

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна
Навчально-науковий інститут комп'ютерних наук та штучного інтелекту
Кафедра комп'ютерних систем та робототехніки

До захисту допущено
Кафедрою комп'ютерних систем та робототехніки
протокол № __ від __ грудня 2025р.

завідувач кафедри _____ Максим ХРУСЛОВ
(підпис)

«__» _____ 2025 р.

Кваліфікаційна робота

здобувача другого (магістерського) рівня вищої освіти

«МОДЕЛЬ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ У СИСТЕМАХ МЕДИЧНОЇ ДІАГНОСТИКИ»

Спеціальність 174 – *Автоматизація, комп'ютерно-інтегровані технології
та робототехніка*

Освітня програма *Комп'ютеризовані системи управління та автоматика*

Виконавець


(підпис)

Владислав НЕСТЕРЕНКО

Науковий керівник


(підпис)

Вікторія СТРІЛЕЦЬ

Харків – 2025

АНОТАЦІЯ

Пояснювальна записка до магістерської атестаційної роботи складається зі вступу, трьох розділів, висновків, списку використаних джерел і чотирьох додатків. Загальний обсяг роботи складає 94 сторінок, із яких 70 сторінок основної частини з 13 рисунками, 2 таблицями, 31 найменуванням списку використаних джерел та чотирма додатками.

Метою кваліфікаційної роботи є підвищення ефективності розпізнавання ультразвукових зображень на основі згорткових нейронних мереж, здатних диференціювати критичні стани, в системах підтримки клінічних рішень.

Об'єкт дослідження – процес автоматизованої медичної діагностики патологій внутрішніх органів (зокрема печінки) на основі аналізу візуальних даних, отриманих за допомогою ультразвукового дослідження.

Предмет дослідження – архітектури та алгоритми функціонування моделей глибокого навчання, зокрема згорткових нейронних мереж (CNN), у задачах багатокласової класифікації ультразвукових зображень печінки.

Проблема, яка вирішується в кваліфікаційній роботі, полягає в подоланні суб'єктивності візуальної інтерпретації діагностичних даних та нівелюванні впливу спекл-шуму на точність діагнозу. Робота вирішує задачу автоматизації диференціації доброякісних та злоякісних новоутворень печінки шляхом розробки програмного комплексу на базі навченої нейромережі Custom CNN, що дозволяє знизити ризик лікарських помилок.

Область застосування – медичні заклади, діагностичні центри, відділення радіології та онкології. Розроблений програмний продукт може використовуватися як система підтримки прийняття рішень («друга думка») для лікарів УЗД-діагностики.

Ключові слова: *Deep Learning, нейронні мережі, CNN, ультразвукова діагностика, печінка, класифікація зображень, Python, Tensorflow, СППР.*

ABSTRACT

The explanatory note to the Master's thesis consists of an introduction, three chapters, conclusions, a list of references, and four appendices. The total volume of the work is 94 pages, of which 70 pages are the main part including 13 figures, 2 tables, 31 references, and four appendices.

The purpose of the thesis is to enhance the efficiency of ultrasound image recognition based on convolutional neural networks capable of differentiating critical conditions in clinical decision support systems.

The object of research is the process of automated medical diagnosis of internal organ pathologies (specifically the liver) based on the analysis of visual data obtained through ultrasound examination.

The subject of research covers the architectures and functioning algorithms of deep learning models, particularly Convolutional Neural Networks (CNN), in tasks of multi-class classification of liver ultrasound images.

The problem solved in the thesis lies in overcoming the subjectivity of visual interpretation of diagnostic data and mitigating the impact of speckle noise on diagnostic accuracy. The work solves the task of automating the differentiation between benign and malignant liver neoplasms by developing a software complex based on the trained Custom CNN neural network, which allows for reducing the risk of medical errors.

The field of application includes medical institutions, diagnostic centers, radiology and oncology departments. The developed software product can be used as a decision support system ("second opinion") for ultrasound diagnostic physicians.

