

**Газдюк К.П.**<https://orcid.org/0000-0002-7568-4422>

Чернівецький національний університет імені Юрія Федьковича

**Мовченко Р.В.**<https://orcid.org/0009-0009-1761-1979>

Чернівецький національний університет імені Юрія Федьковича

## АРХІТЕКТУРА СИСТЕМИ АНАЛІЗУ РЕНТГЕНІВСЬКИХ ЗНІМКІВ З ВИКОРИСТАННЯМ GCP VERTEX AI, KUBERNETES, DATAFLOW ТА MICROSERVICE ARCHITECTURE

У статті представлено комплексний архітектурний підхід до побудови системи автоматизованого аналізу рентгенівських знімків на основі методів глибокого навчання з використанням хмарних технологій. Актуальність дослідження зумовлена стрімким зростанням обсягів медичних зображень, необхідністю скорочення часу їх обробки та дефіцитом кваліфікованих радіологів, що обмежує ефективність традиційних діагностичних процесів. Сучасні алгоритми *deep learning*, зокрема згорткової нейронної мережі та трансформерні архітектури, демонструють високу точність у виявленні патологій на медичних зображеннях, однак їх впровадження у клінічні робочі процеси потребує надійної, масштабованої та керованої обчислювальної інфраструктури.

У роботі запропоновано архітектуру системи, побудовану на основі сервісів *Google Cloud Platform*, яка охоплює повний життєвий цикл машинного навчання – від отримання та попередньої обробки рентгенівських даних до тренування, розгортання та моніторингу моделей. Платформа *Vertex AI* використовується для організації процесів навчання, версіювання та деплоювання моделей глибокого навчання, а також для контролю їх якості в процесі експлуатації. Оркестрація контейнеризованих компонентів системи реалізується за допомогою *Kubernetes*, що забезпечує автоматичне масштабування, відмовостійкість і гнучке керування ресурсами. Для обробки великих потоків рентгенівських знімків і їх підготовки до машинного навчання застосовується сервіс *Dataflow*, який підтримує як потоковий, так і пакетний режими обчислень.

Запропонована архітектура поєднує принципи мікросервісного підходу, *MLOps* та хмарних обчислень, що дозволяє забезпечити стабільну роботу системи, ефективну інтеграцію з медичними інформаційними системами та відповідність вимогам масштабованості. Отримані результати свідчать, що використання запропонованого підходу дає змогу суттєво скоротити час аналізу рентгенівських знімків, підвищити точність діагностичних рішень і адаптувати систему до змін навантаження відповідно до потреб медичних закладів. Запропоноване рішення може бути використане як архітектурна основа для створення інтелектуальних систем медичної діагностики та має потенціал для подальшого розвитку, зокрема шляхом розширення на інші типи медичних зображень.

**Ключові слова:** *Deep Learning*, рентгенівські знімки, *Google Cloud Platform*, *Vertex AI*, *Kubernetes*, *Dataflow*, мікросервісна архітектура, *MLOps*, хмарні обчислення, автоматизований аналіз зображень.

**Постановка проблеми.** Стрімке зростання обсягів медичних зображень, зокрема рентгенівських, потребує впровадження ефективних інструментів автоматизації діагностики, оскільки традиційні системи вже не здатні забезпечити необхідну швидкість та точність опрацювання даних. Наявні рішення часто мають обмежену масштабованість, недостатню інтеграцію з клінічними інформацій-

ними системами та не підтримують безперервну обробку великих потоків знімків. Основна наукова проблема полягає у досягненні високої точності моделей глибокого навчання у поєднанні з можливістю їх стабільного розгортання в умовах зростання навантаження. Саме тому особливого значення набуває використання хмарних технологій та мікросервісної архітектури, що здатні



забезпечити гнучкість, масштабованість, надійність та ефективну інтеграцію системи в реальні медичні робочі процеси

