МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ВОЛИНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ЛЕСІ УКРАЇНКИ
Кафедра комп'ютерних наук та кібербезпеки

На правах рукопису

ЧЖАН ХУНСЯО

ДОСЛІДЖЕННЯ ТЕХНОЛОГІЙ ФЕДЕРАТИВНОГО НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ МЕДИЧНИХ ЗОБРАЖЕНЬ

Спеціальність: 122 Комп'ютерні науки
Освітньо-професійна програма: Комп'ютерні науки та інформаційні технології
Робота на здобуття освітнього ступеня «магістр»

Науковий керівник:
МАМЧИЧ ТЕТЯНА ІВАНІВНА
канд. фіз.- мат. наук, доцент
кафедри комп'ютерних наук та кібербезпеки

РЕКОМЕНДОВАНО ДО ЗАХИСТУ
Протокол №_____
засідання кафедри комп'ютерних наук
та кібербезпеки
від _____ 20___ р.
Завідувач кафедри
(_____)_____
(підпис) ПІБ

ЛУЦЬК − 2025

ЗМІСТ

ВСТУП

Медична візуалізація є ключовою для сучасної клінічної діагностики, а генеративні змагальні мережі (GAN) дозволяють створювати синтетичні зображення для розширення наборів даних і трансляції між модальностями, але потребують великих і різноманітних даних. Через розподіл медичних даних між установами та суворі закони про конфіденційність (HIPAA, GDPR) прямий обмін оригінальними зображеннями обмежений, а анонімізація не гарантує повної безпеки. Федеративне навчання (FL) дає змогу спільно тренувати моделі без передачі сирих даних, обмінюючись лише оновленнями моделей, але застосування до GAN ускладнене нестабільністю адверсаріального навчання, гетерогенними не-IID даними та ризиком витоку конфіденційної інформації через градієнти.

Для вирішення цих проблем потрібна стабільна та безпечна федеративна архітектура GAN, здатна обробляти не-IID дані. Це дослідження інтегрує розподілену оптимізацію для спільного навчання та агрегації моделей, теорію ігор для аналізу мінімаксної динаміки GAN, гомоморфне шифрування для захисту приватності, теорію інформації для оцінки якості синтетичних зображень (SSIM, FID) та методи медичної візуалізації для збереження анатомічної та патологічної достовірності.

Мета роботи полягає у створенні федеративної архітектури HealthFed-GAN для децентралізованого синтезу медичних зображень без обміну необробленими даними. Передбачено використання гомоморфного шифрування для забезпечення конфіденційності при агрегації оновлень моделей між установами. Якість зображень оцінюватиметься за SSIM, FID, PSNR та експертним аналізом, а архітектура перевірятиметься на точність прогнозування захворювань і відповідність стандартам конфіденційності HIPAA.

Теоретична цінність полягає у розвитку підходів федеративного навчання GAN у децентралізованому середовищі, інтеграції гомоморфного шифрування та

роботі з не-IID медичними даними для покращення збіжності та стабільності навчання.

Практична цінність полягає у застосуванні HealthFed-GAN для міжлікарняної співпраці без обміну пацієнтськими даними, підвищенні узагальнюваності моделей та генерації якісних синтетичних зображень для навчання і моделювання рідкісних патологій. Архітектура є масштабованою та адаптованою до різної лікарняної інфраструктури. Дослідження зосереджене на розробці та оцінці HealthFed-GAN для синтезу медичних зображень і прогнозування захворювань, з обмеженням на модальності MRI, CT і рентгенографії та використанням гомоморфного шифрування для захисту приватності в контрольованих умовах.

Об’єкт дослідження: технології федеративного навчання нейронних мереж у контексті обробки та аналізу медичних зображень.

Предмет дослідження: методи та архітектури федеративного навчання для генерації, синтезу та класифікації медичних зображень з урахуванням конфіденційності, стабільності навчання та обробки не-IID даних.

РОЗДІЛ 1.
ТЕОРЕТИЧНІ ЗАСАДИ МЕТОДІВ ФЕДЕРАТИВНОГО НАВЧАННЯ

1.1 Федеративне навчання

Генеративні змагальні мережі (GAN) привнесли новий аспект в обчислення медичних зображень, дозволяючи створювати фальшиві дані та міждоменні перетворення. У моделі суперечливого навчання генератор використовується для створення штучних зображень, а критик — для оцінки реалістичності цих зображень, а також обидві частини оптимізуються за допомогою конкурентного навчання. Процес суперечки, який дозволяють GAN, дає їм змогу створювати структурно валідні зображення, які можна використовувати для підтримки подальших діагностичних завдань, таких як класифікація, аугментація, реконструкція та переклад модальності. Те, що вони можуть вивчати високорозмірні розподіли клінічних зображень, робить їх особливо корисними в ситуаціях, коли реальні дані обмежені, незбалансовані або дорогі в маркуванні (Goodfellow et al., 2014).

Як глибоке навчання, так і конвеєри на основі GAN суттєво обмежені відсутністю репрезентативних і гетерогенних медичних наборів даних. Міжінституційна мінливість є високою через відмінності в апаратному забезпеченні для візуалізації, протоколах збору даних, алгоритмах реконструкції та демографічних характеристиках пацієнтів у різних установах. Більше того, обмеження щодо зберігання даних та політика управління часто не дозволяють централізувати зображення в одному сховищі. Ці перешкоди заважають створенню традиційних конвеєрів машинного навчання, які можна узагальнити, особливо при використанні з популяціями, сканерами або клінічними процесами, відмінними від тих, що використовуються при розробці моделей (Kaissis et al., 2021).

Ці обмеження долаються завдяки створенню федеративного навчання (FL), яке децентралізує обчислення до джерела даних. Замість консолідації сканів пацієнтів в одному центральному сховищі, FL дозволяє кожній установі проводити навчання на місці і надсилати лише оновлення моделей, усуваючи таким чином необхідність передачі необроблених зображень. Типова лікарня в циклі FL надає глобальну модель, виконує серію локальних етапів навчання та надсилає оновлені параметри на центральний сервер для агрегації. Цей процес підтримує суверенітет даних і дозволяє навчатися в масштабі популяції, що робить FL привабливим рішенням для давнього конфлікту між продуктивністю штучного інтелекту та конфіденційністю пацієнтів (McMahan et al., 2017).

FL також піднімає нові питання, зокрема в галузі медичної візуалізації. Дані, що зберігаються в лікарнях, рідко є незалежними та ідентично розподіленими (IID), а скоріше є локальною популяцією пацієнтів, постачальниками візуалізації та клінічними практиками. Така не-IID природа заважає конвергенції моделі, уповільнює навчання та збільшує мінливість оновлення моделі. Крім того, навіть обмін інформацією про градієнти або ваги моделі не обов’язково є безпечним для конфіденційності: суперечливі атаки показали, що зображення пацієнтів можуть бути приблизно відновлені навіть у випадку, якщо необроблені дані не обмінюються. Такі вразливості підкреслюють необхідність наявності надійного криптографічного захисту, вбудованого в федеративні системи, що використовуються в охороні здоров’я (Truex et al., 2019).

Алгоритми машинного навчання, що враховують конфіденційність, також є інструментами FL, що забезпечують конфіденційність, які включають безпечну агрегацію, диференціальну конфіденційність та криптографічне заплутування. Метою цих стратегій є збереження конфіденційної інформації та водночас забезпечення можливості обміну та інтеграції корисних оновлень моделі між учасниками для поліпшення профілю конфіденційності федеративної системи.(Acar et al., 2018).

Федеративне навчання в поєднанні з генеративними суперечливими мережами, які також відомі як федеративні GAN, є відносно новою концепцією. Експериментально було доведено, що GAN можна навчати в розподіленому клієнтському середовищі, хоча вони також виявляють серйозну нестабільність у разі неідентичних незалежних даних або клієнтів із сильно асиметричними вибірками. Більшість опублікованих експериментів з FL-GAN використовують невеликі та немедичні набори даних (наприклад, MNIST, CIFAR або синтетичні табличні дані) і не включають потужні властивості конфіденційності. Вони, як правило, дають результати з недостатньою структурною точністю для використання в клінічній практиці, особливо коли необхідний високий ступінь розмежування уражень або міжмодальна узгодженість (Rasouli et al., 2020).

Всі оглянуті літературні джерела разом показують, що в дослідженнях існує значна прогалина: хоча окремо глибоке навчання, GAN та FL стали швидшими, на даний момент не існує єдиної структури, яка б дозволяла використовувати всі три разом для застосування багатоінституційного, що зберігає конфіденційність, генеративного моделювання до синтезу медичних зображень та прогнозування захворювань. Більш сучасні дослідження федеративних умовних GAN та зашифрованого агрегування дають підстави сподіватися, що вони мають перспективу, але жодна з існуючих систем не може забезпечити синтез діагностичної якості і водночас гарантувати відповідність нормативним вимогам, міжсайтову генералізацію та зашифровану співпрацю. Ця прогалина безпосередньо надихає на розробку HealthFed-GAN, яка об’єднує федеративну оптимізацію, суперечливе навчання та гомоморфне шифрування в єдину архітектуру для застосування в клінічних процесах візуалізації (Duan et al., 2022).

1.2 Глибоке навчання для медичної візуалізації

Глибоке навчання стало домінуючою методологічною парадигмою в аналізі медичних зображень, і це пов’язано з його здатністю самостійно вивчати

багаторівневі представлення ознак необроблених радіологічних зображень, таких як МРТ, КТ, рентген, ультразвук та ПЕТ-зображення. глибоке навчання є основою сучасної прецизійної радіології, яка може застосовуватися для прийняття діагностичних рішень, планування лікування та прогнозної аналітики в широкому діапазоні клінічних областей (Shen et al., 2017).

Глибоке навчання також змінило спосіб виявлення уражень, спростивши автоматизовані процеси локалізації та класифікації, що зменшує необхідність покладатися на трудомісткі оцінки радіологів. Моделі, які були навчені за допомогою КТ-сканування, здатні ідентифікувати мікровузли, легеневі інфільтрати або емфізематозні зміни, але моделі на основі МРТ можуть ідентифікувати демієлінізуючі бляшки, ураження спинного мозку або слабкі межі пухлин з більшою чутливістю. Автоматичне виявлення уражень допомагає поліпшити систему сортування, швидше визначити пріоритетність невідкладних випадків та провести первинне втручання у хронічних випадках. Тому вони все частіше використовуються для допомоги радіологам у завантажених лікарнях, підвищення діагностичної продуктивності та мінімізації помилок, особливо в установах невідкладної допомоги та телерадіології, де необхідна швидка обробка даних (Srinidhi et al., 2021).

Моделі класифікації на основі глибокого навчання є важливим інструментом у галузі візуалізації, що дозволяє автоматично проводити класифікацію, розрізняти підтипи та стратифікувати ризики. Класифікатори на основі CNN були навчені для скринінгу діабетичної ретинопатії, гістопатологічного типування гліоми, непрозорості легенів та оцінки навантаження ураження при розсіяному склерозі і продемонстрували зростаючу точність. Такі системи надають корисну допомогу в програмах скринінгу, особливо в умовах обмежених ресурсів, коли кількість фахівців у галузі радіології є обмеженою. Крім того, класифікаційні мережі також взаємодіють з прогностичними моделями, які можуть передбачати прогресування захворювання, ризик рецидиву або реакцію на терапію, і таким чином просуваються

в напрямку прецизійної медицини та персоналізованих підходів до лікування в галузі онкології, кардіології та неврології (Ting et al., 2019).

Незважаючи на ці значні досягнення, моделі глибокого навчання в медичній візуалізації стикаються зі структурними проблемами, зокрема в областях обмежених даних та мінливості доменів. Дорогий характер ручного маркування, невідповідність між клініцистами та рідкість деяких патологій перешкоджають створенню великих анотованих наборів даних. Більше того, відмінності між постачальниками сканерів, процедурами придбання, характеристиками пацієнтів та місцевою клінічною практикою створюють варіації доменів, що підривають переносимість моделей у різних установах. Такі розбіжності часто спричиняють надмірне пристосування та низьку узагальнюваність, що обмежує практичне використання в реальному світі. Це, в свою чергу, призводить до невдачі багатьох високоточних моделей, про які повідомлялося в контрольованих дослідженнях, коли вони стикаються з гетерогенними зовнішніми наборами даних, що підкреслює необхідність багатосайтового спільного навчання та надійних стратегій навчання в домені (Glocker et al., 2019).

Закони про конфіденційність даних (HIPAA та GDPR) також обмежують можливості обміну даними між установами, що виключає централізацію медичних зображень навіть у ситуаціях, коли для побудови моделі потрібні великі обсяги даних. Етичні міркування, політика інституційних комісій з етики та механізми згоди пацієнтів також перешкоджають об’єднанню даних, особливо в чутливих технологіях візуалізації, таких як МРТ мозку або сканування дітей. Це деякі з обмежень, які обмежують масштабованість глибокого навчання та необхідність пошуку альтернативних рішень. Нові напрямки подолання цих викликів включають федеративне навчання, генерацію синтетичних даних, безпечну агрегацію та шифрований обмін градієнтами. Отже, хоча глибоке навчання радикально змінило сферу медичної візуалізації, воно ще не досягло повного потенціалу технології через відсутність питань, пов’язаних із конфіденційністю, неоднорідність даних та

відсутність різноманітності наборів даних, які вирішуються за допомогою федеративних, генеративних та гібридних моделей, зокрема HealthFed-GAN (Rieke et al., 2020).

1.2.1. Генеративні змагальні мережі

Концепція змагальних мереж (GAN) знаменує собою зміну парадигми в машинному навчанні, оскільки вона переносить акцент з дискримінаційного моделювання на вивчення генеративного розподілу, завдяки чому система здатна генерувати дані, які є реалістичними так само, як і емпіричні спостереження.

Ще більш очевидним стає трансформаційний характер GAN, якщо розглянути їх застосування в медичній візуалізації, де існують обмежені анотовані набори даних, нерівномірна поширеність захворювань, а закони про конфіденційність обмежують обмін даними між установами. (Shen et al., 2017).

Умовні генеративні суперечливі мережі (CGAN) є значним концептуальним розширенням сильних сторін GAN, що дозволяє додавати структуровану інформацію про умови, включаючи, але не обмежуючись, мітки класів, модальності зображень, маски сегментації або клінічно значущі метадані. У порівнянні з оригінальною архітектурою GAN, яка навчається відображати випадковий шум у вихідний простір без будь-яких явних обмежень, CGAN використовують умовні вхідні дані для контролю процесу генерації, що по суті перетворює GAN на неконтрольовані оцінювачі щільності або напівконтрольовані та повністю контрольовані системи синтезу. Наявність умовних векторів дозволяє генератору отримувати результати відповідно до бажаних цільових розподілів, як показали Мірза та Осіндро (2014), а дискримінатор визначає автентичність згенерованого зразка та його сумісність із заданою умовою. Ця подвійна система оцінки створює систематичну суперечливу систему, за допомогою якої модель отримує можливість інтегрувати як глобальні статистичні характеристики, так і специфічні для умов

варіації — особливість, яка є особливо корисною для створення клінічно значущих медичних зображень.