Keywords: *Deep Learning, neural networks, CNN, ultrasound diagnostics, liver, image classification, Python, Tensorflow, CDSS.*

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ І УМОВНИХ ПОЗНАЧЕНЬ.....	6
ВСТУП	7
РОЗДІЛ 1 АНАЛІЗ ПРОБЛЕМИ АВТОМАТИЗОВАНОЇ ДІАГНОСТИКИ В УЛЬТРАЗВУКОВИХ ДОСЛІДЖЕННЯХ	9
1.1 Фізичні основи формування ультразвукових зображень та природа шумів	9
1.2 Проблематика візуальної інтерпретації та класифікація патологічних станів.....	11
1.3 Огляд методів автоматизованого аналізу: від класичних алгоритмів до глибокого навчання.....	16
Висновки за розділом 1.....	18
РОЗДІЛ 2 МЕТОДОЛОГІЯ ПОБУДОВИ ТА НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ МЕДИЧНОЇ ДІАГНОСТИКИ	20
2.1 Парадигма глибокого навчання в ультразвуковій діагностиці	20
2.2 Теоретичні основи побудови згорткових нейронних мереж.....	22
2.2.1 Математична модель згорткового шару	22
2.2.2 Агрегація ознак та зниження розмірності	25
2.2.3 Математична формалізація процесу навчання.....	27
2.3 Формування та підготовка набору даних для навчання.....	28
2.3.1 Структура та класифікація даних	29
2.3.2 Попередня обробка даних	31
2.4 Архітектури досліджуваних моделей	33
2.4.1 Власна базова згорткова нейронна мережа (Custom CNN)	33
2.4.2 Еволюція глибини та сприйняття: класичні архітектури	35
2.4.3 Легковагові архітектури для мобільних систем (POCUS).....	38
2.5 Методика навчання та програмна реалізація	41
Висновки за розділом 2.....	44
РОЗДІЛ 3 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ РОЗРОБЛЕНИХ МОДЕЛЕЙ.....	46
3.1 Організація експерименту та метрики оцінювання якості	46
3.1.1. Характеристика програмно-апаратного середовища	46
3.1.2. Система метрик для оцінювання медичної діагностики.....	47
3.2 Порівняльний аналіз результатів навчання моделей.....	50

3.2.1 Аналіз динаміки навчання та збіжності алгоритмів	50
3.2.2 Зведені результати тестування.....	53
3.2.3 Ефективність легковагових архітектур для портативної діагностики	54
3.2.4 Аналіз співвідношення «Точність – Ресурсоемність»	56
3.3 Аналіз діагностичних помилок та клінічна інтерпретація.....	58
3.3.1 Аналіз хибно-негативних результатів (Ризик пропуску патології) ..	59
3.3.2 Аналіз хибно-позитивних результатів (Ризик гіпердіагностики) ...	61
3.3.3 Матриця рішень та вибір оптимальної моделі	64
3.4 Програмна реалізація системи підтримки прийняття рішень	65
3.4.1 Архітектура програмного комплексу та технологічний стек	66
3.4.2 Алгоритмічна реалізація обробки даних	67
3.4.3 Інтерфейсні рішення та продуктивність	68
Висновки за розділом 3.....	70
ВИСНОВКИ.....	72
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	74
ДОДАТКИ.....	79

ПЕРЕЛІК СКОРОЧЕНЬ І УМОВНИХ ПОЗНАЧЕНЬ

УЗД – ультразвукове дослідження / ультразвукова діагностика

ГЦК – гепатоцелюлярна карцинома

СППР – система підтримки прийняття рішень

CAD – Computer-Aided Diagnosis (система автоматизованої діагностики)

CNN – Convolutional Neural Network (згорткова нейронна мережа)

FN – False Negative (хибно-негативний результат / пропуск патології)

FP – False Positive (хибно-позитивний результат / хибна тривога)

POCUS – Point-of-Care Ultrasound (портативна ультразвукова діагностика)

ReLU – Rectified Linear Unit (випрямлена лінійна функція активації)

ROI – Region of Interest (зона інтересу)