**Аналіз останніх досліджень і публікацій.** Обробка рентгенівських медичних зображень у сучасних інформаційно-обчислювальних системах становить складну науково-технічну задачу, що зумовлена як великими обсягами даних, так і високими вимогами до їх безпечного зберігання та обробки [1],[2]. Висока роздільна здатність діагностичних знімків потребує використання продуктивних графічних прискорювачів (GPU/TPU) під час тренування моделей глибокого навчання, а також надійних систем зберігання з низькими затримками доступу [3]. Крім того, медичні зображення містять чутливі персональні медичні дані (Protected Health Information, PHI), що вимагає дотримання міжнародних нормативів безпеки та конфіденційності, зокрема Закону США про перенесеність і підзвітність медичного страхування (Health Insurance Portability and Accountability Act, HIPAA) та Загального регламенту захисту даних Європейського Союзу (General Data Protection Regulation, GDPR) [4]. У цьому контексті хмарні платформи пропонують низку технологічних рішень, які забезпечують масштабовану інфраструктуру, спеціалізовані сервіси для роботи з медичними даними та інструментарій для реалізації комплексних конвеєрів експлуатації систем машинного навчання (Machine Learning Operations, MLOps) [5].

**Google Cloud** за останні роки значно розширив спектр сервісів, орієнтованих на медичну візуалізацію. У 2022 році було представлено Medical Imaging Suite – інтегрований комплекс для зберігання, стандартизованої обробки та аналізу медичних зображень [6]. До його складу входить Cloud Healthcare API з підтримкою веб-протоколу DICOMweb на основі стандарту Digital Imaging and Communications in Medicine (DICOM), інструменти деперсоналізації даних, а також засоби інтеграції з відкритим медичним фреймворком Medical Open Network for Artificial Intelligence (MONAI). Компонент Imaging Lab забезпечує готове середовище на базі інтерактивних блокнотів Jupyter та віртуальних машин глибокого навчання (Deep Learning Virtual Machine, Deep Learning VM) із попередньо інстальованими інструментами, такими як 3D Slicer з модулями MONAI Label [7], що істотно спрощує процеси анування та формування тренувальних вибірок.

Vertex AI, уніфікована платформа штучного інтелекту (Artificial Intelligence, AI) від Google,

підтримує як розробку моделей традиційним програмуванням, так і використання автоматизованого машинного навчання (Automated Machine Learning, AutoML). У низці досліджень, зокрема у роботах 2025 року, продемонстровано ефективність AutoML Vision для задач автоматизованої діагностики на рентгенограмах [8]. Повідомляється про розробку класифікатора пневмотораксу на основі Vertex AI AutoML, який автоматично виконував поділ вибірок, підготовку зображень до стандартного розміру (224×224), нормалізацію та побудову згорткової нейронної мережі з використанням трансферного навчання. Модель показала високу діагностичну ефективність (Accuracy≈0.95, F1≈0.95) та продемонструвала портативність завдяки ідентичній продуктивності як у хмарному середовищі, так і на периферійних обчислювальних пристроях.

Крім того, Google Cloud активно інтегрує результати досліджень Google Research. У 2024 році було представлено спеціалізовані великі медичні моделі (Medical Large Models, MedLM), зокрема MedLM for Chest X-ray – мультимодальну модель, оптимізовану для класифікації грудних рентгенограм та генерації діагностичних висновків [9]. Таким чином, Google Cloud пропонує комплексний набір інструментів для реалізації генеративних та аналітичних методів у медичній радіології.

**Amazon Web Services (AWS)** пропонує розвинуту екосистему для реалізації дослідницьких та промислових AI-рішень у радіології, побудовану навколо сервісу Amazon SageMaker. SageMaker забезпечує повний цикл машинного навчання – від інтерактивної розробки до масштабованого навчання моделей та їх промислового розгортання.

У 2022 році AWS представила AWS HealthImaging (AHI) – високопродуктивну хмарну систему для зберігання та пошуку DICOM-зображень з мінімальними затримками доступу [10]. Завдяки інтеграції з SageMaker формується єдиний технологічний конвеєр: від автоматизованого завантаження знімків із медичних пристроїв через шлюзи інтернету речей до проведення інференсу на розгорнутих моделях.

Amazon SageMaker пропонує різні режими виконання моделей (реального часу, пакетної обробки, потокового керування запитами), що дозволяє адаптувати рішення до конкретних клінічних сценаріїв. Для задач анотації медичних зображень використовується сервіс SageMaker Ground Truth, який забезпечує інструменти для

інтерпретації та маркування рентгенограм і підтримує інтеграцію з візуалізаторами Open Health Imaging Foundation (OHIF) та 3D Slicer. AWS також активно співпрацює з проектом MONAI, пропонуючи приклади та налаштування для тренування моделей сегментації та класифікації у хмарному середовищі.

