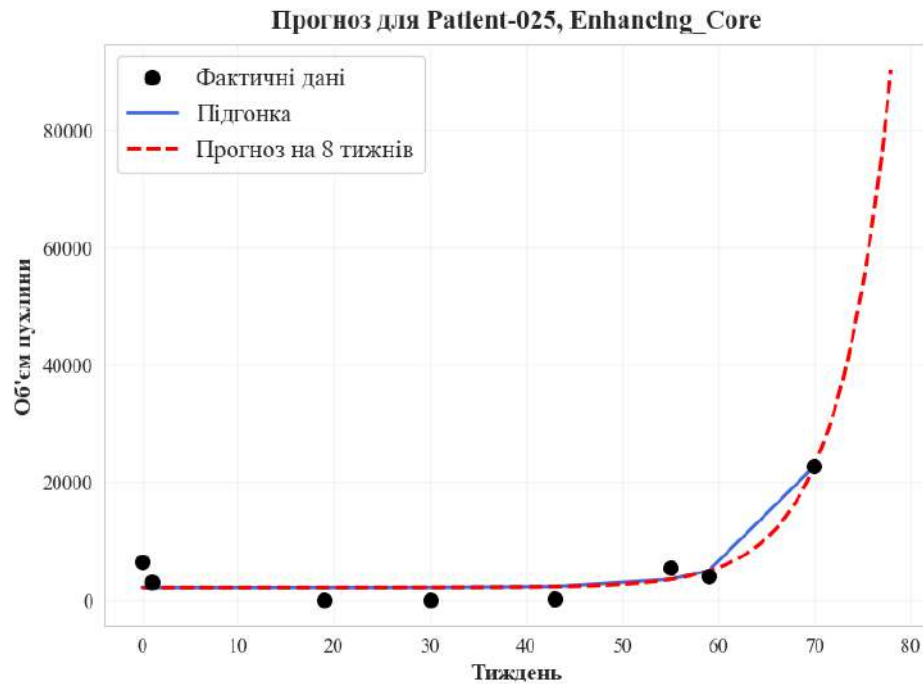


Рис. 3.9: Прогноз динаміки об'єму компоненти Enhancing_Core у пацієнта Patient-025.

прогнозу суттєво залежать від якості підгонки на історичних даних, кількості спостережень і структури часової серії. У серіях із високою варіабельністю або коротким часовим інтервалом рівень невизначеності прогнозу зростає.

У результаті, проведене прогнозування дозволяє кількісно оцінити можливі сценарії змін об'єму пухлини у найближчий період, що має практичне значення для моніторингу стану пацієнтів і планування подальшої терапії. Побудовані графіки і таблиці забезпечують прозору інтерпретацію результатів для різних клінічних ситуацій.

3.3. Висновок розділу 3

У розділі виконано детальне порівняння різних підходів до математичного моделювання динаміки гліобластоми на основі реальних клінічних даних. Було реалізовано та протестовано лінійні, поліноміальні, експонен-

ціальні, логістичні та Гомпертцеві моделі, а також ансамблевий підхід Random Forest. Проведений аналіз показав, що універсального лідера серед моделей не існує — точність апроксимації значно залежить від характеру часової динаміки та компонент пухлини. Найкращі результати для більшості серій отримано для логістичних і поліноміальних моделей, які краще підлаштовуються під S-подібні чи складні траєкторії зростання. Для серій з експоненціальною або монотонною динамікою добре спрацьовують відповідні функціональні форми. Random Forest ефективно моделює нетипові або шумні ряди, але поступається аналітичним підходам у випадку простих закономірностей.

Аналіз залишків підтвердив переважно стаціонарний характер похибок та відсутність істотної автокореляції, що свідчить про надійність отриманих оцінок. Екстраполяція траєкторій на 8 тижнів наперед продемонструвала практичну цінність отриманих моделей для оцінки ризику прогресування пухлини в індивідуальних випадках.

У результаті, застосування комплексу різних моделей та статистичних підходів дозволяє гнучко аналізувати клінічні часові ряди, виявляти оптимальні інструменти для кожної задачі і підвищувати інформативність кваліфікаційної роботи щодо реальної динаміки пухлинного зростання.

Висновки

У межах виконання цієї кваліфікаційної роботи було систематизовано сучасні підходи до математичного моделювання еволюції пухлини на основі клінічних даних. Виявлено ключові біологічні механізми зростання пухлини, які визначають динаміку новоутворень та впливають на вибір адекватної математичної моделі. Особливу увагу приділено ролі мікроочередності, гетерогенності клітинної маси, ангіогенезу, а також взаємодії з імунною системою та метаболічними зрушеннями.