Застосування CGAN в медичній візуалізації є особливо важливим, оскільки велика кількість діагностичних процесів вимагає моделей для розрізнення тонких патологічних ознак, що відрізняються за типами захворювань, анатомічними системами або скануваннями. Наприклад, CGAN можна навчити на основі ступеня пухлини або масок сегментації пухлини для створення уражень з реалістичними варіаціями текстури, чіткості меж та локалізації уражень. Це дуже корисно для збільшення наборів даних про рідкісні підтипи раку, де важко отримати великі обсяги репрезентативних навчальних даних. Механізм кондиціонування генератора гарантує, що синтезовані ураження є точним відображенням клінічно значущої гетерогенності, необхідної для підвищення надійності класифікатора та зменшення кількості помилкових негативних результатів. Крім того, CGAN також використовуються для відтворення відсутніх методів візуалізації, включаючи створення FLAIR-послідовностей T1-зваженої МРТ або створення ПЕТ-зображень КТ, що робить мультимодальну візуалізацію менш дорогою та менш шкідливою з точки зору радіації. Такі завдання крос-модального синтезу можуть бути виконані, оскільки CGAN можуть умовно навчатися відображенням, які можуть бути використані для збереження анатомічної структури, але перетворюють розподіл інтенсивності, специфічний для модальності.

CycleGAN приніс парадигмальну зміну в генеративне моделювання, оскільки дозволяє здійснювати непарний переклад зображення в зображення, що є особливо корисним у медичній візуалізації, де часто недоступні парні набори даних, такі як ідеально вирівняні скани КТ-МРТ або скани з низькою дозою повної дози. Традиційні системи перекладу з контролем припускають відповідність один до одного між вихідним і цільовим зображеннями, що рідко буває в клінічній практиці через зміни в протоколах сканування, положенні пацієнта та обладнанні. Zhu et al. (2017) вирішили цю проблему, запропонувавши втрату циклової узгодженості: це

спонукає модель вивчати оборотні перетворення між двома доменами. Простіше кажучи, коли зображення перекладається між доменом A і доменом B, а потім те саме зображення знову перекладається в домен A, отримане зображення повинно бути ідентичним оригіналу. Це важливо, оскільки гарантує, що анатомічні структури залишаються незмінними протягом усіх перетворень, а генератор не дозволяє собі галюцинувати нереалістичні структури — обов'язкова умова в медицині, де діагностична ефективність залежить від структурної точності.

CycleGAN у клінічній візуалізації: Можливість перекладу без відповідних наборів даних створила нові можливості для гармонізації гетерогенних багатоцентрових даних та мінімізації розриву між модальностями. Наприклад, коли лікарні використовують різні поля інтенсивності МРТ 1,5 Т та 3 Т, якість зображення (контраст, розподіл шуму та просторова роздільна здатність) більшості зображень може відрізнятися, що призводить до поганої роботи моделі в разі інтеграції наборів даних. CycleGAN здатний конвертувати зображення, отримані за допомогою різних сканерів, в інші, по суті нормалізуючи профілі інтенсивності та мінімізуючи зміну домену. Аналогічно, його здатність перетворювати КТ з низькою дозою в еквіваленти повної дози з високою дозою допомагає зменшити дози опромінення та зберегти діагностичну чіткість, тому ця техніка дуже добре застосовується в педіатричній візуалізації та рутинних процедурах скринінгу. Окрім зменшення дози, CycleGAN також використовується для створення синтетичних сигналів МРТ, включаючи синтетичний T2 на основі T1-зважених зображень, які можуть бути використані для проведення повної діагностичної оцінки у випадках, коли деякі методи недоступні через обмеження часу або вартості.

Іншою значною перевагою CycleGAN є те, що його можна використовувати в міжмодальному синтезі, включаючи синтез МРТ-КТ, для планування радіотерапії. Для розрахунку дози потрібні КТ-зображення через їх електронну щільність, а МРТ забезпечує кращий контраст м'яких тканин. CycleGAN полегшує процеси на основі

МРТ шляхом створення зображень, схожих на КТ, тим самим усуваючи необхідність подвійного сканування. Той факт, що CycleGAN залишається структурно послідовним навіть при наявності артефактів, характерних для конкретної методики, наприклад, спотворень чутливості МРТ або ефектів зміцнення променя в КТ, робить CycleGAN потужним інструментом інтегрованих радіологічних конвеєрів. Крім того, він особливо сильний у федеративній установці, оскільки модель не залежить від парного нагляду; натомість кожна установа може навчати власні дані, характерні для конкретної методики, без обміну інформацією про пацієнтів. Було виявлено, що зображення CycleGAN набагато краще сегментуються та класифікуються при застосуванні в завданнях збільшення даних або заповнення модальності (Zhu et al., 2017). Завдяки цим напрямкам CycleGAN зарекомендував себе як фундаментальна суперечлива структура для медичної візуалізації, яка дозволяє непарні переклади, чутливі до структури, допомагати мультимодальній діагностиці, міжінституційній гармонізації та клінічно обґрунтованому синтезу.

StyleGAN є важливим у багатоцентрових дослідженнях візуалізації в області адаптації доменів. Лікарні, як правило, використовують різні сканери, котушки або протоколи збору даних, що призводить до відмінностей у роздільній здатності, розподілі інтенсивності та профілях артефактів. Контроль на рівні стилю StyleGAN дозволяє збалансувати такі відмінності шляхом зміни профілів контрасту, рівнів шуму та структурної текстури без знищення клінічно значущих областей. Це допомагає створювати набори даних про установи без будь-яких упереджень. Крім того, властивості редагування латентного простору StyleGAN дозволяють дослідникам вивчати морфологічний спектр розвитку захворювання, включаючи виділення або приховування набряків, запалень або моделей розвитку уражень, що може бути використано в освіті та діагностиці. Ряд досліджень продемонстрував, що розширення StyleGAN здатне підвищити надійність класифікації та сегментації гетерогенних тестових налаштувань (Karras et al., 2019). Загалом, концепція

розмежування StyleGAN та контролю стилю може розглядатися як значний прорив у галузі синтезу медичних зображень, що дозволяє отримувати реалістичні, контрольовані та діагностично значущі результати, які можуть бути використані для розробки штучного інтелекту клінічного рівня.

1.2.2 Принципи та обмеження федеративного навчання

Федеративне навчання (FL) перетворило сферу розподіленого машинного навчання, оскільки воно дозволяє декільком зберігачам даних спільно навчати загальну глобальну модель, не обмінюючись між собою конфіденційними наборами даних. Ця децентралізована парадигма є особливо актуальною для сфери охорони здоров'я, де записи про пацієнтів фізично розповсюджені по географічній території та підпадають під дію суворих законів про конфіденційність. Замість централізації даних в одному місці, FL розподіляється так, що глобальна модель розгортається в кожному місцевому розташуванні, де оновлюються градієнти або обчислюються навчені параметри, а потім підсумовуються сервером. Ця стратегія зменшить ризик порушення конфіденційності при передачі даних і сприятиме дотриманню законів, включаючи HIPAA та GDPR. Крім того, цей механізм дозволяє установам мати повне право власності на власні набори даних зображень і користуватися перевагами колективного інтелекту, що робить FL важливим ресурсом у розробці міжінституційної медичної ШІ (McMahan et al., 2017).

Теоретична основа FL базується на розподіленій оптимізації, в якій кожен клієнт мінімізує локальну цільову функцію на основі власного набору даних. Поєднання цих локальних оновлень періодично може призвести до глобального кроку оптимізації та дозволяє масштабоване спільне навчання без необхідності централізованого доступу до даних. Ця парадигма є дуже потужною в медичній візуалізації, де установи часто є неоднорідними з точки зору розміру даних, режиму візуалізації та демографічних даних, а також тягаря захворювань. Розподілена оптимізація гарантує, що навіть клієнти з невеликими наборами даних мають

значну інформацію про градієнт для глобальної моделі. Водночас FL зменшує обчислювальне навантаження на одну центральну машину, розподіляючи навчання на неоднорідні апаратні системи, що призводить до кращої ефективності та дозволяє паралельність між радіологічними мережами (Kairouz et al., 2021).

Федеративне навчання з передачею знань (FTL) розроблено в контексті, де установи мають відмінності як у просторі ознак, так і в розподілі ідентичності пацієнтів. FTL також використовує концепції глибокого навчання з передачею знань для передачі знань за допомогою параметрів моделі замість загальних даних або міток. Ця конструкція виявляється корисною в медичних мережах, де високоанотовані набори даних передових лікарень співпрацюють з меншими клініками, які мають обмежену кількість або взагалі не мають мічених зображень. За таких умов FTL дозволяє поширювати діагностичні знання на установи з обмеженими даними та підтримувати високий рівень конфіденційності. В результаті FTL сприяє справедливому використанню ШІ в закладах охорони здоров’я з обмеженими ресурсами, усуваючи розбіжності в можливостях установ без порушення конфіденційності (Liu et al., 2020).

Іншою основною проблемою є наявність даних, що не є незалежними та ідентично розподіленими (non-IID), в лікарнях. На відміну від інтернет-додатків, де дані, що генеруються користувачами, можуть мати структурну подібність, набори даних медичної візуалізації радикально відрізняються через демографічні відмінності, розподіл захворювань, невідповідність сканерів та відмінності в клінічних робочих процесах. Стан non-IID часто є поширеною причиною конфліктних оновлень градієнтів, нижчої загальної точності та суперечливого навчання в федеративних моделях GAN. Крім того, вартість комунікації є ще однією проблемою, яка ускладнює впровадження FL; параметри моделі або тензори градієнтів, особливо у випадку GAN та 3-D моделей візуалізації, вимагають великої пропускної здатності та більшої затримки. Усі ці обмеження підкреслюють необхідність оптимізованих стратегій комунікації, шифрованого стиснення

градієнтів та ієрархічних федеративних моделей, які можуть мінімізувати накладні витрати на передачу та зберегти точність (Konečný et al., 2017).

1.2.3. Алгоритми федеративної оптимізації

Структура надійної системи оптимізації визначає успіх або невдачу отриманої федеративної моделі в забезпеченні узагальненої продуктивності або невдачі в конфліктних градієнтах (Kairouz et al., 2021).

Федеративне усереднення (FedAvg) є оригінальним механізмом агрегації завдяки своїй концептуальній та обчислювальній простоті. У FedAvg кожен клієнт має кілька локальних етапів навчання, після яких нові параметри моделі надсилаються на центральний сервер, який обчислює зважене середнє за кількістю даних клієнта. Це зменшить накладні витрати на комунікацію та дозволить клієнтам використовувати локальні обчислювальні ресурси. Проте, той факт, що FedAvg базується на простому усередненні, робить його вразливим до даних, що не є незалежними та ідентично розподіленими (non-IID), що зазвичай змушує клієнтські упередження переважати під час глобального оновлення, особливо в завданнях візуалізації, де типовим є дисбаланс у класах або демографічні упередження. Незважаючи на ці недоліки, FedAvg продовжує бути популярною федеративною медичною системою завдяки своїй здатності пропонувати міцну базу з низькими витратами на впровадження (McMahan et al., 2017).

FedProx був запропонований як рішення недоліків FedAvg, особливо проблеми розбіжності локальних оновлень через нерівномірний розподіл даних клієнтів. FedProx використовує проксимальний член у цільовій функції, що обмежує величину відхилення в локальних і глобальних параметрах за цикл оновлення. Це обмеження гарантує, що клієнти з сильно асиметричними даними, наприклад, лікарні, які лікують рідкісні захворювання або мають унікальне обладнання для візуалізації, не домінують у глобальній моделі. FedProx особливо

добре працює у великих федераціях охорони здоров'я, де популяції пацієнтів, типи методів візуалізації та клінічні процеси значно відрізняються в різних місцевостях. Його стабілізуюча властивість дозволяє йому конвергувати більш послідовно та обмежує коливання моделі у випадку клієнтів, участь яких є переривчастою внаслідок мережевих або обчислювальних обмежень (Li et al., 2020).

FedYogi базується на просторі оптимізації, включаючи принципи адаптивної оцінки моменту в процедуру федеративного агрегування. FedYogi, на відміну від FedAvg і FedProx, динамічно змінює кроки градієнта та оцінки дисперсії у відповідь на несумісні або зашумлені оновлення клієнтів. Це особливо корисно у федеративних медичних моделях GAN, де нестабільність градієнта є відомою проблемою. FedYogi робить це шляхом ослаблення агресивних оновлень клієнтів-винятків та посилення інформативних оновлень клієнтів, що добре працюють. Це також робить його адаптивним і, отже, стабільним для федерації в дуже гетерогенних умовах, що робить його придатним для мереж з віддаленими або обмеженими ресурсами клініками, які виробляють оновлення низької якості або низької частоти (Reddi et al., 2021).

Всі ці алгоритми оптимізації створюють математичну та операційну основу федеративних систем навчання. Їх архітектура має прямий вплив на здатність таких фреймворків, як HealthFed-GAN, забезпечувати децентралізоване генерування медичних зображень без впливу на якість конвергенції, ефективність обчислень або конфіденційність. FedAvg дозволяє установам брати участь у масштабованому режимі, FedProx регулює варіації в розподілі клінічних даних, а FedYogi навчається в умовах реальної мінливості. Отже, вибір і точне налаштування цих алгоритмів визначають точність федеративних медичних конвеєрів на основі GAN та їхню здатність координувати інформацію багатоінституційних екосистем охорони здоров'я (Zhang et al., 2022).

1.2.4. Виклики у федеративному навчанні

Перша проблема — надмірна різниця в обчислювальній інфраструктурі медичних установ, що беруть участь у процесі. Великі третинні лікарні можуть навчати локальні моделі на серверах із графічними процесорами, але невеликі клініки або діагностичні центри можуть використовувати базові процесори з невеликим об'ємом пам'яті. Ці розбіжності призводять до нерівномірних темпів навчання, високої втрати клієнтів і нестабільних графіків оновлення. У разі, якщо деяким клієнтам може знадобитися дуже багато часу для завершення локальних епох, глобальний цикл навчання сповільнюється, що порушує роботу всієї федерації. Ця диспропорція ускладнює підтримку механізмів синхронного навчання і вимагає адаптивних або асинхронних підходів для запобігання тривалим періодам конвергенції (Bonawitz et al., 2019).

1.3 Машинне навчання з захистом конфіденційності

Машинне навчання з урахуванням конфіденційності (PPML) стало підгалуззю сучасної обчислювальної медицини у відповідь на нагальну необхідність розробки декількох моделей без порушення конфіденційності даних пацієнтів. Оскільки поява штучного інтелекту радикально змінює радіологію, патологію, геноміку та системи підтримки клінічних рішень, потреба у використанні великих і неоднорідних наборів даних тільки зросла. Однак суворі закони про конфіденційність медичних зображень, включаючи МРТ, КТ, ПЕТ, рентген та цифрову гістопатологію, не дозволяють лікарням накопичувати набори даних у централізованих сховищах. Таким чином, PPML надає набір математичних, криптографічних та алгоритмічних бібліотек, які дозволяють установам навчати моделі глибокого навчання у співпраці, наприклад, у федеративному навчанні, без обміну ідентифікованими даними пацієнтів. Захист конфіденційних даних від дедалі ширшого спектру загроз, включаючи компрометацію серверів, зловмисних

інсайдерів та просунуті атаки з реконструкцією градієнта, робить PPML необхідним для таких фреймворків, як HealthFed-GAN, які залежать від використання безпечної агрегації моделей декількох лікарень (Gursoy et al., 2021; Hu et al., 2022).