**Microsoft Azure** забезпечує комплексні можливості для побудови систем штучного інтелекту у сфері охорони здоров'я, зокрема через платформу Azure Machine Learning (Azure ML) та модуль Azure Health Data Services. Сервіс Azure DICOM Service функціонує як хмарний архів, аналог систем архівування та передавання зображень (Picture Archiving and Communication System, PACS), забезпечуючи зберігання та передачу зображень у форматі DICOM з підтримкою географічної реплікації та стандартизованих програмних інтерфейсів (Application Programming Interface, API). Важливою перевагою Azure є нативна інтеграція між клінічними даними у форматі Fast Healthcare Interoperability Resources (FHIR) та DICOM, що створює умови для комплексного аналізу медичних даних, включаючи поєднання зображень та структурованих електронних медичних записів.

Azure ML також підтримує AutoML для задач класифікації медичних зображень. Одним із відомих прикладів є дослідження Microsoft 2020 року щодо автоматичної класифікації стадії COVID-19 за рентгенографіями грудної клітки за допомогою сервісу Custom Vision. У 2023 році в Azure AI Gallery з'явилися спеціалізовані медичні моделі, такі як CXRRreportGen (Chest X-Ray Report Generation) – мультимодальний трансформер для автоматичної генерації текстових описів рентгенограм, що демонструє потенціал Azure у розвитку когнітивних AI-помічників для радіологів.

Аналіз наявних публікацій свідчить, що провідні хмарні платформи – Google Cloud, AWS та Azure – виявляють різні сильні сторони у контексті застосування штучного інтелекту в медичній радіології:

Google Cloud вирізняється спеціалізованою підтримкою медичної візуалізації, тісною інтеграцією з екосистемою Google, доступом до TPU та потужними генеративними моделями.

AWS забезпечує високу гнучкість, модульність і виражену орієнтацію на масштабовані MLOps-конвеєри, пропонуючи широкий спектр сервісів для кастомізованих AI-рішень. Microsoft Azure відзначається глибокою інтеграцією з корпоративними інфраструктурами лікувальних установ та комплексним об'єднанням клінічних і візуальних даних. Узагальнення основних характеристик наведено в таблиці 1.

Усі три платформи демонструють здатність забезпечувати повний цикл розвитку медичних AI-систем – від прототипування до промислового впровадження. Вибір конкретного рішення залежить від наявної інфраструктури медичної установи, характеру даних та вимог до інтеграції, безпеки й масштабованості.

**Постановка завдання.** Попри значний прогрес у розвитку технологій, важливою не вирішеною проблемою є інтеграція повного циклу інфраструктури машинного навчання у робочі процеси медичних закладів. Радіологи й відділи інформаційних технологій (IT-відділи) стикаються з труднощами у підписанні контрактів з численними постачальниками AI-рішень та їх інтеграції у власне IT-середовище; брак стандартів обміну даними та відсутність єдиної архітектури ускладнюють впровадження. Значна частина закладів використовує локальні (on-premise) системи, які обмежені фізичною ємністю, вимагають трива-

Таблиця 1

**Основні технологічні можливості Google Cloud, AWS і Azure для обробки медичних зображень та побудови AI-рішень**

Критерій	Google Cloud	Amazon Web Services	Microsoft Azure
Медична візуалізація	Medical Imaging Suite, DICOMweb, MONAI	AWS HealthImaging (DICOM)	Azure DICOM Service, FHIR
ML / AI-платформа	Vertex AI	SageMaker	Azure Machine Learning
MLOps та автоматизація	Pipelines, Model Registry, Monitoring	Pipelines, Model Monitor	Pipelines, Data Drift Monitor
Інференс і масштабування	Vertex AI Endpoints, GKE, serverless	SageMaker Endpoints, EC2 GPU	Azure ML Endpoints, AKS
Ключова перевага	Спеціалізація для медичних зображень	Гнучкість і масштабованість	Інтеграція з корпоративними системами
Типове застосування	Радіологія, аналіз DICOM-архівів	Радіологія, аналіз DICOM-архівів	Корпоративні медичні платформи