TN – True Negative (істинно-негативний результат)

TP – True Positive (істинно-позитивний результат)

ВСТУП

Стрімкий розвиток медичної діагностики в останні десятиліття нерозривно пов'язаний з інтеграцією новітніх інформаційних технологій. Серед усіх методів візуалізації внутрішніх органів ультразвукове дослідження (УЗД) посідає провідне місце завдяки своїй доступності та безпечності. Особливого значення цей метод набуває при дослідженні органів черевної порожнини, зокрема печінки, де своєчасне виявлення структурних змін є критичним для ефективного лікування. Проте інтерпретація ультразвукових даних залишається складним процесом, що суттєво залежить від досвіду лікаря та технічних характеристик обладнання, що створює передумови для розробки автоматизованих систем аналізу.

Актуальність теми дослідження. Захворювання печінки, зокрема гепатоцелюлярна карцинома та фіброз, є глобальною проблемою охорони здоров'я. Рання діагностика цих патологій ускладнюється специфічною фізичною природою ультразвуку: наявністю спекл-шуму, низькою контрастністю та візуальною схожістю початкових стадій захворювань зі здоровою тканиною. Це призводить до суб'єктивності оцінювання та підвищує ймовірність діагностичних помилок. Сучасні методи глибокого навчання, а саме згорткові нейронні мережі (CNN), дозволяють автоматизувати процес пошуку діагностичних ознак, виявляючи приховані текстурні патерни, недоступні людському оку. Однак існуючі універсальні моделі часто демонструють недостатню ефективність на реальних медичних даних через їх зашумленість та дисбаланс класів. Тому розробка спеціалізованих моделей, адаптованих до специфіки ультразвукових зображень, є актуальним науковим завданням, вирішення якого сприятиме підвищенню якості діагностики.

Об'єкт дослідження – процес автоматизованої медичної діагностики патологій внутрішніх органів (зокрема печінки) на основі аналізу візуальних даних, отриманих за допомогою ультразвукового дослідження.

Предмет дослідження – архітектури та алгоритми функціонування моделей глибокого навчання, зокрема згорткових нейронних мереж (CNN), у задачах багатокласової класифікації ультразвукових зображень печінки.

Метою дослідження є підвищення ефективності розпізнавання ультразвукових зображень на основі згорткових нейронних мереж, здатних диференціювати критичні стани, в системах підтримки клінічних рішень.

Методи дослідження: методи цифрової обробки зображень, методи глибокого навчання та теорія штучних нейронних мереж, методи математичної статистики, методи оцінки якості класифікації, методи об'єктно-орієнтованого програмування.

Завдання дослідження:

1. Виконати аналіз особливостей формування ультразвукових зображень та обмежень існуючого науково-методичного апарата щодо автоматизованої діагностики патологій печінки.

2. Сформувати репрезентативний набір даних шляхом консолідації різномірних джерел та розробити алгоритм попередньої обробки для усунення дисбалансу класів.

3. Обґрунтувати вибір та виконати адаптацію архітектур згорткових нейронних мереж для задачі розпізнавання текстур біомедичних зображень в умовах спекл-шуму.

4. Виконати навчання розроблених моделей, провести оцінку їх ефективності та порівняльний аналіз за метриками точності, чутливості та специфічності.

5. Розробити програмний засіб підтримки прийняття рішень із графічним інтерфейсом для практичної реалізації найкращої моделі діагностики.

РОЗДІЛ 1

АНАЛІЗ ПРОБЛЕМИ АВТОМАТИЗОВАНОЇ ДІАГНОСТИКИ В УЛЬТРАЗВУКОВИХ ДОСЛІДЖЕННЯХ

1.1 Фізичні основи формування ультразвукових зображень та природа шумів

Ультразвукова діагностика (УЗД) базується на реєстрації ехо-сигналів, що виникають при відбитті акустичних хвиль від меж розділу тканин з різним акустичним імпедансом. Основним режимом візуалізації для діагностики патологій печінки є В-режим (Brightness Mode) – двовимірна томографія, де амплітуда зареєстрованого ехо-сигналу трансформується у яскравість відповідного пікселя на екрані. У цьому режимі структури з високою відбивною здатністю (наприклад, капсула печінки або стінки судин) відображаються як гіперехогенні (світлі) ділянки, а рідинні утворення, що майже не відбивають звук (кісти, судини) – як анехогенні (темні). Проміжні значення формують шкалу сірого, яка відображає акустичну щільність тканини.