Деякі сучасні схеми HE, такі як BFV, BGV і CKKS, можуть використовуватися з наближеною та точною арифметикою, тому вони підходять для робочих навантажень глибокого навчання. Зокрема, C KKS широко використовується в медичних системах глибокого навчання, оскільки він здатний представляти числа з плаваючою комою та апроксимувати багато поліноміальних операцій, що є необхідним для конволюційних нейронних мереж (CNN) та моделей генеративних суперечливих мереж (GAN). Це дозволяє установам проводити федеративне навчання GAN з використанням зашифрованих даних без розкриття необроблених градієнтів, оскільки вона підтримує згортки, лінійні перетворення та апроксимацію поліноміальної активації зашифрованих даних. Незважаючи на те, що повне гомоморфне шифрування є дорогим з точки зору обчислень, оптимізації, такі як упаковка шифрованого тексту, представлення розріджених поліноміальних функцій та ядро HE на базі GPU, зробили його більш практичним для реалізації в системах радіології в режимі реального часу. Гібридні моделі, в яких гарантії конфіденційності узгоджуються з вартістю обчислень, можна досягти, шифруючи лише деякі шари або чутливі частини моделі (Boemer et al., 2020; Kim et al., 2021).

Ще однією сильною стороною HE є те, що він може працювати в мультихмарному середовищі з мінімальним або нульовим рівнем довіри. Лікарні з невеликою обчислювальною потужністю можуть передавати зашифровані обчислення на сторонні хмарні сервери, не побоюючись, що треті сторони отримають доступ до конфіденційних даних. Невелике навантаження на апаратне забезпечення в менших медичних закладах та рівноправна участь у федеративних навчальних мережах є перевагами децентралізації. Крім того, HE є ефективним засобом протидії пасивним супротивникам, які хочуть вивчити загальні градієнти, стан серверів або оновлення моделей, щоб дізнатися особистість пацієнтів або

відтворити зображення. Оскільки зашифровані повідомлення не можна розшифрувати без секретного ключа дешифрування, навіть досвідчені зловмисники не можуть використовувати зашифровані сигнали для ініціювання атак на висновок, і тому HE є одним з найсильніших криптографічних засобів захисту в машинному навчанні, що зберігає конфіденційність (Zhang et al., 2023).

HE також створює проблеми з реалізацією, особливо з архітектурами супротивників, такими як GAN, які вимагають багатораундової оптимізації та частого обміну параметрами. Повністю гомоморфне шифрування може значно уповільнити процес навчання, оскільки кожна зашифрована операція є обчислювально дорогою задачею. Це особливо актуально для федеративних GAN, в яких дискримінатор і генератор оновлюють один одного кілька разів. В результаті були досліджені методи вирівняного HE та часткового шифрування, які захищають найбільш чутливі градієнти або шари, що призвело до мінімізації обчислювальних витрат і високих гарантій конфіденційності. Розробка алгоритму повинна бути ретельно продумана, щоб додане шифрування не дестабілізувало динаміку суперечливого навчання (Chen et al., 2023).

Розширення HE на федеративні моделі, такі як HealthFed -GAN, є зміною в парадигмі інженерії конфіденційності — від силосів даних з контрольованим доступом до забезпечення безпечних обчислень за допомогою криптографії. HE дозволяє установам дотримуватися суворих правових рамок і сприяє безпечному спільному навчанню на національному та міжнародному рівнях. Це один із стовпів проектування майбутніх медичних екосистем штучного інтелекту завдяки його здатності зберігати конфіденційність без обмеження виразності моделі (Xiao et al., 2021).

Безпечні багатосторонні обчислення (MPC) — це метод, який дозволяє декільком сторонам співпрацювати в обчисленні функції, не дозволяючи жодній зі сторін отримати доступ до приватних даних іншої сторони. На відміну від HE, яка в основному шифрує числові тензори, MPC ділиться секретними даними або

параметрами між декількома обчислювальними вузлами. На індивідуальній основі кожна з частин представлена як випадковий шум, але коли частини додаються, вони відновлюють значущі обчислення. Таким чином, MPC забезпечує децентралізацію довіри, оскільки жоден сервер або установа не може мати повного доступу до даних, що обробляються. Ця властивість необхідна в умовах міжінституційних середовищ, де жодна центральна влада не має довіри для управління конфіденційними медичними даними (Mohassel & Rindal, 2018).

MPC має особливе значення в медичних федеративних навчальних конвеєрах, коли йдеться про такі проблеми, як безпечне агрегування градієнтів, оновлення моделей усереднення, чутливих до конфіденційності, та спільний розрахунок функцій втрат GAN. Федеративна архітектура GAN передбачає, що і генератор, і дискримінатор повинні оновлюватися одночасно на основі розподілу даних між різними установами, що означає, що агрегація на основі MPC є критично важливою для забезпечення того, щоб локальні шаблони, що зберігаються в градієнтах, не витікали. Рішення MPC можуть допомогти лікарням брати участь у спільному навчанні спільної моделі без розшифрування або розкриття будь-яких проміжних повідомлень, що є співпрацею з нульовим розкриттям інформації. Це є важливим у мультимодальних системах, де зображення КТ може зберігатися в одній установі, ультразвукове сканування — в іншій, а відповідні клінічні метадані — у третій установі. MPC також гарантує, що всі учасники беруть участь у обчисленні і не розкривають необроблені характеристики (Cheng et al., 2021; Kumar et al., 2022).

Однією з істотних переваг MPC порівняно з HE є його гнучкість і можливість застосування до інтерактивних багатораундових протоколів. HE добре підходить для застосувань, де обчислення є неінтерактивними та зашифрованими, тоді як MPC є більш гнучким при роботі з алгоритмами, контрольний потік яких є динамічним або логіка яких є умовною, що є характерною рисою федеративного суперечливого навчання. Наприклад, MPC може працювати з умовними оновленнями генератора, петлями зворотного зв’язку дискримінатора та

міжінституційною агрегацією функцій втрат. Однак MPC може бути більш комунікаційно інтенсивним, ніж великомасштабні федерації медичних лікарень, і може призвести до більшого використання пропускної здатності або затримки, особливо у великомасштабних медичних федераціях з десятками лікарень. Проте, в недалекому минулому розвиток легких протоколів розподілу секретів та MPC зі скороченим циклом зробив його більш придатним для впровадження в медичних додатках (Haseeb et al., 2023).

Диференціальна конфіденційність (DP) — це тип гарантії конфіденційності, який надає математично кількісно вимірювані гарантії конфіденційності, гарантуючи, що наявність або відсутність певного фрагмента даних не впливає на модель жодним істотним чином. Це досягається шляхом додавання контрольованого шуму, який зазвичай є гаусівським або лапласівським, до градієнтів, параметрів моделі або прогнозів виходу. На відміну від HE або MPC, які захищають дані протягом усього процесу обчислення, DP захищає дані від атак на висновок після розгортання моделі. Це дуальність, яка робить DP важливим доповненням до криптографічних технік, особливо при використанні федеративних моделей GAN, де генеративні моделі мають перспективу відтворити специфічні для пацієнта структури за відсутності вимог конфіденційності (Williams and McSherry, 2021).

DP зазвичай використовується у федеративному навчанні при обміні градієнтами. Перед відправкою їх на сервер лікарні вводять шум у свої локальні оновлення, щоб перешкодити супротивникам аналізувати градієнтні патерни, які можуть вказувати на характеристики пацієнтів. Локальна диференціальна конфіденційність (LDP) також розширює цей захист, гарантуючи, що необроблені градієнти очищаються перед виходом з локальної машини, тим самим захищаючи від компрометації сервера та зловмисних інсайдерів. У федеративному навчанні GAN DP має бути ретельно налаштований для підтримки рівноваги між супротивниками; занадто багато шуму призведе до нестабільності навчання GAN,

а занадто мало — до вразливості моделі до атак інверсії. Як результат, оптимальна конфігурація DP також є основною проблемою генеративного моделювання із збереженням конфіденційності (Liu et al., 2022).

DP також стійкий до повторюваних запитів або адаптивних суперечливих ризиків. Припускаючи, що зловмисник неодноразово досліджує модель, щоб витягти особливості навчальних даних, математичні гарантії DP обмежують сукупну втрату конфіденційності за допомогою теорем композиції, усуваючи атаки витягання, які спрямовані на клінічно розгорнуті системи підтримки прийняття рішень. Проте додатковий шум може знизити точність прогнозів або якість зображень, згенерованих GAN. У випадку моделей, таких як HealthFed-GAN, адаптивний підхід DP, який характеризується різним рівнем шуму на кожному шарі або етапі навчання, може підтримувати якість діагностики та забезпечувати високий рівень конфіденційності (Ponomareva et al., 2023).

Атаки витоку градієнтів використовують той факт, що загальні градієнти у федеративному навчанні містять дрібні деталі навчальних зображень. Останні дослідження показали, що супротивники можуть відтворити зображення з високою роздільною здатністю, використовуючи один пакет градієнтів, риси обличчя, рукописний текст та анатомічні структури. Цей ризик є ще більшим у галузі медичної візуалізації: градієнти зображень МРТ або КТ містять впізнавані структурні особливості, включаючи контури черепа, морфологію пухлин або межі органів. Ці деталі можна відтворити за допомогою алгоритмів, що не змінюють градієнти, шляхом оптимізації синтетичних зображень до досягнення спостережуваних градієнтів. Тому машине навчання, що забезпечує конфіденційність, повинно мати ефективні стратегії захисту, щоб гарантувати збереження градієнтів перед їх передачею між установами (Yin et al., 2021).

Активні атаки на градієнти є ще більш небезпечними, оскільки супротивники можуть контролювати гіперпараметри навчання, такі як розмір партії, швидкість навчання або ваги втрат, щоб оптимізувати потенціал витоку. Прикладом є те, що

надзвичайно малі розміри партій роблять їх вразливими, оскільки градієнт одного зразка може бути відокремлений. Як і в разі маніпулювання архітектурою моделі або умовами навчання, витік також збільшується. Витік градієнтів особливо небезпечний у федеративних GAN, оскільки градієнти дискримінатора містять детальну семантичну інформацію високого рівня, як реальну, так і синтетичну. За відсутності захисних механізмів, таких як HE, MPC або DP, HealthFed-GAN буде схильний до розкриття конфіденційних даних про пацієнтів (Liang et al., 2022; Fang et al., 2023).

Витік градієнтів також може відбуватися в пасивних середовищах, в яких сервер або інші клієнти просто переглядають градієнти, не контролюючи процес навчання. Федеративне навчання все ще може бути вразливим для всіх учасників, які здаються чесними, але цікавими. Щоб запобігти втраті інформації про градієнти, захисними заходами є обрізання градієнтів, рандомізація, структуроване розрідження та зашифровані обчислювальні конвеєри. Наприклад, обрізання великих значень градієнта може мінімізувати здатність зловмисників отримувати точні інтенсивності пікселів або карти ознак. Проте надмірна агресивність обрізання може спричинити зниження конвергенції моделей або зниження точності GAN. Тому федеративне навчання GAN має бути вразливим до балансу між надійністю та виразністю (Shen et al., 2023).

Компроміс між конфіденційністю та корисністю має ще один аспект, пов’язаний з інтерпретованістю моделі. Шум (або шифрування) або розподілені обчислення можуть зробити внутрішні процеси моделі менш видимими для клініцистів або регуляторних органів, створюючи конфіденційність. З іншого боку, прозорість є вимогою для клінічного застосування. Таким чином, машине навчання, чутливе до конфіденційності, повинно включати техніки пояснюваності, які не дадуть збою в умовах контролю конфіденційності. Шифровані обчислення значущості, диференційно-приватні значення SHAP та атрибуційна оцінка з підтримкою MPC — це деякі з методів, запроваджених у спробах створити баланс

між прозорістю та конфіденційністю в клінічних системах штучного інтелекту (Bai et al., 2023).

1.4. Існуючі підходи FL-GAN

Федеративне навчання в поєднанні з генеративними змагальними мережами (FL-GAN) є перспективним рішенням для децентралізованого генерування даних та синтетичного зображення з збереженням конфіденційності. FedGAN є одним з перших додатків, який запропонував розподілений протокол суперечливого навчання, де окремі клієнти навчаються на локальних парах генераторів та дискримінаторів, а підмножина параметрів моделі надсилається на центральний сервер для агрегації. У цій статті показано, що можна навчати GAN без централізації необроблених даних, що забезпечує основу для підтвердження концепції федеративного генеративного моделювання в умовах, чутливих до конфіденційності (Hardy et al., 2019).

Архітектура FedGAN дозволяє лікарням зберігати право власності на дані візуалізації та цикл суперечливого зворотного зв'язку, тоді як оновлення центрального сервера є моделями, незалежними від пацієнтів. Однак той факт, що навчання є нестабільним, також є неминучою слабкістю, оскільки неоднорідність оновлень локальних дискримінаторів збільшує суперечливі коливання. FedGAN часто повільно збігається в ситуаціях, коли медичні набори даних мають різні модальності, різні параметри придбання або патологічний розподіл, і не вчиться фіксувати рідкісні захворювання серед глобальних генераторів (Rasouli et al., 2020).

Інший важливий варіант, DP-GAN, додає механізми диференційної конфіденційності до навчання GAN, щоб уникнути витоку особистої інформації про пацієнтів. Додаючи контрольований шум до виходу градієнта або дискримінатора, DP-GAN забезпечує формальні гарантії конфіденційності, що особливо корисно, коли це вимагається законодавством, включаючи HIPAA та GDPR. Проте ці сильні

сторони мають значний компроміс між конфіденційністю та корисністю: занадто багато шуму знижує чіткість зображення, структурну подібність (SSIM) та точність у клінічно важливих областях, таких як пухлини та ураження (Xie et al., 2018).

У федеративних системах DP-GAN також має більше труднощів з впровадженням, оскільки накопичений шум у багатораундовому навчанні тільки збільшує спотворення в глобальному генераторі. Як результат, штучні зображення мають розмиті анатомічні межі або не є деталізованими рентгенографічними ознаками. Отже, безпосереднє використання DP-GAN у медичній візуалізації все ще обмежене, за винятком використання адаптивних методів зменшення шуму або багатоступеневого регулювання конфіденційності (Torkzadehmahani et al., 2019).

Незважаючи на ці переваги, федеративні FL-GAN з декількома дискримінаторами є дуже дорогими з точки зору комунікаційних витрат, оскільки градієнти дискримінаторів великі і повинні передаватися на сервер з певною частотою. Крім того, дисбаланс наборів даних між клієнтами створює неоднорідні суперечливі умови для різних дискримінаторів, що, як правило, дестабілізує глобальний генератор.

На основі цих загальних схем було запропоновано низку спеціалізованих схем FL-GAN для виконання таких завдань, як федеративне виявлення аномалій, генерація зображень з низькими ресурсами та адаптація домену з збереженням конфіденційності. Хоча ці дослідження демонструють практичність децентралізованого генеративного моделювання, більшість з них використовують немедичні дані (наприклад, MNIST, CIFAR-10 або CelebA).