лої установки нового обладнання та не здатні до автоматичного масштабування, що призводить до простоїв. Якість потоків зображень та результатів моделей часто не контролюється належним чином; літературні огляди вказують на рідкісне застосування систем технічної та клінічної перевірки й відсутність систематичних методів оцінювання якості пояснюваного штучного інтелекту (Explainable Artificial Intelligence, XAI). Крім того, глибинні моделі залишаються «чорними скриньками», що обмежує довіру лікарів; хоча XAI може частково вирішити проблему, досі не існує стандартів для його використання, що гальмує клінічні застосування. Ці проблеми потребують комплексних досліджень та розробки стандартизованих рішень у майбутньому.

Загальною метою дослідження розробка архітектурного підходу для побудови системи аналізу рентгенівських знімків, що поєднує можливості глибокого навчання з сучасними хмарними технологіями. Система повинна забезпечити масштабованість шляхом розгортання контейнеризованих компонентів у Kubernetes, що дозволить автоматичне збільшення або зменшення ресурсів відповідно до навантаження та створить основу для фінального розгортання моделей. Автоматизовані конвеєри машинного навчання (Machine Learning, ML-пайплайни) на платформі Vertex AI потрібні для навчання, оптимізації та моніторингу моделей, а також для організації повного життєвого циклу машинного навчання. Для підготовки та обробки великих масивів рентгенівських зображень слід використовувати сервіс Dataflow, який забезпечує ефективні та масштабовані потоки даних для ML-процесів. Мікросервісна архітектура забезпечить модульність, відмовостійкість (fault tolerance) та можливість швидкого оновлення окремих компонентів. Особлива увага приділяється досягненню високої точності діагностики завдяки використанню передових моделей глибокого навчання та застосуванню практик MLOps. Крім того, необхідно передбачити засоби безшовної інтеграції системи з медичними інформаційними системами та внутрішніми процесами закладів охорони здоров'я, що сприятиме широкому впровадженню та адаптації запропонованого рішення, надаючи лікарям надійний інструмент для прискорення роботи.

**Виклад основного матеріалу.** Архітектура запропонованої інформаційно-аналітичної системи побудована на принципах глибокої інтеграції методів машинного навчання з хмарними технологіями Google Cloud Platform (GCP) та

застосування мікросервісної парадигми. Система охоплює повний життєвий цикл обробки даних медичної візуалізації – від отримання та стандартизації рентгенівських знімків до їх зберігання, попередньої обробки, тренування моделей глибокого навчання, розгортання та подальшого моніторингу. Центральним компонентом архітектури є конвеєр обробки даних, реалізований засобами сервісу Dataflow на основі фреймворку Apache Beam, що забезпечує уніфікований підхід до потокової та пакетної обробки великих масивів даних.

**Завантаження та стандартизація даних.** Джерелами вхідної інформації для системи є клінічні архіви медичних зображень – системи архівування та передавання зображень (Picture Archiving and Communication System, PACS), системи обміну медичними повідомленнями у форматі Health Level Seven (HL7), а також сховища зображень у форматі DICOM. Формат DICOM, розроблений у 1993 році, забезпечує стандартизоване подання медичних зображень, включаючи структуру метаданих, мережеві протоколи та формати файлів. Для інтеграції з адміністративними інформаційними системами застосовується стандарт HL7, який формалізує процеси обміну службовими повідомленнями (наприклад, реєстрація пацієнта, направлення на обстеження, підтвердження завершення процедури тощо). Для автоматизованої передачі зображень у хмарне середовище використовується конвеєр процесів вилучення, перетворення та завантаження даних (Extract–Transform–Load, ETL-конвеєр), реалізований на основі Dataflow. Подібні підходи продемонстровано в проєкті INCISIVE, де ETL-модулі виконують читання DICOM-файлів, їх попередню обробку та передавання до серверів PACS у контейнеризованому середовищі. У GCP інтеграція локальних PACS із хмарними сховищами DICOM здійснюється через Cloud Healthcare API, що забезпечує стандартизовану та захищену передачу даних. Для підтримки подієвої архітектури нові зображення автоматично публікуються у темах служби публікації-підписки (publish/subscribe, Pub/Sub), на які підписані аналітичні сервіси, що дає можливість запускати обчислювальні процедури одразу після надходження даних.