Критичною особливістю ультразвукових даних, що фундаментально відрізняє їх від оптичних фотографій або КТ-знімків, є наявність спекл-шуму (Speckle Noise). Це інтерференційний феномен, який виникає внаслідок когерентного додавання хвиль, відбитих від численних хаотично розташованих мікроструктур тканини (розсіювачів), розмір яких менший за довжину хвилі. У результаті конструктивної та деструктивної інтерференції на зображенні формується характерна зерниста текстура, яка не є шумом у класичному розумінні (як адитивний шум камери), а є внутрішньою властивістю методу візуалізації.

З точки зору комп'ютерного зору, спекл-шум створює значні виклики: він має мультиплікативну природу (залежить від сигналу), знижує контрастність та "розмиває" межі об'єктів, що робить класичні методи

сегментації (наприклад, детектори країв) малоефективними. Однак для глибокого навчання спекл є джерелом корисної інформації. Статистичні характеристики розподілу спеклу корелюють з реальною мікроструктурою тканини паренхіми. Саме ця специфічна "зернистість" дозволяє досвідченим лікарям та навченим нейронним мережам диференціювати здорову тканину від тканини з ознаками фіброзу чи цирозу, навіть якщо явні вогнищеві утворення відсутні. Тому сучасні підходи, зокрема згорткові нейронні мережі (CNN), не намагаються просто "відфільтрувати" цей шум, а вчаться виділяти з нього приховані текстурні патерни.

Проблеми УЗД. Спекл-шум має мультиплікативну природу, тобто його інтенсивність залежить від інтенсивності корисного сигналу. Математично модель зашумленого зображення $g(x, y)$ можна представити формулою (1.1):

$$g(x, y) = f(x, y) \cdot n(x, y) + a(x, y), \quad (1.1)$$

де $f(x, y)$ – ідеальне зображення без шуму; $n(x, y)$ – мультиплікативна компонента шуму; $a(x, y)$ – адитивна компонента шуму [1].

Така природа шуму призводить до того, що яскраві ділянки зображення є більш зашумленими, що ускладнює застосування стандартних методів фільтрації. Спекл-шум «забруднює» розподіл рівнів сірого, знижуючи контраст між патологічним вогнищем та здоровою тканиною, а хаотична структура шуму робить контури об'єктів переривчастими та нерівними. Більше того, розмір «зерна» спеклу часто співрозмірний з дрібними діагностичними ознаками, такими як мікрокальцинати або дрібні метастази, що робить їх невидимими для класичних алгоритмів аналізу.

Окрім шумів, УЗД-зображення містять артефакти (рис. 1.1), що виникають через порушення припущень про прямолінійне поширення звуку [2]. У контексті машинного навчання ці артефакти мають подвійну природу: з одного боку, вони спотворюють анатомію, з іншого – є важливими діагностичними ознаками. Наприклад, акустична тінь, що виникає позаду об'єктів з високим поглинанням (кальцинати, кістки) і виглядає як чорна

смуга, для нейромережі є надійною ознакою наявності твердих включень [2]. Іншим прикладом є дорсальне посилення, яке виникає позаду рідинних утворень (кісти) через слабе поглинання звуку рідиною. Цей артефакт є ключовим маркером для диференціації доброякісних кіст від солідних пухлин.