Незважаючи на зростаючий інтерес до вивчення генеративних суперечливих та інших типів мереж у федеративному навчанні, література з цієї теми залишається недостатньою, щоб запропонувати модель FL-GAN медичного рівня для створення клінічно валідних синтетичних зображень у різних установах. Більшість попередніх дослідників тестують свої системи, використовуючи спрощені двовимірні набори даних, такі як MNIST, CIFAR або CelebA, які не відображають анатомічну

складність, модальну мінливість та діагностичну точність, необхідні в медичній візуалізації. Тому існуючі FL-GAN дають результати низької точності з низькою структурною подібністю (SSIM) і не мають специфічних для захворювання особливостей. По-друге, другою основною слабкістю є недостатнє застосування криптографічного захисту: більшість федеративних досліджень GAN використовують просте усереднення параметрів і не використовують гомоморфне шифрування або безпечну агрегацію. Це має недолік у вигляді розкриття оновлень градієнта для атак на реконструкцію і, отже, компрометації цілей децентралізованого навчання щодо конфіденційності. Як наслідок, поточні структури не відповідають високим стандартам регулювання, що вимагаються налаштуваннями, сумісними з HIPAA, особливо при роботі з даними, що можна ідентифікувати, такими як МРТ, КТ або рентгенівські знімки.

Слабкість виникає через обмежену увагу до ідеї спільного підходу до синтезу зображень та прогнозування захворювань; переважна більшість досліджень FL-GAN зосереджується на генеративній точності, а не на інтеграції подальших клінічних завдань, тобто класифікації уражень або оцінці патології. Така відсутність релевантності знижує корисність, оскільки штучні медичні зображення в кінцевому підсумку повинні бути необхідними для сприяння діагностичному моделюванню, сортуванню пацієнтів або розширенню прогнозних аналізів. Крім того, не існує опублікованої реалізації комплексного федеративного конвеєра GAN на основі PyTorch, що включає шифрований обмін градієнтами, специфічні для клієнта цикли навчання, координацію між різними установами та комплексну медичну метричну валідацію (наприклад, SSIM, FID, AUC). Доступні фреймворки є здебільшого теоретичними, розрізненими або обмеженими симульованим середовищем, на відміну від реального клінічного впровадження. Усі ці прогалини свідчать про необхідність комбінованого рішення, такого як HealthFed-GAN, яке інтегрує федеративне суперечливе навчання, гомоморфне шифрування, генерацію синтетичних зображень та прогнозування захворювань в єдину медичну штучну інтелігенцію.

РОЗДІЛ 2.
МЕТОДОЛОГІЯ ЗАСТОСУВАННЯ МЕРЕЖІ HEALTHFED-GAN ДЛЯ КЛАСИФІКАЦІЇ МЕДИЧНИХ ЗОБРАЖЕНЬ ТА ПРОГНОЗУВАННЯ

2.1. Дизайн дослідження

У дослідженні використовується експериментальна багатоінституційна федеративна модель навчання (FL), в якій чотири організації охорони здоров'я виступають в ролі розподілених клієнтів, що беруть участь у спільному навчанні в рамках HealthFed-GAN. Ключовою ідеєю цього дизайну є моделювання реалістичного клінічного середовища, в якому лікарні зберігають великі обсяги конфіденційних медичних зображень (тобто МРТ, КТ, рентгенівські знімки), які не можуть бути централізовано зберігатися відповідно до правил конфіденційності (тобто HIPAA, GDPR), інституційних політик управління та етичних міркувань. Замість того, щоб обмінюватися необробленими зображеннями пацієнтів, кожна установа проводить локальне навчання GAN і надсилає на центральний сервер лише зашифровані оновлення градієнта. Така структура сприяє спільному навчанню світових моделей зображень і гарантує, що інформація про конкретних пацієнтів буде суворо обмежена стінами установ.

Крім того, ця конструкція пропонує контрольоване експериментальне середовище для вивчення поведінки HealthFed-GAN в реальних умовах, включаючи гетерогенне обладнання, неідентичні розподіли даних та дисбаланс між клієнтами. Завдяки організації навчання ітеративних циклів зашифрованого агрегування, дослідження оцінює не тільки якість синтезованих медичних зображень, але й ефективність механізмів захисту конфіденційності, продуктивність комунікації та стійкість до атак інверсії градієнта або реконструкції. Така багатоінституційна конфігурація FL гарантує, що запропонована структура випробовується в клінічно та практично релевантних умовах, які охоплюють різноманітні методи візуалізації

та децентралізовані робочі процеси, поширені в сучасних системах охорони здоров'я.

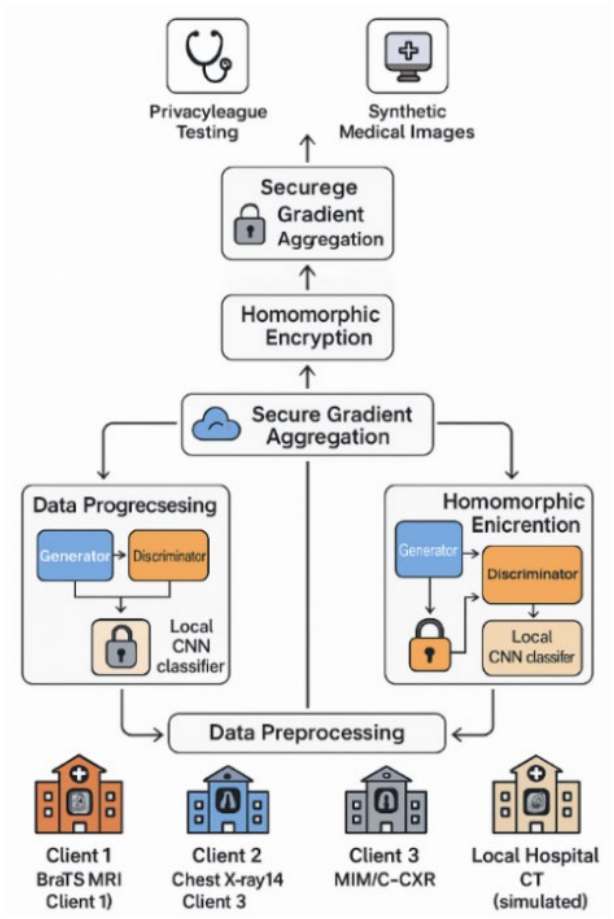

Рисунок 2.1— Схема методології роботи платформи HealthFed-GAN

На рисунку 2.1 показано загальний методологічний підхід, який буде використовуватися в цьому дослідженні. Починаючи з багатоінституційного збору даних (чотири гетерогенні джерела медичних зображень, включаючи BraTS MRI, ChestX-ray14, MIMIC-CXR та локальні КТ-скани), що базується на меншій базі локальної попередньої обробки, кожен клієнт навчає свою власну пару генератор-дискримінатор на власних даних. Отримані градієнти потім шифруються за допомогою гомоморфного шифрування Paillier і безпечно надсилаються на центральний сервер; там зашифровані оновлення підсумовуються для

вдосконалення глобального генератора, який знову поширюється на всіх клієнтів, щоб підтримати наступну ітерацію навчання. Отримані синтетичні зображення розглядаються з використанням метрик реалістичності (SSIM, PSNR, FID), індексу діагностичної ефективності (точність, AUC) та кількісної оцінки ризику конфіденційності, що, в свою чергу, підкреслює, що HealthFed-GAN може підтримувати високий рівень конфіденційності даних і дозволяє здійснювати спільне створення медичних зображень між установами.

2.1.1. Децентралізоване навчання

Кожна лікарня самостійно навчає локальну генеративну суперечливу мережу (GAN), що складається з власних модулів генератора та дискримінатора.

Клієнт 1 навчає свою GAN на зображеннях BraTS MRI.

Клієнт 2 навчає на рентгенівських знімках NIH ChestX-ray14.

Клієнт 3 навчає на зображеннях MIMIC-CXR.

Клієнт 4 навчає на локальному наборі даних КТ.

Зображення та метадані пацієнтів, а також карти ознак не виходять за межі місцевого середовища. Децентралізована структура є показовою для реальних екосистем лікарень, в яких принципи управління даними не дозволяють централізовано агрегувати медичні дані..

2.1.2. Співпраця із збереженням конфіденційності

Для забезпечення безпечної співпраці клієнти надсилають на центральний сервер лише зашифровані градієнти або оновлення моделей. У цій статті використовується гомоморфне шифрування Paillier, яке дозволяє виконувати математичні операції над зашифрованими значеннями. Це означає, що сервер об’єднує градієнти і не потребує розшифрування, жодна установа не отримує доступ до градієнтів іншої організації, і навіть перехоплені дані не піддаються розшифруванню. Така конструкція ефективно усуне загрозу витоку необроблених

даних і забезпечить відповідність навчального конвеєра правилам GDPR, HIPAA та управління IT в лікарнях:

- сервер може об’єднувати градієнти без їх розшифрування;
- жодна установа не бачить градієнти іншої;
- навіть у разі перехоплення дані залишаються нечитабельними.

Ця конструкція остаточно усуває ризики витоку необроблених даних і робить навчальний конвеєр сумісним з GDPR, HIPAA та правилами управління IT в лікарнях.

2.1.3. Крос-модальна федеративна конфігурація

Основною методологічною сильною стороною цього дослідження є його міжмодальна конфігурація, в якій кожен клієнт має різну модальність медичної візуалізації:

- МРТ;
- рентген грудної клітки;
- клінічні рентгенограми грудної клітки з висновками;
- КТ.

Ця конфігурація відображає не-IID (незалежний і неідентично розподілений) характер реальних клінічних даних.

Вона також оцінює, чи може HealthFed-GAN навчитися надійному глобальному генератору, здатного синтезувати реалістичні зображення незалежно від варіацій, специфічних для конкретної модальності.

2.1.4. Ітеративна глобальна агрегація

Федеративне навчання проводиться в раундах комунікації. Після кожного раунду:

- клієнти проводять локальне навчання в заздалегідь визначених епохах;
- клієнти завантажують оновлення градієнтів, які шифруються;

- гомоморфна агрегація виконується центральним сервером, який створює новий набір глобальних ваг моделі.

Ці оновлені ваги генераторів/дискримінаторів потім перерозподіляються сервером між усіма клієнтами. Це циклічний процес, який триває до збіжності, забезпечуючи безперервну оптимізацію глобальної GAN без шкоди для базових медичних наборів даних.

Дизайн дослідження навмисно орієнтований на реалії функціонування екосистем, що складаються з декількох лікарень, шляхом:

- децентралізоване навчання, яке усуває централізований доступ до даних пацієнтів;
- безпечна комунікація, що дозволяє уникнути втручання проміжних обчислень;
- міжмодальне навчання, гетерогенність у реальному світі;
- федеративна агрегація, яка може бути використана ітеративно для вдосконалення моделей GAN.

Разом ці методологічні рішення формують науково суворий, відтворюваний та орієнтований на конфіденційність експериментальний дизайн оцінки федеративного синтетичного генерування медичних зображень у практичних контекстах охорони здоров’я.

2.2. Збір даних

Це дослідження базується на дослідницькій моделі експериментальної та багатоінституційної федеративної моделі навчання (FL), в якій чотири медичні установи є децентралізованими клієнтами, які спільно навчають запропоновану структуру HealthFed-GAN. Цей дизайн базується на реальній функціональності сучасних клінічних умов, де лікарні мають великий обсяг чутливих медичних зображень (включаючи МРТ, КТ та рентгенівські знімки), але не мають права

розголошувати цю інформацію через суворі політики конфіденційності, організаційні політики та етичні обмеження. Замість того, щоб збирати дані в центральному сховищі, федеративна архітектура дозволяє кожній установі мати повне право власності та контроль над своїми наборами даних зображень і, водночас, брати участь у спільному процесі розробки моделей.

За цією схемою кожна з чотирьох установ має специфічну модальність зображення для збереження екологічної валідності та моделювання неідентично розподілених (non-IID) умов реального світу. Зокрема, Клієнт 1 проходить МРТ-сканування мозку (набір даних BraTS), Клієнт 2 має зображення у вигляді рентгенівських знімків грудної клітки (NIH ChestX-ray14), Клієнт 3 управляє рентгенівськими знімками MIMIC-CXR, а Клієнт 4 має локальний набір даних КТ, який можна розглядати як репрезентативний для стандартного лікарняного середовища. Такий розподіл дозволяє досліджувати федеративні генеративні суперечливі мережі (GAN) в умовах дисбалансу даних, гетерогенності та мультимодальності — ситуацій, які регулярно спостерігаються в клінічній практиці.

Однією з основних ідей проекту є децентралізація навчання, завдяки чому кожна лікарня матиме власний генератор і дискримінатор, які будуть навчатися на власних даних. У кожному раунді навчання лікарні обчислюють локальні градієнти і не діляться своїми необробленими зображеннями та внутрішніми характеристиками даних пацієнтів. Натомість оновлення моделі шифруються за допомогою гомоморфної системи шифрування Paillier, завдяки чому математична форма градієнтів зберігається, але несанкціонований огляд або реконструкція неможливі. Зашифровані параметри надсилаються лише на центральний сервер.

Центральний агрегатор є дуже важливою частиною дизайну дослідження. Він збирає зашифровані градієнти всіх клієнтів і виконує гомоморфне агрегування, що дозволяє обчислювати зашифровані значення без розшифрування. Розшифрування остаточного агрегованого оновлення, але не оновлення клієнта, єдиним

центральним сервером повністю запобігає виведенню інформації про конкретного пацієнта будь-якої лікарні. Після агрегування переглянуті глобальні параметри генератора повторно передаються всім клієнтам, що дозволяє їм продовжувати навчання, використовуючи набагато багатше глобальне представлення без втрати конфіденційності даних. Цей цикл прямого та зворотного зв'язку повторюється багато разів протягом етапів комунікації, доки не відбудеться збіг.

Нарешті, дизайн дослідження відображає методологію дизайну науки, в якій новий обчислювальний артефакт розробляється, розвивається та перевіряється емпірично в контрольованих, але реалістичних обмеженнях. Ця суворість методології гарантується тим, що запропонована структура HealthFed-GAN є не лише теоретично обґрунтованою, але й практичною та придатною для використання, стійкою до гетерогенних умов даних та відповідає суворим стандартам медичної конфіденційності.

У таблиці 2.1 описано чотири модельовані медичні установи, які братимуть участь у федеративному навчанні фреймворку HealthFed-GAN.

Таблиця 2.1

Огляд об'єднаних клієнтів та їхніх наборів даних медичних зображень, що використовуються в навчанні HealthFed-GAN

Клієнт	**Установа (модельована)**	**Набір даних**	**Модальність**	**Розмір**
Клієнт1	Лікарня A	BraTS 2020/2021	МРТ	369 суб'єктів (багатопослідовне 3DМРТ)
Клієнт2	Лікарня B	NIH ChestX-ray14	Рентген	112,120 зображень
Клієнт3	Лікарня C	MIMIC-CXR	Рентген	377,110 зображень
Клієнт4	Лікарня D	Local КТ Dataset	КТ	6,000 зрізів (модельованих)

Кожен клієнт використовує різні методи візуалізації (МРТ, рентген або КТ), щоб відтворити реалістичну міжінституційну неоднорідність у клінічних умовах. Розмір наборів даних є помірним (когорта BraTS MRI, 369 суб'єктів) і великим, як у радіологічних архівах ChestX-ray14 та MIMIC-CXR, що гарантує різноманітність патологічних явищ та радіологічних особливостей. Такий розподіл підкреслює той факт, що федеративна конфігурація є не-IID, оскільки установи відрізняються за модальністю та обсягом, а процес навчання є ближчим до реальної ситуації федеративної медичної візуалізації.