Особлива увага приділяється деперсоналізації та забезпеченню конфіденційності пацієнтів. Cloud Healthcare API підтримує вбудовані механізми де-ідентифікації даних, а інтеграція з журналами аудиту дозволяє відстежувати всі операції над медичними зображеннями в межах нормативних вимог.

**Сховище даних та аналітична інфраструктура.** Сховище даних та аналітична інфраструктура. Після завантаження та попередньої очистки медичні зображення зберігаються у сервісі Google Cloud Storage (GCS). Відповідно до рекомендацій GCP Healthcare Solutions Playbook, архітектура передбачає використання окремих GCS-сховищ для архівів медичних знімків з урахуванням класифікації даних за рівнем доступності, географічним розташуванням та вимогами безпеки. Для зберігання метаданих та агрегованих результатів аналізу використовується аналітичний сервіс BigQuery, що дозволяє формувати масштабовані аналітичні набори даних і виконувати модельне навчання за допомогою розширення BigQuery Machine Learning (BigQuery ML).

**Моделювання з використанням методів глибокого навчання.** На етапі моделювання застосовуються сучасні алгоритми глибокого навчання, зокрема згорткові нейронні мережі (Convolutional Neural Networks, CNN) та трансформерні архітектури для обробки зображень, такі як Vision Transformer (ViT). Vertex AI забезпечує два основні підходи до тренування моделей – AutoML та custom training. AutoML орієнтований на автоматизовану побудову моделей, включаючи поділ даних на вибірки, нормалізацію, масштабування до стандартного розміру та використання трансферного навчання. У дослідженнях 2025 року AutoML-модель для виявлення пневмотораксу досягла точності близько 95 % і показала здатність до ефективного виконання як у хмарному середовищі, так і на периферійних пристроях. Підхід custom training надає розробникам ширшу гнучкість у виборі архітектур (ResNet, EfficientNet, ViT тощо), налаштуванні гіперпараметрів і застосуванні спеціалізованих оптимізаторів. Vertex AI Workbench забезпечує інтегроване середовище для експериментів із використанням GPU або TPU та сумісність з бібліотеками на кшталт MONAI.

Питання пояснюваності моделей має принципове значення у контексті медичних застосувань. Для інтерпретації результатів застосовуються методи Gradient-weighted Class Activation Mapping (Grad-CAM) та інтегрованих градієнтів (integrated gradients), які дають можливість відобразити внесок окремих областей (пікселів) у фінальний прогноз моделі, що полегшує клінічну валідацію та підвищує довіру лікарів до автоматизованих рішень.

**Розгортання, масштабування та мікросервісна архітектура.** Після тренування моделі розгортаються у сервісі Vertex AI Endpoints, що

дозволяє опублікувати їх як веб-служби у стилі Representational State Transfer (REST-служби) для інференсу в режимах реального часу чи пакетної обробки. Платформа підтримує версіювання моделей, гнучке керування трафіком та механізми безпечного оновлення (canary-deployment).

Для реалізації високопродуктивних сервісів інференсу використовується сервіс Google Kubernetes Engine (GKE), де моделі виконуються у контейнерному середовищі. Kubernetes-оркестрація підтримує автоматичне масштабування, самовідновлення та ізоляцію ресурсів. Взаємодія між мікросервісами реалізується через Pub/Sub, що забезпечує асинхронний транспорт повідомлень із високою доступністю. Компонент API Gateway виконує функції єдиної точки входу до системи, забезпечує автентифікацію, авторизацію, лімітування запитів та централізоване логування.

**Моніторинг та MLOps-підходи.** Оскільки життєвий цикл моделі не завершується на етапі розгортання, система передбачає комплексний підхід до моніторингу та MLOps. Сервіс Vertex AI Model Monitoring відстежує якість прогнозів, дрейф даних і відхилення між розподілами тренувальних і сервісних даних (training-serving skew). Для безперервної інтеграції та доставки (Continuous Integration/Continuous Delivery, CI/CD) застосовується Cloud Build у поєднанні з Vertex AI та Cloud Build Triggers. Цей процес охоплює тестування коду, автоматичний запуск сесій тренування, реєстрацію моделей у сховищі Model Registry та їх поступове розгортання. Dataflow надає інструменти для моніторингу потоків даних, включаючи автоматичне масштабування та діагностику помилок, що дозволяє оптимізувати використання ресурсів і забезпечувати стабільну роботу системи.