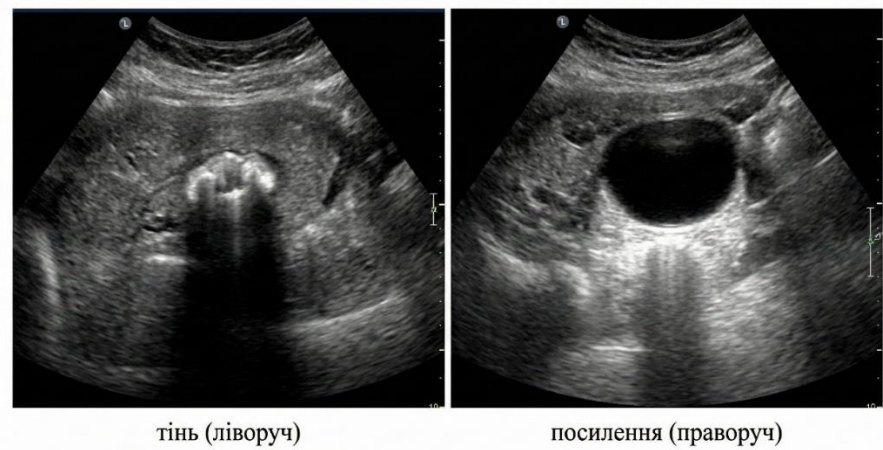


Рисунок 1.1 – Приклади акустичних артефактів

Таким чином, ультразвукове зображення з погляду комп'ютерного аналізу є складним сигналом з низьким співвідношенням сигнал/шум та низькою контрастністю. Корисна інформація про патологію часто «закодована» не лише у геометрії об'єктів, але й у статистичних характеристиках розподілу спекл-шуму та наявності супутніх артефактів. Це робить застосування класичних методів обробки зображень малоефективним і зумовлює необхідність використання більш потужних підходів, таких як глибоке навчання, здатних виділяти складні нелінійні залежності з зашумлених даних.

1.2 Проблематика візуальної інтерпретації та класифікація патологічних станів

Процес інтерпретації ультразвукових зображень не є пасивним зчитуванням інформації, а являє собою складний когнітивний процес, що поєднує візуальне сприйняття, розпізнавання патернів та прийняття рішень в

умовах невизначеності. Якщо для КТ або МРТ існують чіткі денситометричні критерії (наприклад, одиниці Гаунсфілда), то ультразвукова діагностика залишається значною мірою якісним, описовим методом. Це породжує фундаментальну *проблему суб'єктивності*, яка стає головним бар'єром для стандартизації діагностики та однією з основних причин розробки систем автоматизованого аналізу.

Центральною проблемою традиційного підходу є *висока варіабельність результатів інтерпретації*. Розрізняють інтер-операторську варіабельність (inter-observer variability), коли різні лікарі приходять до відмінних висновків, аналізуючи одні й ті самі скани, та інтра-операторську варіабельність (intra-observer variability), коли рішення одного й того ж експерта може змінюватися залежно від зовнішніх факторів, таких як втома, часовий тиск або контекст попередніх діагнозів [3]. Дослідження показують, що рівень узгодженості (agreement) між радіологами при диференціації складних випадків, наприклад, між атиповою гемангіомою та гепатоцелюлярною карциномою, може бути незадовільно низьким.

Суб'єктивність діагностики посилюється *когнітивними упередженнями*, притаманними людському сприйняттю. Зокрема, поширеним є «упередження підтвердження» (confirmation bias), коли лікар підсвідомо шукає на зображенні ознаки, що підтверджують попередній діагноз, ігноруючи суперечливі деталі. Іншим небезпечним явищем є «задоволення від пошуку» (satisfaction of search), коли після виявлення однієї очевидної патології (наприклад, кісти) процес ретельного сканування припиняється, внаслідок чого пропускаються менш помітні, але більш небезпечні супутні утворення. Системи комп'ютерного зору, на відміну від людини, забезпечують детермінований аналіз всього поля зображення, що дозволяє нівелювати вплив цих психологічних факторів.

Для побудови ефективних алгоритмів класифікації необхідно *формалізувати візуальні ознаки* основних станів печінки. З погляду

комп'ютерного аналізу, задача зводиться до розпізнавання трьох основних класів: норма, доброякісні утворення та злоякісні новоутворення. Кожен з цих класів характеризується специфічним набором текстурних та морфологічних ознак.