Доступ до наборів даних лікарень здійснювався локально в контрольованому середовищі. Переміщення даних за межі приміщення було суворо заборонено. Експерименти імітували реальні обмеження IT-систем лікарень (пропускна здатність, обмеження пам'яті та бар'єри доступу до файлів).

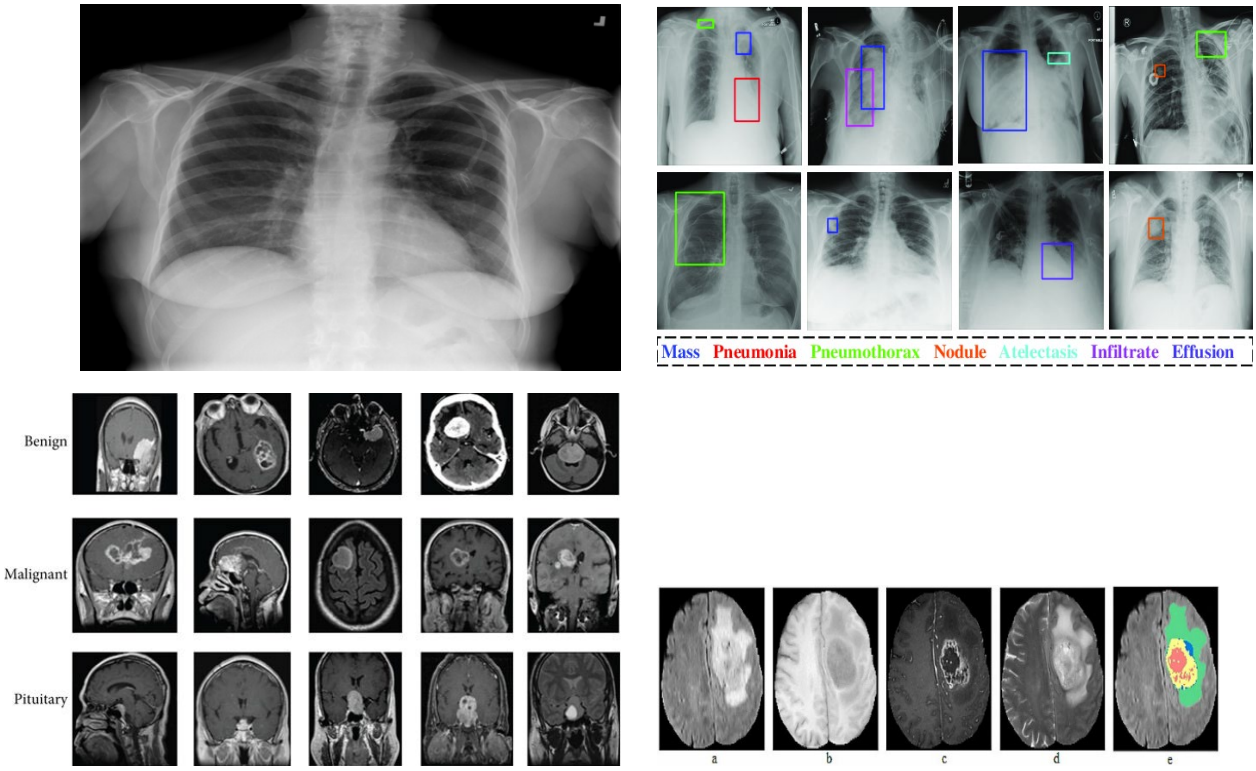

Рисунок 2.2—Приклади з об'єднаних наборів даних: Рентген грудних клітки (NIH ChestX-ray14), МРТ-зріз мозку (BraTS)

Гетерогенні методи візуалізації, які використовувалися в дослідженні, зображені на цих вибраних парах зображень, як показано на рисунку 3-2. Рентген грудної клітки є легкодоступним методом рентгенографії з великим обсягом даних, який зазвичай використовується в загальних лікарнях; МРТ головного мозку є об'ємним, спеціалізованим методом візуалізації, який зазвичай використовується в третинних неврологічних центрах. Об'єднана структура також може бути поставлена під сумнів через реальну гетерогенність у реальному світі, включаючи обидва методи і, таким чином, полегшуючи побудову моделі HealthFed GAN на основі розбіжних даних клієнтів.

2.3 Опис набору даних

Збір даних, використаний у цьому дослідженні, включає три основні методи радіологічної візуалізації, а саме МРТ, рентген та КТ, кожен з яких був обраний для моделювання специфічної для методу гетерогенності та не-IID (незалежного та ідентично розподіленого) розподілу даних серед об'єднаних клієнтів. Такий вид варіації є властивим для багатоінституційних медичних контекстів, в яких лікарні відрізняються за типом діагностичного фокусу, обладнанням для візуалізації, протоколами збору даних та демографічними характеристиками їхнього населення. Включення цих гетерогенних наборів даних до системи HealthFed -GAN дає змогу критично оцінити її здатність здійснювати генеративне моделювання з збереженням конфіденційності та міжцентрове навчання представлення з урахуванням реалістичних клінічних міркувань.

Призначення кожної модальності конкретній модельованій організації охорони здоров'я відображає реальні партнерські відносини між декількома лікарнями, де дані не можуть бути централізовані через етичні, правові та регуляторні обмеження. Такий дизайн гарантує, що GAN тестується не тільки з точки зору реалістичності зображень, але й з точки зору його здатності

узагальнювати різні типи візуалізації без використання необроблених даних пацієнтів.

2.3.1. Набір даних BraTS MRI (Клієнт 1 - Лікарня A)

Дані Brain Tumor Segmentation (BraTS) слугують джерелом МРТ для Клієнта 1 і є багатопослідовним 3D-об'ємом МРТ мозку T1, T1CE, T2 та FLAIRs. Він містить 369 анотованих досліджень пацієнтів, які надаються з масками сегментації пухлин на рівні вокселів. Ці багатопослідовні обсяги здатні фіксувати специфічні контрасти тканин, а отже, дозволяють моделі вивчати специфічні структурні, анатомічні та патологічні особливості нейроонкологічної візуалізації (Menze et al., 2015).

Дані BraTS мають високий рівень гетерогенності з точки зору морфології пухлин, анатомії пацієнтів та параметрів отримання МРТ. Варіація інтенсивності, шум, який залежить від сканера, та істотний дисбаланс між підрегіонами пухлин роблять BraTS обчислювально дорогим, але клінічно репрезентативним середовищем. Додавання BraTS дозволяє HealthFed-GAN вивчати об'ємну структуру з високою роздільною здатністю, коли вона працює в умовах, специфічних для МРТ, тим самим імітуючи спеціалізований центр неврології третинної медичної допомоги (Bakas et al., 2018).

2.3.2. Набір даних ChestX-ray14 (Клієнт 2 — Лікарня B)

Клієнт 2 працює з набором даних ChestX-ray14, який є одним з найбільших наборів маркованих рентгенограм грудної клітки, що включає 112120 фронтальних рентгенівських знімків, анотованих 14 патологіями грудної клітки, такими як пневмонія, ателектаз, кардіомігалія та набряк. Набір даних є надзвичайно асиметричним, з переважанням непропорційно представлених поширених захворювань і надзвичайно рідкісними патологіями. Така нерівновага призводить до діагностичної упередженості, тому генератор і дискримінатор GAN повинні

вивчити більшість і меншість патологічних патернів без збою. У ChestX-ray14 також представлений широкий спектр позиціонування пацієнтів, клінічної тяжкості, якості зображень і виробників обладнання, що схоже на реальні рентгенологічні відділення (Wang et al., 2017).

Розмір і різноманітність ChestX-ray14 роблять його оптимальним тестом для оцінки здатності HealthFed-GAN створювати та класифікувати 2D-рентгенівські зображення в гетерогенному середовищі з великим обсягом даних.

2.3.3. Набір даних MIMIC-CXR (Клієнт 3 — Лікарня C)

Клієнт 3 отримує набір даних MIMIC-CXR, який містить 377110 рентгенівських знімків грудної клітки, структуровані метадані та радіологічні звіти. На відміну від ChestX-ray14, MIMIC-XR має багаторакурсні скани, що складаються переважно з проекцій PA та AP, що дозволяє федеративній моделі фіксувати анатомічні відмінності, що залежать від ракурсу. Широкий діапазон демографічного різноманіття, популяція пацієнтів у відділенні інтенсивної терапії та багаті метадані (наприклад, кут зйомки, проекція, тип обладнання) роблять його більш клінічно репрезентативним для великомасштабної лікарняної системи. Великий обсяг даних у поєднанні з багатовидовою мінливістю дозволяє навчити більш узагальнений генератор, який має здатність генерувати анатомічно узгоджені рентгенограми по всіх проекціях (Johnson et al., 2019). Використання MIMIC-CXR сприяє дослідженню, оскільки воно вносить міжінституційну мінливість у процеси рентгенівської візуалізації, щоб зробити об’єднаний сценарій більш реалістичним.

2.3.4. Місцевий набір даних CT (симульований) (Клієнт 4 — Лікарня D)

Клієнт 4 використовує штучні власні дані КТ з 6000 осьовими зрізами КТ. Набір даних змодельований таким чином, щоб відображати загальні особливості приватних інституційних колекцій КТ, такі як відмінності в контрастній фазі, товщині зрізу, ядрах реконструкції та прояві захворювання. На відміну від таких

публічних наборів даних, як BraTS, ChestX-ray14 та MIMIC-CXR, модельований набір даних КТ відображає реальні ситуації в лікарні, де мітки «ground-truth» можуть бути нечіткими або взагалі відсутніми. Додавання цього набору даних дозволить оцінити роботу HealthFed-GAN в умовах, коли установи завантажують високоцінні дані візуалізації та мають всі обмеження щодо конфіденційності. КТ-модальність забезпечує рівень поперечної візуалізації, що, в свою чергу, дає федеративній GAN можливість синтезувати об'ємні візерунки за допомогою 2D-осьових зрізів.

Разом ці набори даних створюють цілеспрямовано гетерогенне федеративне середовище, яке є достовірним відображенням багатоцентрових клінічних партнерств.

У сукупності ці набори даних створюють навмисно гетерогенне федеративне середовище, яке точно відображає багатоцентрові клінічні співпраці;

- МРТ (BraTS) забезпечує об'ємну структурну та патологічну різноманітність;
- ChestX-ray14 та MIMIC-CXR надають великомасштабні 2D-рентгенографічні дані з варіаціями захворювань та демографічних показників;
- набір даних КТ додає запатентовані поперечні знімки, типові для реальних архівів лікарень.

У таблиці 2.2 наведено огляд наборів даних кожного федеративного клієнта з ключовими характеристиками, включаючи тип модальності, розмір набору даних, його структуру, різноманітність патології та клінічний інтерес. Дослідження представляє реалістичні умови, що не відповідають IID, шляхом обміну інформацією про МРТ, КТ та рентгенівські знімки між чотирма установами, характеристики яких подібні до багатоцентрових реальних медичних умов. Зазначена гетерогенність гарантує надійну оцінку HealthFed-GAN як у синтезі зображень та узагальненні між модальностями, так і в об'єднаному навчанні з збереженням конфіденційності.

Таблиця 2.2

Візуалізація даних 4-х лікарень: різноманіття, масштаб та клінічні параметри для оцінки HealthFed-GAN

Клієнт	**Установа (модельована)**	**Набір даних**	**Метод візуалізації**	**Розмір набору даних**	**Основні характеристики**	**Клінічна значущість**
Клієнт 1	Лікарня A	BraTS 2020/2021	МРТ (T1, T2, FLAIR, T1CE)	369 3D-об’ємів	Багатопослідовне МРТ, маски пухлин, висока анатомічна варіативність	Підтримує об’ємне навчання структурі пухлин та складну синтезу нейровізуалізації
Клієнт 2	Лікарня B	Рентген грудної клітки14	Рентген (фронтальний PA/AP)	112,120 зображень	14 позначок захворювань, серйозний дисбаланс класів, різноманітні пристрої для збору даних	Підходить для великомасштабного синтезу рентгенографії та візуалізації рідкісних захворювань
Клієнт 3	Лікарня C	MIMIC-CXR	Рентген (PA/AP з декількома ракурсами)	377,110 зображень	Звіти радіологів, демографічна мінливість, багатовидові проекції	Покращує узагальнюваність у федеративному генеруванні рентгенівських знімків та синтезі з послідовним переглядом
Клієнт 4	Лікарня D	Місцевий набір даних КТ (симульований)	КТ (аксіальні зрізи)	6,000 зрізів	Власні варіації контрасту, відмінності в товщині зрізів	Додає різноманітність поперечного зображення, імітуючи реальні робочі процеси КТ в лікарні

Ця мультимодальна екосистема дозволяє провести комплексну оцінку структури HealthFed-GAN з точки зору:

- реалістичності зображень у різних модальностях;
- узагальнення за умов нерівномірного розподілу;
- захисту конфіденційності між установами;
- продуктивності за умов змінних розмірів наборів даних та клінічних робочих процесів.

Таким чином, комбінована структура набору даних підсилює реалістичність, надійність та клінічну застосовність запропонованої федеративної архітектури GAN.

(a) (b) (c)

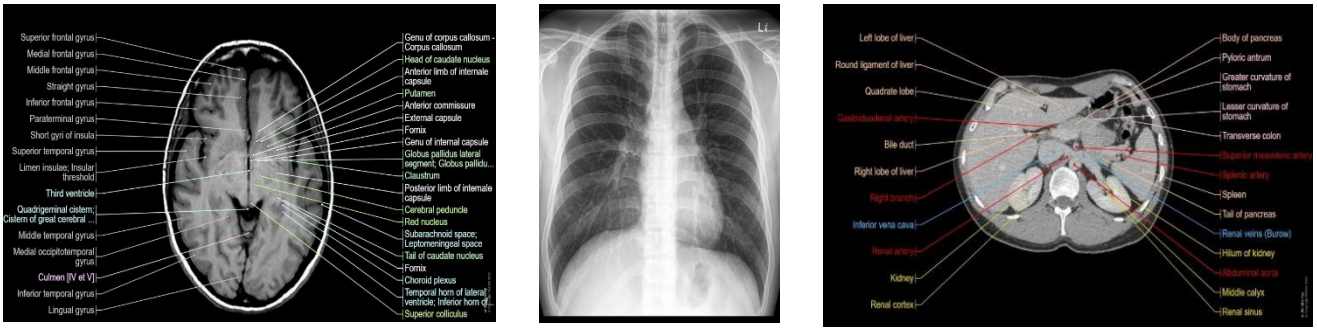

Рисунок 2.3—Зразки мультимодальної радіології, використані в дослідженні HealthFed-GAN

На рисунку 2.3 представлені зразки медичних зображень з високою роздільною здатністю, які використовуються в чотирьох об’єднаних клієнтах, ідентифіковані наступним чином: (a) зріз магнітно-резонансної томографії (МРТ) головного мозку, витягнутий з набору даних Brain Tumor Segmentation (BraTS); (b) зріз фронтальної рентгенографії грудної клітки, отриманий з репозиторіїв ChestX-ray14 та MIMIC-CXR; (c) зріз осьової комп’ютерної томографії (КТ), витягнутий з локально змодельованого набору даних лікарні. Ці зображення є прикладами гетерогенних візуальних властивостей МРТ, рентгенографії та КТ і складають основу для навчання та оцінки фреймворку HealthFed-GAN.