Запропонована архітектура розгортання, масштабування та MLOps у середовищі Google Cloud Platform (GCP) (Рис. 1) забезпечує конкретні технічні переваги порівняно з локальними та монолітними рішеннями, що традиційно застосовуються в медичних ІТ-системах. Використання керованих сервісів Vertex AI і Kubernetes дає змогу реалізувати автоматичне горизонтальне масштабування сервісів інференсу моделей, а не просто вертикальне нарощування потужності окремих серверів. Це дозволяє стабільно обробляти пікові навантаження (наприклад, масову обробку рентгенівських знімків у години максимальної діагностичної активності) без деградації продуктивності та з контрольованими витратами на обчислювальні ресурси.

Мікросервісна архітектура, доповнена асинхронною комунікацією через Pub/Sub, забезпечує ізоляцію функціональних компонентів (збір даних, передобробка, інференс, постобробка, зберігання результатів) і дає змогу незалежно масштабувати саме ті сервіси, які є «вузьким місцем» у ланцюжку обробки. Це спрощує оновлення окремих модулів без зупинки всієї системи та знижує ризики каскадних відмов, характерних для монолітних застосунків.

У сфері MLOps ключовою перевагою є наявність інтегрованого конвеєра CI/CD на базі Cloud Build і Vertex AI, який автоматизує тренування, тестування, реєстрацію та розгортання моделей. Це забезпечує кероване версіонування, відтворюваність експериментів і можливість контрольованого повернення до попередніх версій у разі виявлення деградації якості. Використання Vertex AI Model Monitoring дає змогу виявляти дрейф даних та зміну продуктивності моделей на реальних клінічних даних, що принципово важливо для довготривалої експлуатації системи в умовах змін популяції пацієнтів та налаштувань рентгенологічного обладнання.

Додатково, поєднання API Gateway, сервісного контролю доступу та централізованого журналювання формує цілісний контур безпеки та аудиту операцій над медичними даними й моделями. На відміну від типових локальних PACS-

орієнтованих рішень, це полегшує доведення відповідності нормативним вимогам і забезпечує простежуваність усіх критичних дій у системі.

**Забезпечення безпеки та відповідності нормативам.** Оскільки система працює з медичними даними, вона має відповідати вимогам стандартів HIPAA та GDPR. GCP підтримує комплекс заходів безпеки: шифрування даних на дисках (AES-256), шифрування у транзиті (TLS), управління ключами через Cloud KMS, диференційовані IAM-ролі, захист мережі на рівні VPC та DDoS-захист. Для гарантування конфіденційності дані проходять деперсоналізацію перед передачею у хмарне середовище. Організації, що експлуатують систему, зобов'язані проводити оцінку впливу на конфіденційність та укладати відповідні угоди щодо обробки даних.

**Висновки.** У роботі розв'язано актуальну науково-прикладну задачу побудови масштабованої та надійної системи автоматизованого аналізу рентгеновських знімків на основі методів глибокого навчання та сучасних хмарних технологій. Запропоновано цілісний архітектурний підхід, що поєднує інструменти Google Cloud Platform, зокрема Vertex AI, Kubernetes і Dataflow, у межах мікросервісної парадигми та концепції MLOps. Це дозволяє забезпечити повний життєвий цикл роботи з медичними зображеннями – від збору та попередньої обробки даних до тренування, роз-