Нормальна паренхіма печінки на ультразвуковому зображенні характеризується однорідною, дрібнозернистою ехоструктурою. Ключовим параметром для алгоритмів є ехогенність – середня яскравість пікселів у зоні інтересу. У нормі печінка є ізоехогенною або злегка гіперехогенною відносно кори правої нирки, що слугує еталоном для калібрування яскравості. Контури органу мають бути рівними та чіткими, а судинний малюнок – рівномірним, без ознак деформації або збіднення [4]. Будь-яке порушення цієї гомогенності розглядається системою як потенційна аномалія.

Суттєвим фактором, що ускладнює автоматизовану діагностику, є наявність супутніх дифузних змін паренхіми, які змінюють базові характеристики «норми». Найпоширенішим станом є стеатоз (жирова інфільтрація), який підвищує загальну ехогенність тканини, роблячи її «яскравою», та значно збільшує коефіцієнт загасання ультразвуку [4]. Це призводить до того, що глибокі відділи зображення стають затемненими, а контрастність об'єктів у цих зонах критично знижується. Ще більшим викликом є цироз печінки, який трансформує гладку текстуру органу в грубозернисту та вузлувату. На фоні такої неоднорідної (гетерогенної) текстури алгоритмам детекції стає значно важче відрізнити патологічні вогнища від регенераторних вузлів самої тканини, що різко підвищує ймовірність хибно-позитивних спрацювань.

Клас доброякісних утворень включає патології, які не несуть безпосередньої загрози життю пацієнта, але вимагають коректної диференціації від раку. Найпоширенішими представниками цього класу є прості кісти та гемангіоми. Прості кісти мають «ідеальні» з точки зору алгоритмічного розпізнавання ознаки: вони є анехогенними (повністю

чорними), мають чіткі, тонкі контури та сферичну форму. Важливою ознакою є ефект дорсального посилення сигналу позаду кісти, який виникає через відсутність поглинання ультразвуку рідиною. Гемангіоми у типових випадках візуалізуються як гіперехогенні (яскраво-білі), однорідні утворення з чіткими межами.

Проте візуальна картина доброякісних утворень не завжди є однозначною. Існує широкий спектр атипових проявів, які стають причиною діагностичних помилок. Наприклад, гемангіоми великих розмірів можуть втрачати свою однорідність через тромбоз або фіброз центральної частини, набуваючи гетерогенної структури, схожої на злоякісну пухлину. Крім того, на фоні жирового гепатозу, коли навколишня паренхіма стає аномально світлою, звичайна світла гемангіома може виглядати як гіпоехогенне (темне) утворення, імітуючи метастаз або гепатоцелюлярну карциному. Складні кісти з перегородками або внутрішнім вмістом (наприклад, при крововиливі) втрачають свою характерну анехогенність, що унеможливорює їх класифікацію простими пороговими методами сегментації.

Найбільшу складність для діагностики становить клас злоякісних новоутворень, до якого відносяться гепатоцелюлярна карцинома (ГЦК) та метастази. ГЦК називають «хамелеоном» ультразвукової діагностики через виражений поліморфізм її проявів. Малі пухлини часто виглядають як гіпоехогенні (темні) вогнища з нечіткими контурами, що робить їх важкопомітними на фоні паренхіми. Зі зростанням пухлина набуває гетерогенної, «мозаїчної» структури, що відображає зони некрозу, фіброзу та активного росту. Специфічною ознакою злоякісності є «Halo sign» – тонкий гіпоехогенний обідок навколо утворення, що відповідає капсулі пухлини або зоні компресії тканини. Метастатичні ураження також демонструють значну варіабельність, часто маючи вигляд «мішені» (Bull's eye) з гіпоехогенним обідком та гіперехогенним центром [15].