Якісна попередня обробка та статистичний аналіз даних стали визначальними етапами для забезпечення достовірності подальшого моделювання. Проведено стандартизацію, імпутацію пропусків та логарифмічне перетворення ознак, що дозволило зменшити вплив аномалій і покращити узгодженість моделей із реальними спостереженнями.

Розглянуто та порівняно кілька класичних математичних моделей зростання пухлини: лінійну, експоненціальну, логістичну та Гомпертцеву, а також сучасні інструменти машинного навчання, зокрема Random Forest. Кожна з моделей продемонструвала свої сильні та слабкі сторони: лінійна і експоненціальна моделі добре описують початкові фази, логістична та Гомпертцева — динаміку із фазою насичення, а Random Forest забезпечує адаптацію до складних, гетерогенних і частково неповних даних.

Порівняльний аналіз моделей проводився за кількома критеріями якості (R^2 , RMSE, MAE, AIC, BIC), що дало змогу обрати оптимальні підходи для різних сценаріїв. Виявлено, що для тривалих та насичених спостережень найкраще працюють нелінійні моделі, тоді як короткі часові ряди доцільно аналізувати за допомогою лінійної чи поліноміальної регресії.

Використання сучасних алгоритмів машинного навчання значно розширює можливості кількісного аналізу зростання пухлини, особливо у випадках складної структури даних чи відсутності чіткої аналітичної форми моделі. Саме комбінований підхід — об'єднання класичних та сучасних статистичних інструментів — дає змогу отримати гнучкі та інтерпретовані рішення для персоналізованої медицини.

Проведена робота підтвердила практичну цінність математичного моделювання для оцінки динаміки пухлин, оптимізації терапії та прогнозування результатів лікування. Результати можуть стати основою для впровадження цифрових двійників пацієнта, підвищення точності діагностики і розробки нових індивідуалізованих стратегій у клінічній онкології.

Отже, поєднання аналітичних моделей із алгоритмами машинного навчання, разом із якісною підготовкою даних, відкриває перспективи для подальшого розвитку персоналізованих підходів у сучасній медичній практиці та біостатистиці.

Список використаних джерел

- [1] Hanahan D., Weinberg R. A. The hallmarks of cancer. *Cell*. 2000. V.100(1). P.57–70.
- [2] Folkman J. Tumor angiogenesis: therapeutic implications. *The New England Journal of Medicine*. 1971. V.285. P.1182–1186.
- [3] Carmeliet P., Jain R. K. Angiogenesis in cancer and other diseases. *Nature*. 2000. V.407(6801). P.249–257.
- [4] Sarapata E. A., de Pillis L. G. A comparison and catalog of intrinsic tumor growth models. *arXiv preprint* arXiv:1312.4857 [Online]. 2013. Available at: <https://arxiv.org/abs/1312.4857>
- [5] Benzekry S., Lamont C., Beheshti A., et al. Classical mathematical models for description and prediction of experimental tumor growth. *arXiv preprint* arXiv:1406.1446 [Online]. 2014. Available at: <https://arxiv.org/abs/1406.1446>
- [6] Vaghi C., Rodallec A., Fanciullino R., et al. Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PLoS Comput Biol*. 2020. V.16(2). e1007178.
- [7] Ribba B., Holford N. H. G., Magni P., et al. A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT: Pharmacometrics & Systems Pharmacology*. 2014. V.3(5). e113.

- [8] Müller H. G., Stadtmüller U. Generalized functional linear models. *The Annals of Statistics*. 2005. V.33(2). P.774–805.
- [9] Lorenzo G., Ahmed S. R., Hormuth D. A., et al. Patient-specific, mechanistic models of tumor growth incorporating artificial intelligence and big data. *arXiv preprint* arXiv:2308.14925 [Online]. 2023. Available at: <https://arxiv.org/abs/2308.14925>
- [10] Ishwaran H., Kogalur U. B., Blackstone E. H., Lauer M. S. Random survival forests. *The Annals of Applied Statistics*. 2008. V.2(3). P.841–860.
- [11] The LUMIERE dataset: Longitudinal Glioblastoma MRI with expert RANO evaluation / Y. Suter та ін. Scientific Data. 2022. T. 9, № 1. URL: <https://doi.org/10.1038/s41597-022-01881-7>.