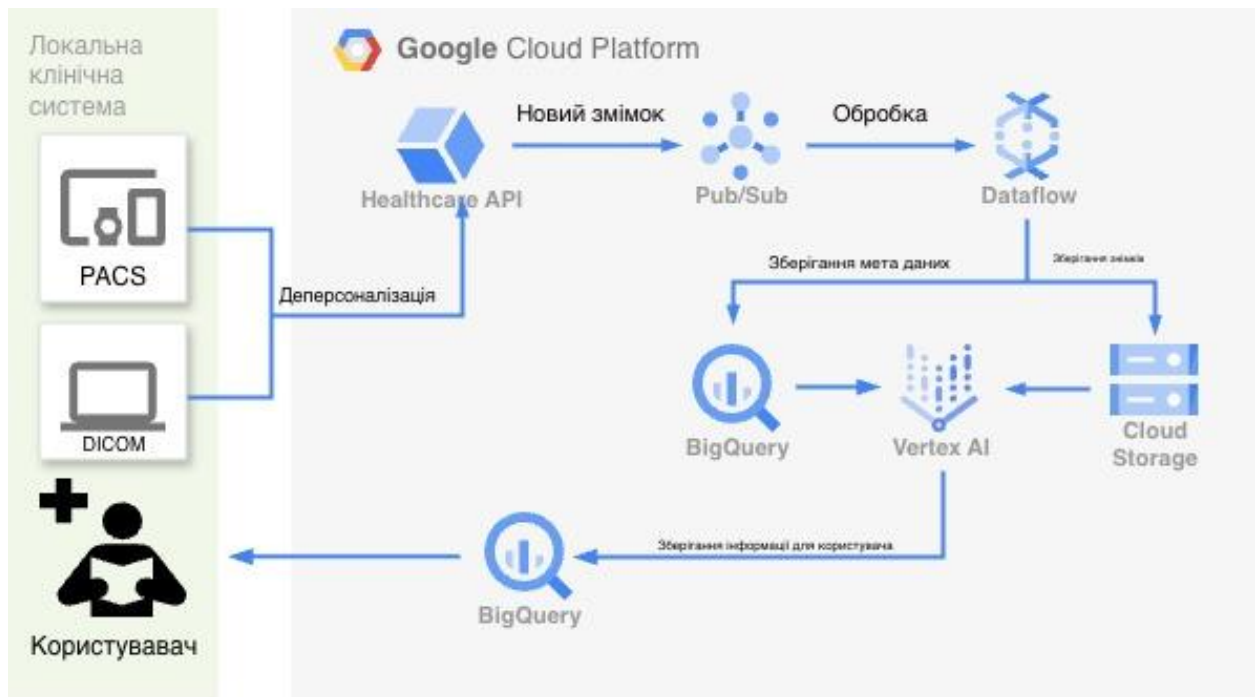


Рис. 1. Інфраструктура розгортання та моніторингу моделей глибокого навчання у хмарному середовищі GCP

гортання, масштабування й моніторингу моделей машинного навчання.

Науковий здобуток роботи полягає в системному обґрунтуванні архітектури, яка забезпечує інтеграцію алгоритмів глибокого навчання у реальні клінічні робочі процеси з урахуванням вимог до масштабованості, відмовостійкості, безпеки та нормативної відповідності (HIPAA, GDPR). Показано, що використання керованих хмарних сервісів у поєднанні з контейнеризацією та асинхронною взаємодією мікросервісів дає змогу суттєво скоротити час аналізу рентгенівських знімків і підвищити точність діагностичних рішень.

Наукова новизна роботи полягає в запропонованні та обґрунтуванні інтегрованої архітектури системи аналізу рентгенівських знімків, яка поєднує:

- масштабовану обробку великих потоків медичних зображень на основі Dataflow;
- уніфіковане середовище тренування, деплоюменту та моніторингу моделей глибокого навчання за допомогою Vertex AI;
- мікросервісну оркестрацію на базі Kubernetes з автоматичним масштабуванням;
- впровадження практик MLOps і засобів пояснюваного штучного інтелекту для підвищення довіри до результатів моделей.

Отримані результати мають теоретичне та практичне значення й можуть бути використані як методологічна основа для розробки та впровадження інтелектуальних систем медичної діагностики, а також для подальших досліджень у напрямі мультимодальних моделей, федеративного навчання та розширення запропонованого підходу на інші види медичних зображень.