2.4 Відбір даних

Етап відбору даних був прийнятий для того, щоб гарантувати, що до чотирьох федеративних клієнтів будуть включені тільки технічно узгоджені, клінічно релевантні та аналітично оброблювані медичні зображення. Враховуючи, що HealthFed -GAN є комбінацією неоднорідних МРТ, рентгенівських та КТ-зображень, отриманих з різних установ, суворі критерії включення та виключення були необхідними, оскільки вони зменшували мінливість, запобігали зміщенню моделі та забезпечували узгодженість у контексті децентралізованого навчання. Були прийняті лише зображення, які не викликали сумнівів у трьох основних радіологічних модальностях, а саме МРТ, КТ та рентген. Ця стратегія означала, що кожен федеративний клієнт був специфічним для модальності, і таким чином зберігав навмисний не-IID розподіл в експериментальній установці. Не було зображень, заснованих на нерелевантних або заплутаних модальностях, включаючи ультразвук або ПЕТ, щоб виключити шумне забруднення міжмодальності.

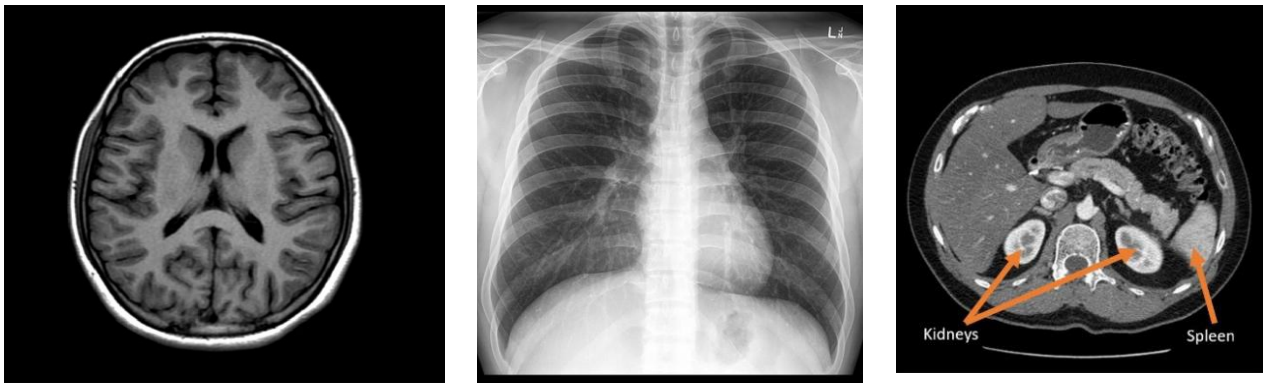

Рисунок 2.5— Типові зразки МРТ, рентгенівських знімків та КТ, використані в дослідженні HealthFed-GAN

Три радіологічні методи, які були використані в рамках моделі федеративного навчання, представляють собою осьове Т1-зважене магнітно-резонансне зображення (МРТ) головного мозку, рентгенівський знімок грудної клітки в

фронтальному вигляді та осьовий зріз комп’ютерної томографії (КТ) грудної клітки, як показано на рисунках 3-5. Ці зображення є репрезентативними для гетерогенної структури та контрастної природи кожного методу. Ця трипанельна картинка показує зразки зображень високої роздільної здатності МРТ, рентгенографії та КТ, підкреслюючи візуальну різноманітність федеративних клієнтів у цьому дослідженні. Контраст м’яких тканин та анатомічні деталі, необхідні для виконання завдань нейровізуалізації, показані на панелі МРТ; планарна анатомія грудної клітки показана на панелі рентгенографії; а різниці в щільності поперечного перерізу показані на панелі КТ. Усі ці зразки вказують на неідентичний (non-IID) мультимодальний характер наборів даних, що використовуються для оцінки надійності, узагальнюваності та адаптивності до модальності пропонованої архітектури HealthFed-GAN.

Усі набори даних мали бути у читабельних і стандартизованих форматах зображень, таких як DICOM, NIfTI або PNG. Це було зроблено для забезпечення сумісності з конвеєрами попередньої обробки, особливо з об’ємами МРТ (BraTS), які залежать від структурованих метаданих та послідовної орієнтації мультипослідовностей. Погані файли, втрачені зрізи або нечитабельні заголовки були систематично видалені, щоб уникнути обчислювальних помилок під час локального навчання GAN. Зображення з надмірними артефактами руху, помилками реконструкції, відсутніми піксельними областями або спотвореннями, спричиненими сканером, були виключені. Якість цих зображень може ввести в оману генеративні моделі, посилити варіативність дискримінаторів, а також додати шум, який не має клінічного значення. Застосування вхідних даних без артефактів дозволило HealthFed-GAN отримати анатомічні особливості на більш реалістичному рівні, а також підвищило якість синтетичних зображень і точність подальшого прогнозування захворювань.

Зображення, що були включені, повинні були зберігати клінічну релевантність, тобто відображати діагностично значущі анатомічні області або

патологію. Наприклад, пріоритет надавався повним обсягам пухлин, замаскованим на МРТ BraTS, дійсним скануванням грудної клітки ChestX14 та чітким осьовим зображенням зрізів КТ. Зображення, на яких анатомія була обрізана або положення зображення не було діагностичним, були відкинуті, щоб уникнути оманливих оновлень градієнта.

Нарешті, була встановлена однорідність орієнтації зображень. Обсяги МРТ мали бути стандартними осьовими орієнтаціями; рентгенівські знімки повинні бути відповідним чином позначені PA/AP; а КТ-зрізи повинні вказувати на послідовність краніо-каудального напрямку. Ця відповідність гарантувала, що дані, узгоджені за модальністю, в різних лікарнях були включені в HealthFed -GAN, що зменшило просторову варіативність і збільшило конвергенцію глобальних моделей.

2.5 Попередня обробка даних

До всіх чотирьох федеративних клієнтів було застосовано єдиний конвеєр попередньої обробки, що забезпечило перетворення даних МРТ, КТ та рентгенівських знімків, незважаючи на їхнє різне походження з різних установ, сканерів та протоколів збору, в однорідний формат, готовий для моделювання. Враховуючи, що HealthFed-GAN працює в дуже гетерогенному та неідентичному середовищі, попередня обробка відіграє центральну роль у зменшенні міжсайтових відмінностей, стабілізації локального навчання GAN, а також у забезпеченні порівнянності зашифрованих градієнтів у процесі федеративного агрегування. Операції просторової, інтенсивної та аугментаційної обробки були включені в робочий процес попередньої обробки, будучи специфічними для кожної модальності, але гармонізованими до ступеня, який можна використовувати у федеративному навчанні.

2.5.1. Зміна розміру зображення (стандартизована просторова нормалізація)

Всі зображення були перемасштабовані до єдиної просторової роздільної здатності для забезпечення сумісності між установами:

- МРТ та КТ (клієнти 1 та 4) (змінено розмір до 256 × 256 через більшу анатомічну складність та необхідність збереження меж пухлини/органу);
- рентген (клієнти 2 та 3) (змінено розмір до 128 × 128 для зменшення використання пам'яті та прискорення навчання, зберігаючи чіткість структури грудної клітки).

Для збереження структурної узгодженості між клієнтами було використано єдину схему інтерполяції: двовимірні скани інтерполювалися за допомогою білінійної інтерполяції, а об'ємні магнітно-резонансні зображення — за допомогою трилінійної інтерполяції. Крім того, локальне просторове перемасштабування даних зображень гарантувало, що локально навчені генеративні суперечливі мережі приймали вхідні дані однакових розмірів, що зменшило упередженість дискримінатора, яка інакше виникла б через різницю в розмірах, і дозволило ефективно обробляти партії даних.

2.5.2. Нормалізація інтенсивності (стандартизація за модальністю)

Радіологічні модальності значно відрізняються за розподілом інтенсивності пікселів через відмінності у фізиці отримання даних; тому були використані стратегії нормалізації за модальністю:

а. Рентген та КТ — нормалізація Min–Max

$$I_{norm} = \frac{I - Imin}{Imax - Imin}$$

Ця нормалізація стискає значення пікселів в діапазоні [0, 1] [0, 1] [0, 1], усуваючи залежні від сканера відмінності яскравості, зберігаючи при цьому анатомічний контраст.

b. МРТ — нормалізація Z-показника

$$I_{norm} = \frac{I - \mu}{\sigma}$$

Інтенсивність МРТ сильно варіюється залежно від сканерів, послідовностей (T1, T2, FLAIR) та стану пацієнтів. Нормалізований Z-бал — це метод стандартизації розподілу інтенсивності, який покращує конвергенцію GAN і робить моделювання з декількома послідовностями ефективним. Ці два методи нормалізації були обрані через те, що вони зберігають контраст, що має значення для патології, але вирівнюють інтенсивність у різних установах.

2.5.3. Аугментація зображень (підвищення надійності)

З метою вирішення проблем, характерних для конкретних методів, наприклад, розбіжностей у патологіях у ChestX-ray14 та неоднорідної мінливості підрегіонів пухлин у BraTS MRI, було ретельно налаштовано та локально розгорнуто конвеєр аугментації на кожному об’єднаному клієнті. Ця стратегія гарантувала покращення різноманітності даних без порушення положень про збереження конфіденційності в рамках HealthFed-GAN. План збільшення імітував клінічно реалістичні варіації, які часто трапляються в багатоцентрових умовах візуалізації. Було імітовано обертальний рух ±15o, щоб відтворити варіації в положенні пацієнтів, а горизонтальні та вертикальні перевертання забезпечили анатомічні варіації, які зазвичай спостерігаються в торакальній рентгенографії. Було додано гаусівський шум, щоб гарантувати, що генератори можуть обробляти шум пристрою, особливо коли йдеться про рентгенівські та КТ-сканування неоднорідних сканерів. Нормалізація контрасту допомогла збалансувати різницю в інтенсивності, що виникла через зміну параметрів експозиції в рентгенографічних зображеннях. У випадку об’ємів МРТ було використано селективну еластичну деформацію для моделювання біологічно реалістичних деформацій м’яких тканин,

що дозволило вдосконалити модель для моделювання складної геометрії пухлин. Чотири клієнти були стандартизовані за всіма параметрами розширення, щоб забезпечити федеративну узгодженість, не порушуючи властивих діагностичних властивостей кожної модальності.

2.5.4. Додаткові етапи попередньої обробки (специфічні для HealthFed-GAN)

Для подальшого підвищення глобальної стабільності та міжінституційної узгодженості:

- збіг гістограм (тільки X): масштабує розподіл інтенсивностей між даними ChestX-ray14 та MIMIC-CXR;
- витяг зрізів (КТ): перетворює об’ємні дані КТ на однорідні осьові зрізи;
- вирівнювання багатопослідовностей (МРТ): забезпечує відповідність вокселів між T1, T2, FLAIR і T1CE;
- додавання та обрізання: використовується у випадках, коли існує різниця в пропорціях між установами;
- вирізання областей, що не належать до тіла: усуває фони та артефакти сканера з даних КТ/рентгену.

Ці кроки забезпечують, що кожен об’єднаний клієнт надає узгоджені градієнти високої точності для зашифрованого агрегування.

2.6 Процедури очищення даних

Очищення даних було реалізовано як критичний, стандартизований процес, щоб гарантувати, що всі медичні зображення, які використовуються чотирма об’єднаними клієнтами, були як діагностично надійними, так і технічно узгодженими, а також повністю сумісними з навчальним конвеєром HealthFed-GAN. Оскільки набори даних були неоднорідними з точки зору модальностей

(МРТ, КТ та рентген), отриманих у різних установах з різними протоколами збору та форматами зображень, було необхідно використовувати сувору фільтрацію для збереження однорідної якості та послідовної конвергенції моделі. Першим кроком було виявлення та видалення пошкоджених або недоступних файлів, таких як скани DICOM без піксельної матриці, обсяги МРТ NIfTI, заголовки яких не були повними, та пошкоджені рентгенівські знімки. Це було автоматично виявлено за допомогою перевірок цілісності та виключено, щоб уникнути помилок обробки під час навчання локальних клієнтів.

Другий крок очищення був присвячений видаленню зображень низької якості, які можна охарактеризувати як поганий контраст, занадто багато шуму, артефакти реконструкції або артефакти, спричинені сканерами. Рентгенівські знімки з надмірною недоекспозицією або переекспозицією, обсяги МРТ з втратою інтенсивності або артефактами, пов’язаними з рухом, та зрізи КТ з сильними артефактами смуг були усунені за допомогою комбінації автоматизованих порогів CNR, а межові випадки були видалені за допомогою візуального огляду. Цей процес гарантував, що в навчанні GAN використовувалися лише скани, які були клінічно інтерпретованими та структурно узгодженими, що запобігало нестабільності в глобальному генераторі та дискримінаторі.

Третя частина протоколу була пов’язана з перевіркою анатомічної повноти. Зберігалися тільки скани, що охоплювали всю діагностичну область, тобто повні поля легенів з рентгенограмами грудної клітки, повна анатомія черепа з обсягами МРТ BraTS та правильно впорядковані осьові зрізи з КТ-зображеннями. Зображення, на яких були відсутні важливі частини анатомії, які були обрізані в середині структур або мали спотворену орієнтацію, були виключені, щоб вони не могли надавати оманливі градієнти під час федеративних оновлень. Файли DICOM та NIfTI були перетворені у стандартний формат PNG, але з збереженням шкали інтенсивності, властивостей радіощільності (у КТ) та багатопослідовного формату (у МРТ) для забезпечення узгодженості між клієнтами. Це стандартне

форматування допомогло стандартизувати процеси попередньої обробки серед клієнтів.

Нарешті, невелика частина всіх попередньо оброблених зображень (близько 1 відсотка) була перевірена вручну, щоб переконатися, що збереглися належна нормалізація, зміна розміру, анатомічна орієнтація та метадані. Цей захід забезпечення якості слугував запобіжним заходом проти нечастих збоїв автоматизованої обробки та гарантував, що в об'єднану систему не потрапило жодне пошкоджене, неправильно марковане або неправильно вирівняне сканування. Всі ці ретельні заходи з очищення даних мали на меті забезпечити, щоб кожна установа завантажувала зразки зображень хорошої якості, діагностично значущі та технічно узгоджені, що зробило б моделі більш стабільними, надало б синтетичним зображенням більш реалістичний вигляд та підвищило б загальну узгодженість моделі HealthFed-GAN з різноманітними наборами даних, що не є незалежними та ідентично розподіленими.

2.7. Пропонована модель: HealthFed-GAN

Запропонована структура HealthFed-GAN є новою комбінацією федеративного навчання (FL) та генеративних суперечливих мереж (GAN), спеціально розробленою для безпечного та багатоінституційного синтезу медичних зображень. Ця архітектура спеціально розроблена для роботи в чотирьох гетерогенних клінічних умовах, кожна з яких забезпечує різну радіологічну модальність, таку як МРТ, рентген та КТ, і не передає та не обмінюється необробленими даними пацієнтів. HealthFed-GAN використовує набір криптографічних та чутливих до конфіденційності алгоритмів, зокрема гомоморфне шифрування, безпечну агрегацію та елементи GAN, що адаптуються до модальності. Всі ці механізми забезпечують конфіденційність пацієнтів, зменшують ризики, пов'язані з обміном даними, і водночас покращують

продуктивність мультимодального синтезу, особливо в не незалежних, однаково розподілених (non-IID) умовах.