Окремою проблемою для комп'ютерного зору є *феномен ізоехогенності*, коли акустичний імпеданс пухлинної тканини майже не відрізняється від імпедансу здорової паренхіми. У таких випадках вогнище стає фактично невидимим у градаціях сірого і може бути детектуване виключно за непрямими ознаками: деформацією судинного малюнка або локальним порушенням текстурного патерну (ентропії зображення). Межі таких утворень не мають чіткого градієнта яскравості, а являють собою плавну зону переходу, що робить застосування класичних детекторів країв (наприклад, оператора Кенні) неефективним [5]. Нейронна мережа в цьому контексті повинна навчитися реагувати не на перепад інтенсивності, а на зміни у просторовому розподілі спекл-шуму, що є задачею значно вищого рівня складності.

Головним викликом для автоматизованої класифікації є *перетин просторів ознак* (feature overlap) між доброякісними та злоякісними утвореннями. Наприклад, атипова гемангіома на фоні стеатозу печінки може виглядати гіпоехогенною, імітуючи злоякісну пухлину. І навпаки, високоехогенна форма ГЦК може бути помилково класифікована як гемангіома. Крім того, наявність дифузних захворювань, таких як цироз, кардинально змінює фонову текстуру органу, роблячи її грубозернистою та неоднорідною, що значно ускладнює виділення вогнищевих утворень.

Значний вплив на можливість коректної інтерпретації зображення мають *технічні параметри сканування*, які залежать від налаштувань оператора та конституції пацієнта. Некоректно налаштоване підсилення (Gain) може призвести до «засвічення» зображення або втрати інформації в тінях, що нівелює текстурні відмінності між класами. У пацієнтів з надмірною вагою виникає ефект розмиття зображення через розсіювання ультразвуку в підшкірно-жировій клітковині, що знижує роздільну здатність та робить контури патологій ще менш чіткими. Крім того, динамічний характер дослідження призводить до появи артефактів руху (наприклад, від дихання пацієнта), які можуть спотворювати геометрію об'єктів. Модель машинного

навчання повинна бути інваріантною до цих збурень, тобто вміти виділяти стабільні діагностичні ознаки незалежно від рівня шуму чи контрастності конкретного кадру.

Ця проблема поліморфізму та візуальної схожості різних класів патологій демонструє обмеженість класичних методів аналізу, що базуються на жорстких правилах (наприклад, «якщо об'єкт чорний і круглий – це кіста»). В реальних клінічних умовах, зашумлених спекл-артефактами та варіабельністю налаштувань сканера, такі евристики часто призводять до помилок. Саме тому виникає необхідність застосування методів глибокого навчання, здатних виявляти нелінійні, високорівневі залежності між текстурними патернами та клінічним діагнозом, що дозволить об'єктивізувати процес прийняття рішень та знизити ризик діагностичних помилок.

1.3 Огляд методів автоматизованого аналізу: від класичних алгоритмів до глибокого навчання

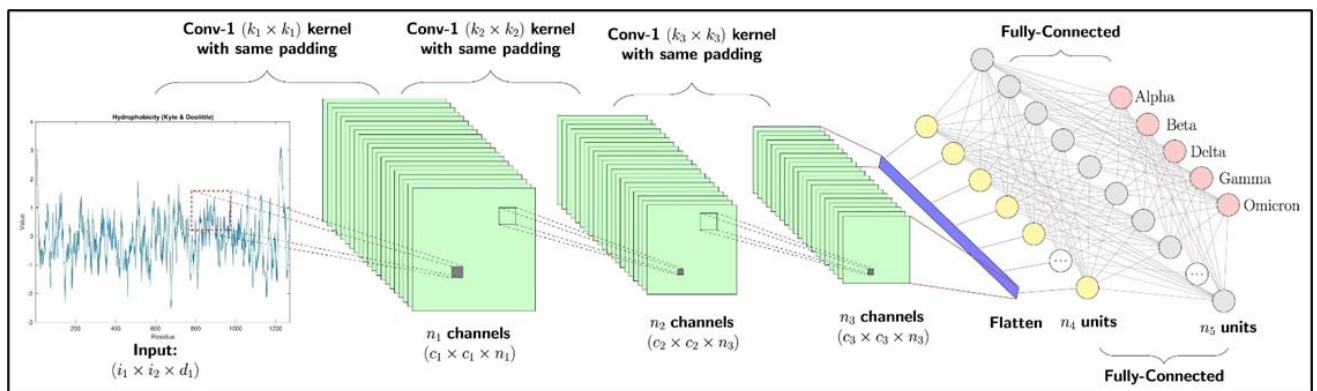
Розвиток систем автоматизованої діагностики (CAD – Computer-Aided Diagnosis) пройшов еволюційний шлях від спроб формалізувати правила візуального сприйняття лікаря до створення самонавчальних систем штучного інтелекту. Історично склалося два покоління методів аналізу ультразвукових зображень: класичні методи, що базуються на ручному конструюванні ознак (Feature Engineering), та сучасні методи глибокого навчання (Deep Learning), що базуються на автоматичному навчанні представлень [6].