#### Список літератури:

1. Litjens G., Kooi T., Bejnordi B. E., Setio A. A., et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. DOI: <https://doi.org/10.1016/j.media.2017.07.005>
2. Shen D., Wu G., Suk H.-I. "Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017. DOI: <https://doi.org/10.1146/annurev-bioeng-071516-044442>
3. Abadi M., Barham P., Chen J., Chen Z., et al. TensorFlow: Large-scale machine learning on heterogeneous systems. *Google Research*, 2016. URL: <https://www.tensorflow.org>
4. European Union. General Data Protection Regulation (GDPR). *Official Journal of the European Union*, 2016. URL: <https://gdpr.eu>
5. Lee J., Zhai Y., Afzal N., Ganapathy S. Cloud-based medical imaging frameworks: A review. *Journal of Digital Imaging*, vol. 34, pp. 1–15, 2021. DOI: <https://doi.org/10.1007/s10278-020-00417-8>
6. Google Cloud. Introducing Medical Imaging Suite. *Google Cloud Blog*, 2022. URL: <https://cloud.google.com/blog/products/healthcare-life-sciences/introducing-medical-imaging-suite>
7. Cardoso M. J., Li W., Brown R., Ma N., et al. MONAI: A comprehensive framework for deep learning in healthcare imaging. *arXiv preprint*, arXiv:2211.02701, 2022. URL: <https://arxiv.org/abs/2211.02701>
8. Google Cloud AI. *AutoML Vision for medical imaging: Case studies*. Vertex AI Documentation, 2024–2025. URL: <https://cloud.google.com/vertex-ai/docs>
9. Google Research. *MedLM: Large medical imaging models for clinical tasks*. Research Release, 2024. URL: <https://research.google>
10. AWS. *Introducing AWS HealthImaging – optimized storage and retrieval for medical imaging*. AWS News Blog, 2022. URL: <https://aws.amazon.com/blogs/aws/introducing-aws-healthimaging>
11. Newman S. *Building Microservices: Designing Fine-Grained Systems*. Себастопол: O'Reilly Media, Inc., 2021. 500 с.
12. Spring AMQP. URL: <https://spring.io/projects/spring-amqp>. (дата звернення: 29.06.2024).

#### **Hazdiuk K.P., Movcheniuk R.V. ARCHITECTURE OF AN X-RAY IMAGE ANALYSIS SYSTEM USING GCP VERTEX AI, KUBERNETES, DATAFLOW, AND A MICROSERVICE ARCHITECTURE**

*The article presents a comprehensive architectural approach to the development of an automated X-ray image analysis system based on deep learning methods and cloud technologies. The relevance of the study is driven by the rapid growth in the volume of medical images, the need to reduce image interpretation time, and the shortage of qualified radiologists, which limits the effectiveness of traditional diagnostic workflows. Modern deep learning algorithms, particularly convolutional neural networks and transformer-based architectures, demonstrate high accuracy in detecting pathologies in medical images; however, their integration into real clinical environments requires a reliable, scalable, and well-managed computational infrastructure.*

*This paper proposes a system architecture built on Google Cloud Platform services, covering the complete machine learning lifecycle—from acquisition and preprocessing of X-ray data to model training, deployment,*

*and monitoring. Vertex AI is used to organize the training, versioning, and deployment of deep learning models, as well as to monitor their performance during operation. The orchestration of containerized system components is implemented using Kubernetes, which provides automatic scaling, fault tolerance, and flexible resource management. Dataflow is applied for processing large streams of X-ray images and preparing them for machine learning in both batch and streaming modes.*

*The proposed architecture integrates the principles of microservice-based design, MLOps, and cloud computing, ensuring stable system operation, effective integration with medical information systems, and compliance with scalability requirements. The obtained results demonstrate that the proposed approach significantly reduces the time required for X-ray image analysis, improves diagnostic accuracy through the use of advanced deep learning models, and enables the system to adapt dynamically to changing workloads according to the needs of medical institutions. The proposed solution can serve as an architectural foundation for intelligent medical diagnostic systems and has strong potential for further development, including extension to other types of medical imaging*

**Keywords:** *Deep Learning, X-ray images, medical diagnostics, Google Cloud Platform, Vertex AI, Kubernetes, Dataflow, microservice architecture, MLOps, cloud computing, automated image analysis.*

Дата першого надходження статті до видання: 10.01.2026

Дата прийняття статті до друку після рецензування: 12.02.2026

Дата публікації (оприлюднення) статті: 08.04.2026