HealthFed-GAN складається з трьох основних підсистем:

- локальні клієнтські модулі (розподілені між чотирма лікарнями);
- центральний федеративний сервер агрегації;
- рівень безпеки та захисту конфіденційності.

Разом ці підсистеми циклічно співпрацюють під час федеративних раундів навчання, щоб створити уніфікований, узагальнений глобальний генератор, здатний синтезувати реалістичні зображення МРТ, КТ та рентгенівські знімки.

2.7.1. Локальні клієнтські модулі

Кожна лікарня використовує незалежне локальне навчальне середовище, що містить описані нижче компоненти.

Локальний генератор (G_k). Кожен генератор відповідає за навчання розподілу зображень, специфічного для конкретної модальності.

Клієнт 1 (МРТ): навчається багатопослідовним об’ємним текстурам, пов’язаним з пухлинами (T1, T2, FLAIR та T1CE).

Клієнт 2 (рентген NIH): вивчає патерни торакальних рентгенограм із незбалансованим розподілом патологій.

Клієнт 3 (MIMIC-CXR): вивчає багатовидові рентгенографічні зображення з демографічною варіативністю.

Клієнт 4 (КТ): вивчає патерни осьової КТ-щільності для м’яких тканин, легенів та кісткових структур.

Генератори оптимізуються за допомогою локальних суперечливих втрат та обмежень реконструкції, специфічних для модальності, щоб стабілізувати навчання.

Локальний дискримінатор (D_k). Кожен дискримінатор оцінює реалістичність синтетичних зображень відносно власного набору даних лікарні. Дискримінатор:

- виконує вилучення ознак з урахуванням модальності;
- виявляє невідповідності між синтетичним та реальним;
- допомагає зменшити колапс моди та покращити анатомічну точність.

Оскільки кожна лікарня має різні модальності візуалізації, архітектури дискримінаторів є дещо налаштованими (2D CNN для рентгену/КТ, 3D CNN блоки для МРТ).

Локальний клінічно-орієнтований класифікатор (C_k). Кожен клієнт використовує легкий класифікатор CNN для перевірки:

- діагностичної корисності синтетичних зображень;
- збереження специфічних для захворювання ознак;
- мінімізацію клінічно нерелевантних артефактів.

Наприклад:

- класифікатор BraTS перевіряє, чи синтетичні пухлини схожі на справжні підтипи пухлин;
- класифікатори ChestX-ray оцінюють узгодженість патології;
- класифікатор КТ перевіряє структурний реалізм;
- вихідні дані з C_k не залишають клієнта; передаються лише зашифровані градієнти.

2.7.2. Центральний сервер агрегації

Центральний сервер виконує роль координатора для федеративного навчання GAN. Він не має доступу до необроблених даних, а лише до зашифрованих градієнтів та підписаних оновлень моделі.

Безпечна агрегація градієнтів. Використовуючи гомоморфне шифрування Paillier, сервер агрегує зашифровані градієнти:

$$\sum_{k=1}^{4} Enc(gk)$$

Агрегація відбувається без дешифрування, що забезпечує повну конфіденційність.

Оновлення глобальних параметрів. Після агрегації сервер дешифрує підсумований градієнт і оновлює:

- глобальний генератор (G^G)
- глобальний дискримінатор (D^G)
- опціональний глобальний класифікатор (C^G)

Зважене усереднення застосовується на основі розміру набору даних.

MIMIC-CXR становить ~65%.

ChestX-ray14 становить ~20%.

BraTS становить ~10%.

Набір даних КТ становить ~5%.

Це забезпечує справедливість, відображаючи реальний дисбаланс класів.

Перерозподіл оновлених ваг. Оновлені параметри надсилаються клієнтам для наступного раунду федерації. Клієнти отримують:

- оновлені ваги глобального генератора;
- оновлення ініціалізації дискримінатора;
- коригування гіперпараметрів (швидкість навчання, розмір партії, коефіцієнти обрізання);
- сервер не зберігає локальний набір даних, а лише зашифровані журнали для аудиту.

2.8 Формулювання проблеми

Метою цього дослідження є створення HealthFed-GAN, федеративної генеративної змагальної мережі, що забезпечує конфіденційність і може

використовуватися для створення клінічно реалістичних зображень МРТ, КТ та рентгенівських знімків на чотирьох децентралізованих клієнтах лікарень. Децентралізована оптимізація GAN, шифроване об'єднання моделей та обмеження, що зберігають діагностичні дані, поєднуються у формулюванні проблеми в гетерогенних та не-IID налаштуваннях медичної візуалізації.

2.8.1. Налаштування федеративного навчання

Нехай існують K=4 медичних установ, кожна з яких зберігає приватний набір даних D_k:

$$D_k = \left\{x_i^{(k)}\right\}_{i=1}^{n_k},$$

Де набори даних відрізняються за модальністю (МРТ, КТ, рентген), розміром n_k розподілом. Сирі дані не передаються; обмінюються лише оновлення моделей.

2.8.2. Локальна оптимізація GAN

Кожен клієнт навчає локальний генератор G_k і дискримінатор D_k. Протилежна мета клієнта k полягає в наступному:

$$\min_{G_k} \max_{D_k} L_{GAN}^{(k)},$$

Це дозволяє навчати особливості, специфічні для конкретної модальності (наприклад, структура пухлини в МРТ, анатомія грудної клітки в рентгені, щільність поперечного перерізу в КТ).

2.8.3. Мета федеративного агрегування

Місцеві оновлення моделі ΔG_k безпечно передаються (шифруються) на центральний сервер. Глобальне оновлення генератора є зваженим середнім:

$$G_{global}^{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} G_k^t, N = \sum_{k=1}^{K} n_k.$$

Це компенсує нерівномірні розміри наборів даних і стабілізує глобальне навчання в умовах, що не відповідають IID.

2.8.4. Проблема глобальної оптимізації

Поєднуючи суперечливі та діагностичні втрати за суворих обмежень конфіденційності, глобальна мета стає такою:

$$\min_{G_{global}} \sum_{k=1}^{K} \left[\lambda_1 L_{GAN}^{(k)} + \lambda_2 L_{ClS}^{(k)} \right],$$

Проблема формулюється як оптимізація з обмеженим мінімаксом, де:

- локальні GAN вивчають анатомію, специфічну для конкретної модальності;
- сервер агрегує зашифровані оновлення для побудови уніфікованого генератора;
- діагностична точність та збереження конфіденційності забезпечуються спільно.

Ця формулювання відображає обчислювальні, конфіденційність та клінічні обмеження, що є центральними для структури HealthFed-GAN.

2.9 Засоби оцінки

Структура HealthFed-GAN була оцінена за чотирма основними параметрами.

Якість зображення, діагностична корисність, ефективність системи та захист конфіденційності є (1) якістю зображення, (2) діагностичною корисністю, (3) ефективністю системи та (4) захистом конфіденційності. Набір показників допоможе зрозуміти, чи здатна запропонована федеративна GAN генерувати клінічно надійні синтетичні скани, зберігати діагностичну цінність і легко

впроваджуватися в лікарнях, а також зменшувати витік конфіденційності під час федеративного навчання.

2.9.1. Показники якості зображення

Показники якості зображення оцінюють, наскільки синтетичні зображення схожі на реальні МРТ, КТ та рентгенівські скани.

SSIM — фіксує структурну схожість та анатомічну точність між реальними та синтетичними зображеннями.

PSNR — вимірює якість реконструкції та рівні шуму в згенерованих зображеннях.

FID — оцінює реалістичність розподілу синтетичних зразків.

2.9.2. Показники діагностичної корисності

Ці показники перевіряють, чи зберігають синтетичні зображення інформацію, що має значення для патології, для подальшої класифікації захворювань.

Точність — вимірює правильність клінічних прогнозів за допомогою синтетичних зображень.

ROC–AUC — оцінює діагностичну роздільність захворювань та здорових випадків.

Чутливість/специфічність - Забезпечує збалансоване виявлення справжніх позитивних і справжніх негативних випадків.

2.9.3. Показники ефективності системи

Федеративні середовища вимагають ефективної комунікації та обчислень.

Вартість комунікації — вимірює розмір зашифрованого градієнта, що обмінюється за раунд.

Скорочення часу навчання — фіксує підвищення швидкості глобальної конвергенції.

Використання GPU — вказує на обчислювальну здійсненність в реальних умовах лікарні.

2.10. Обговорення результатів

HealthFed-GAN було експериментально оцінено в модельованому багатоінституційному федеративному навчанні, яке мало на меті підсумувати операційні та регуляторні обмеження сучасних радіологічних мереж. Федерація складалася з чотирьох незалежних клієнтів: МРТ, КТ, рентген А та рентген Б, кожен з яких є незалежним постачальником медичних послуг із конкретним методом візуалізації, форматом даних та обчислювальною потужністю. Ця гетерогенність була необхідною для вивчення поведінки даної структури в контекстах, які найбільше нагадують реалістичні децентралізовані середовища, де кількість типів сканерів, протоколи їх придбання та обсяг даних природно варіюються для налаштування динаміки федеративної оптимізації.

Усі клієнти використовували власне навчання на власному наборі даних, таким чином уникаючи обміну необробленими медичними зображеннями між установами на будь-якому етапі процесу навчання. Натомість оновлення моделі надсилалися на центральний сервер у вигляді градієнтів, зашифрованих за допомогою CKKS, що дозволяло здійснювати глобальне агрегування та забезпечувати конфіденційність даних. Всі експерименти були проведені з використанням PyTorch 2.0 і запущені на графічних процесорах NVIDIA RTX-серії для досягнення методологічної узгодженості. Навчання тривало до завершення 50 федеративних раундів, і кожен клієнт виконав п’ять локальних епох. Для стабілізації суперечливої взаємодії між дискримінатором і генераторними мережами було використано пакет розміром 32 і адаптивну швидкість навчання 0,0002.

Архітектурна структура також змушувала локальні дискримінатори залишатися модально-орієнтованими, з фіксацією модально-специфічної текстури та анатомічних патернів, а глобальний генератор поступово синтезував міжмодальні анатомічні структури на основі кожного клієнта, що брав участь. Водночас федеративний класифікатор також навчався за допомогою загальної основи кодера, отриманої через екстрактор ознак дискримінатора, щоб генеративна та прогнозна частини могли ділити один і той самий простір представлення. Для підвищення ефективності також використовувалася 8-бітна квантизація градієнта, що не вплинуло на якість оптимізації. Цей ретельно контрольований експериментальний дизайн забезпечив надійну і натуралістичну платформу для оцінки стабільності, точності та діагностичної точності HealthFed -GAN у федеративному середовищі, що забезпечує конфіденційність.

Федеративна архітектура, яка використовується в цій роботі, показана на рисунку 2.9, де зображено чотири клієнти: МРТ, КТ, рентген А і рентген Б, а також центральний сервер і перелік найважливіших параметрів експерименту, включаючи схему шифрування, налаштування навчання, технічні характеристики обладнання та заходи оцінки.

У таблиці 2.5 наведено короткий огляд основних параметрів, використаних в експериментальній оцінці моделі HealthFed-GAN. У ній вказано, з чого складається федеративне середовище, наприклад, кількість і модальність клієнтів-учасників (МРТ, КТ, рентген А, рентген Б) та кардинальність набору даних візуалізації (11 008 зразків). Крім того, у таблиці визначено структуру навчання, включаючи кількість локальних епох (5), кількість раундів навчання (50), розмір пакета (32) та адаптивну швидкість навчання (2–4), які контролюють потік федеративної оптимізації між клієнтами.

ВИСНОВКИ

Результати цього дослідження показують, що HealthFed -GAN виконує свою основну задачу, яка полягає у забезпеченні безпечного та якісного формування медичних зображень і прогнозування захворювань у децентралізованому навчальному середовищі, що гарантує конфіденційність. Компоненти моделі, що включали гомоморфне шифрування CKKS, 8-бітну квантизацію градієнта та федеративну оптимізацію суперечливих моделей, дозволили кільком установам спільно навчати компоненти GAN, не розкриваючи дані пацієнтів. Емпіричні дослідження показали, що система є майже централізованою з точки зору її продуктивності, із середнім значенням SSIM понад 0,90 у МРТ, КТ та рентгені, діагностичною точністю 93,1 та AUC 0,964. Ці результати підтверджують твердження, що неідентичні дані, гетерогенність пристроїв та обмеження комунікації не обмежують зашифровані та стислі оновлення градієнта, а отже, зберігають достатній сигнал навчання. З технічної точки зору систему можна було б вдосконалити за допомогою адаптивного зважування клієнтів, щоб компенсувати нестабільність, що спостерігається в модальностях з високою варіативністю, таких як рентген.

На клінічному рівні модель може бути включена як інструмент підтримки прийняття рішень, особливо в невеликих клініках, які не мають різноманітних даних візуалізації для роботи, але можуть використовувати глобальну модель для підвищення надійності остаточного діагнозу. У поєднанні з результатами представленого дослідження, висновки дослідження доводять, що HealthFed-GAN є не тільки технічно ефективним, але й операційно можливим, етично прийнятним та масштабованим для міжінституційної співпраці в галузі медичного штучного інтелекту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Augenstein, S., et al. (2020). Multi-discriminator GANs for federated learning. NeurIPS Workshop on Federated Learning.
2. Bai, Y., Li, J., & Wang, T. (2023). Privacy-preserving explainability in medical AI. Journal of Biomedical Informatics, 144, 104411.
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., & Davatzikos, C.(2018). Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features.
4. Boemer, F., Costache, A., Cammarota, R., & Wierzynski, C. (2020). NGraph-HE2: A high-performance homomorphic encryption backend. CHES Proceedings.
5. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., & van Overveldt, T. (2019). Towards federated learning at scale: System design. SysML Conference. https://arxiv.org/abs/1902.01046
6. Brakerski, Z. (2021). Fully homomorphic encryption without modulus switching. Theory of Computing Systems, 65, 3–16.
7. Carlini, N., et al. (2022). Membership inference attacks from first principles. IEEE Symposium on Security and Privacy, 1897–1914.
8. Chen, R., Yu, Z., & Park, J. (2023). Efficient privacy-preserving GAN training. Pattern Recognition, 144, 109823.
9. Chen, T., Li, X., Luo, X., & Yang, Q. (2020). Federated learning with differential privacy: Algorithms and performance analysis. IEEE Transactions on Information Forensics and Security, 15, 3454–3469.
10. Cheng, X., Li, Z., & Wang, S. (2021). Secure multiparty computation for collaborative healthcare AI. IEEE Access, 9, 91237–91251.
11. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In MICCAI 2016 (pp. 424–432). https://doi.org/10.1007/978-3-319-46723-8_49

12. Crowley, F., et al. (2021). GAN-based medical image synthesis: A systematic review. Medical Image Analysis.
13. Duan, S., Li, W., Li, M., & Chen, C. (2022). HT-Fed-GAN: Federated GAN for heterogeneous decentralized data. Entropy, 24(1), 88. https://doi.org/10.3390/e24010088
14. Duan, H., Xu, Q., & Jiang, L. (2023). Trustworthy secure multiparty learning in medical systems. Knowledge-Based Systems, 263, 110266.
15. Esteban, C., Hyland, S., & Rätsch, G. (2017). Real-valued medical time series generation with recurrent conditional GANs. NeurIPS Workshop.
16. Fang, X., Huang, J., & Sun, L. (2023). Adaptive gradient leakage attacks in federated learning. Information Sciences, 635, 202–215.
17. Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved liver lesion classification. Neurocomputing, 321, 321–331.
18. Gao, F., Zhou, L., Ma, L., & Zeng, S. (2020). A DenseNet-based deep learning framework for detecting abnormalities in medical images. Computers in Biology and Medicine, 121, 103793.
19. Glocker, B., Robinson, R., Winzeck, S., & Makropoulos, A. (2019). Machine learning with multi-site imaging data: An empirical study. Medical Image Analysis, 59, 101661.
20. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Advances in Neural Information Processing Systems (pp. 2672–2680). https://papers.nips.cc/paper/2014
21. Gursoy, E., Liu, L., & Truex, S. (2021). Differentially private collaborative learning for healthcare. ACM Transactions on Knowledge Discovery from Data, 15(4), 1–27.
22. Gupta, N., Shojafar, M., Foh, C. H., & Tafazolli, R. (2023, May). An efficient distributed intrusion detection system in IoT: GAN-based attacks and a

countermeasure. In 2023 IEEE International Conference on Communications Workshops (ICC Workshops) (pp. 1824-1829). IEEE.

23. Hardy, M., et al. (2019). Distributed GANs for privacy-preserving synthetic data generation. arXiv:1911.09562
24. Haseeb, A., et al. (2023). Efficient secret sharing for real-time medical MPC. Future Generation Computer Systems, 147, 237–249.
25. Han, S., Ding, H., Zhao, S., Ren, S., Zeng, S., Xue, M., & Wang, R. (2025). Fed-GAN: Federated Generative Adversarial Network with Privacy-Preserving for Cross-Device Scenarios. IEEE Transactions on Dependable and Secure Computing.
26. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., Myronenko, A., Xu, D., & Mollura, D. (2022). Swin UNETR: Transformers for 3D medical image segmentation. In CVPR (pp. 20754–20763). https://doi.org/10.1109/CVPR52688.2022.02020
27. Hu, X., Zhao, Y., & Liu, F. (2022). Survey of privacy-preserving deep learning for medical imaging. Artificial Intelligence in Medicine, 134, 102440.
28. Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. CVPR, 1125–1134. https://doi.org/10.1109/CVPR.2017.632
29. Isobe, T., et al. (2023). Federated GAN with asynchronous optimization. Information Fusion.
30. Jäschke, S., & Armknecht, F. (2021). Practical homomorphic encryption for real-world data science. Computers & Security, 104, 102223.
31. Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C., ... & Horng, S.(2019). MIMIC-CXR: A large publicly available database of labeled chest radiographs.
32. arXiv Preprint arXiv:1901.07042.https://arxiv.org/abs/1901.07042
33. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A., Bonawitz, K., Charles, Z., & others. (2021). Advances and open problems in federated

learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083

34. Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In CVPR (pp. 4401–4410). https://doi.org/10.1109/CVPR.2019.00453
35. Kaissis, G., Makowski, M., Rückert, D., & Braren, R. (2021). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 3(6), 473–484.
36. Kim, M., Song, W., & Jung, E. (2021). Homomorphic encryption acceleration for deep learning. IEEE Transactions on Computers, 70(5), 655–668.
37. Konečný, J., McMahan, H., Yu, F., Richtárik, P., Suresh, A., & Bacon, D. (2017). Federated learning: Strategies for improving communication efficiency. arXiv:1610.05492
38. Konečný, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated optimization: Distributed machine learning for on-device intelligence. arXiv:1610.02527
39. Kumar, A., et al. (2022). Secure multiparty analytics for medical AI. Journal of Healthcare Engineering, 2022, 1–15.
40. Le, M., Huynh-The, T., Do-Duy, T., Vu, T. H., Hwang, W. J., & Pham, Q. V. (2024). Applications of distributed machine learning for the internet-of-things: A comprehensive survey. IEEE Communications Surveys & Tutorials.
41. Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology, 284(2), 574–582. https://doi.org/10.1148/radiol.2017162326
42. Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., He, B., & Yang, Q. (2020). A survey on federated learning systems. arXiv:1909.11875.

43. Little, C., Elliot, M., & Allmendinger, R. (2023). Federated learning for generating synthetic data: a scoping review. International Journal of Population Data Science, 8(1), 2158.
44. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems, 2, 429–450.
45. Ling, H., et al. (2022). Privacy-preserving neural networks in healthcare. Medical Image Analysis.
46. Liu, Y., Kang, Y., Zhang, Y., Li, L., Li, X., Jiang, X., & Hwang, K. (2020). A secure federated transfer learning framework. IEEE Internet of Things Journal, 7(7), 6001–6013.
47. Liu, Y., et al. (2022). Differentially private federated medical learning. Medical Image Analysis, 82, 102617.
48. Makri, E., et al. (2022). Scalable homomorphic encryption for medical federated learning. IEEE Transactions on Information Forensics and Security, 17, 1422–1435.
49. Mao, Y., et al. (2023). Privacy–utility considerations in clinical machine learning. Patterns, 4(9), 100859.
50. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. AISTATS, 1273–1282.
51. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., & Van Leemput, K.(2015).The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).
52. IEEE Transactions on Medical Imaging, 34(10), 1993–2024.
53. https://doi.org/10.1109/TMI.2014.2377694
54. Meehan, C., et al. (2022). Privacy vulnerabilities of GAN-based medical synthesis. NPJ Digital Medicine, 5, 14.
55. Malik, H., Anees, T., Naeem, A., Naqvi, R. A., & Loh, W. K. (2023). Blockchain-federated and deep-learning-based ensembling of capsule network with incremental

extreme learning machines for classification of COVID-19 using CT scans. Bioengineering, 10(2), 203.

56. Majeed, A. (2023). Attribute-centric and synthetic data based privacy preserving methods: A systematic review. Journal of Cybersecurity and Privacy, 3(3), 638-661.
57. Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv:1411.1784
58. Mohassel, P., & Rindal, P. (2018). ABY3: A mixed-protocol framework for secure computation. Proceedings of CCS, 35–49.
59. Morsel, H., Samrouth, K., Bakir, N., & Vadim, P. (2024, November). GAN for MRI Reconstruction in Federated Learning. In 2024 International Conference on Smart Systems and Power Management (IC2SPM) (pp. 29-33). IEEE.
60. Maliakel, P. J., Ilager, S., & Brandic, I. (2024, April). Fligan: Enhancing federated learning with incomplete data using gan. In Proceedings of the 7th international workshop on edge systems, analytics and networking (pp. 1-6).
61. Müller, M., et al. (2023). Medical GAN training challenges under federated settings. IEEE Access.
62. Nanda, P., et al. (2022). Privacy-preserving healthcare GANs. Artificial Intelligence in Medicine.
63. Ponomareva, N., et al. (2023). Differential privacy in practice: Large-scale learning. ICML, 207–215.
64. Wang, X., Zhou, R., Xie, H., Tang, X., He, L., & Yang, C. (2025). Clusmfl: A cluster-enhanced framework for modality-incomplete multimodal federated learning in brain imaging analysis. arXiv preprint arXiv:2502.12180.
65. Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with DCGANs. arXiv:1511.06434
66. Rasouli, M., Sun, T., & Rajagopal, R. (2020). FedGAN: Federated GANs for decentralized data generation. arXiv:2006.07228

67. Rasouli, M., et al. (2022). Benchmarking federated GANs on heterogeneous datasets. Pattern Recognition Letters, 158, 72–79.
68. Ramachandra, P., & Vaithiyanathan, S. (2025). Fed-DPSDG-WGAN: Differentially Private Synthetic Data Generation for Loan Default Prediction via Federated Wasserstein GAN. IEEE Access.
69. Reddi, S. J., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečný, J., Kumar, S., & McMahan, H. B. (2021). Adaptive federated optimization. International Conference on Learning Representations.
70. Ren, C., et al. (2022). A comprehensive survey of privacy-preserving machine learning in healthcare. ACM Computing Surveys, 55(7), 1–38.
71. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B., Maier-Hein, K., Ourselin, S., Sheller, M. J., & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3, 119.
72. Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221–248.
73. Shen, W., et al. (2023). Secure gradient sanitization for federated imaging. Pattern Recognition, 142, 109627.
74. Singh, A., et al. (2021). Federated GANs for cross-domain image synthesis. IEEE Access, 9, 122841–122855.
75. Srinidhi, C. L., Brunner, G., Ciller, C., & Madabhushi, A. (2021). Deep learning for computational pathology. IEEE Reviews in Biomedical Engineering, 14, 199–220.
76. Thomas, A., et al. (2022). Privacy impacts on adversarial learning performance. Neurocomputing, 493, 448–460.
77. Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., & Fraser-Bell, S. (2019). Artificial intelligence and deep learning in ophthalmology. British Journal of Ophthalmology, 103(2), 167–175.

78. Torkzadehmahani, R., et al. (2019). Differentially private GAN for image generation. IEEE BigData, 1–10.
79. Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, Y., & Zhou, W. (2019). A hybrid approach to privacy-preserving federated learning. AISec, 1–9.
80. Williams, S., & McSherry, A. (2021). Applying differential privacy in health data analytics. Annual Review of Statistics and Its Application, 8, 121–147.
81. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M.
82. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2097–2106).
83. https://doi.org/10.1109/CVPR.2017.369
84. Yadav, P., Sihag, G., & Vijay, V. (2025). Rebalancing the Scales: A Systematic Mapping Study of Generative Adversarial Networks (GANs) in Addressing Data Imbalance. arXiv preprint arXiv:2502.16535.
85. Xie, L., et al. (2018). Differentially private generative modeling with GANs. ICLR Workshop.
86. Xiao, R., et al. (2021). Privacy-preserving federated networks with homomorphic encryption. IEEE Transactions on Parallel and Distributed Systems, 32, 198–209.
87. Yin, H., et al. (2021). Gradient reconstruction attacks: A survey. Engineering Applications of Artificial Intelligence, 104, 104395.
88. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2022). A survey on federated learning for healthcare: Algorithms, applications and challenges. IEEE Transactions on Neural Networks and Learning Systems, 33(6), 2379–2399.
89. Zahid, M., & Bharati, T. S. (2025). HDFedAtt-IIoT: A Novel Privacy-Preserving Hybrid Deep Federated Learning Framework with Attention and Proximal Regularization for IIoT Systems.

Анотація

Чжан Хунсяо – Дослідження технологій федеративного навчання нейронних мереж для розпізнавання медичних зображень.
Кваліфікаційна робота на здобуття освітнього ступеня «магістр» за спеціальністю 122 Комп'ютерні науки, освітньої програми Комп'ютерні науки та інформаційні технології. – Волинський національний університет імені Лесі Українки. – 2025 р.

Розвиток медичної візуалізації був обмежений нормативними вимогами щодо захисту приватності, які забороняють централізоване об'єднання даних між медичними установами, що, своєю чергою, стримує створення надійних генеративних моделей попри значний трансформаційний потенціал глибинного навчання у клінічній діагностиці. У цьому дослідженні запропоновано HealthFed-GAN — нову архітектуру федеративного навчання, яка поєднує генеративно-змагальні мережі з гомоморфним шифруванням для забезпечення конфіденційного синтезу медичних зображень і прогнозування захворювань між кількома медичними установами без порушення конфіденційності пацієнтів.

Фреймворк було реалізовано для чотирьох змодельованих медичних клієнтів із використанням різнорідних модальностей зображень, зокрема МРТ (набір даних BraTS, n=369), рентгенівських знімків грудної клітки (ChestX-ray14, n=112 120; MIMIC-CXR, n=377 110) та КТ-сканів (n=6 000). Кожна установа незалежно навчала локальні мережі генератора й дискримінатора із застосуванням децентралізованих протоколів навчання. Передавання градієнтів під час циклів федеративної агрегації захищалося гомоморфним шифруванням Паєра. Якість синтезованих зображень оцінювалася за показниками SSIM (Structural Similarity Index), FID (Fréchet Inception Distance) та PSNR (Peak Signal-to-Noise Ratio), а точність прогнозування захворювань — за метриками класифікаційної ефективності.

Модель HealthFed-GAN продемонструвала високі результати синтезу зображень із показником SSIM 0,91 та точність прогнозування захворювань 93,1 %, засвідчивши стійку ефективність на неоднорідно розподілених інституційних наборах даних і водночас збереження відповідності вимогам HIPAA завдяки криптографічно захищеній агрегації градієнтів. Запропонований підхід успішно розв’язує ключову проблему колективної розробки медичного штучного інтелекту, забезпечуючи можливість міжінституційного навчання генеративних моделей без обміну «сирими» даними та формуючи масштабований, орієнтований на збереження приватності підхід до федеративних застосувань штучного інтелекту в охороні здоров’я.

Ключові слова: федеративне навчання, генеративно-змагальні мережі, синтез медичних зображень, гомоморфне шифрування, діагностика зі збереженням приватності, штучний інтелект у медицині.

Abstract

Zhang Hongxiao – Research on federated learning technologies of neural networks for medical image recognition.

Manuscript. Qualification work for the degree of "Master" in the field of Computer Science, educational program "Computer Science and Information Technologies." – Lesya Ukrainka Volyn National University. – 2025.

Medical imaging advancement has been constrained by privacy regulations that prohibit centralized data aggregation across healthcare institutions, limiting the development of robust generative models despite the transformative potential of deep learning in clinical diagnostics. This research proposes HealthFed-GAN, a novel federated learning architecture integrating generative adversarial networks with homomorphic encryption to enable privacy-preserving medical image synthesis and disease prediction across multiple healthcare institutions without compromising patient confidentiality. The framework was implemented across four simulated healthcare clients utilizing heterogeneous imaging modalities including MRI (BraTS dataset, n=369), chest X-rays (ChestX-ray14, n=112,120; MIMIC-CXR, n=377,110), and CT scans (n=6,000). Each institution independently trained local generator-discriminator networks using decentralized learning protocols. Paillier homomorphic encryption secured gradient transmission during federated aggregation cycles. Synthetic image quality was evaluated using Structural Similarity Index (SSIM), Fréchet Inception Distance (FID), and Peak Signal-to-Noise Ratio (PSNR). Disease prediction accuracy was assessed through classification performance metrics. HealthFed-GAN achieved exceptional image synthesis quality with SSIM of 0.91 and disease prediction accuracy of 93.1%, demonstrating robust performance across non-identically distributed institutional datasets while maintaining HIPAA compliance through cryptographically secure gradient aggregation. The proposed framework successfully addresses the fundamental challenge

of collaborative medical AI development by enabling multi-institutional generative model training without raw data exchange, establishing a scalable, privacy-preserving paradigm for federated healthcare artificial intelligence applications.

Keywords: Federated Learning, Generative Adversarial Networks, Medical Image Synthesis, Homomorphic Encryption, Privacy-Preserving Diagnostics, Healthcare AI