Класичний підхід до аналізу медичних зображень традиційно реалізується у вигляді послідовного конвеєра (pipeline), що складається з чотирьох етапів: попередня обробка, сегментація, виділення ознак та класифікація. Критичним етапом тут є пригнічення спекл-шуму, який розглядається класичними алгоритмами як завада. Для цього застосовуються адаптивні фільтри (Лі, Фроста, Куана) та методи анізотропної дифузії (SRAD), які намагаються згладити однорідні ділянки, зберігаючи межі об'єктів. Після

фільтрації виконується сегментація зони інтересу (ROI) методами порогової обробки, розростання регіонів або активних контурів («змійок»).

На етапі виділення ознак класичні системи покладаються на математичні дескриптори, розроблені експертами. Найбільш поширеними є текстурні ознаки першого (гістограма яскравості) та другого порядку (матриця суміжності рівнів сірого – GLCM, ознаки Хараліка), а також локальні бінарні патерни (LBP) та вейвлет-перетворення. Сформований вектор ознак подається на вхід традиційних класифікаторів, таких як Машина опорних векторів (SVM), Випадковий ліс (Random Forest) або k-найближчих сусідів (k-NN). Головним недоліком цього підходу є його залежність від якості попередньої сегментації (яка часто є помилковою на зашумлених знімках) та обмежена здатність вручну сконструйованих ознак описувати складні, нелінійні варіації патологій.

Револьюційним кроком у подоланні цих обмежень стала поява *згорткових нейронних мереж* (CNN – Convolutional Neural Networks) (рис. 1.2). В основі роботи CNN лежить біонічний принцип ієрархічної обробки візуальної інформації, притаманний зоровій корі головного мозку [7]. На відміну від класичних методів, CNN не потребують окремого етапу виділення ознак. Мережа самостійно формує набір фільтрів (ядер згортки) у процесі навчання на розмічених даних, оптимізуючи їх для вирішення конкретної діагностичної задачі.



Архітектура CNN складається з послідовності згорткових шарів (Convolutional layers), шарів активації (зазвичай ReLU) та шарів підвибірки (Pooling). На початкових шарах мережа автоматично навчається детектувати прості примітиви – межі, кути та плями ("blob detection"). Глибші шари комбінують ці примітиви у складні текстурні патерни та фрагменти об'єктів [7]. Ця здатність до узагальнення дозволяє CNN ефективно працювати зі спекл-шумом: замість того, щоб фільтрувати його, мережа вчиться ігнорувати високочастотні флуктуації або використовувати статистику шуму як додаткову ознаку щільності тканини.

Сучасні дослідження в області ультразвукової діагностики печінки фокусуються на використанні глибоких архітектур (VGG, ResNet, DenseNet) та методу трансферного навчання (Transfer Learning). Цей підхід дозволяє використовувати знання (ваги), отримані моделями на великих наборах даних (наприклад, ImageNet), для задач з обмеженою кількістю медичних зображень. Крім того, розвиток "легковагових" архітектур, таких як MobileNet та EfficientNet, відкриває можливості для впровадження систем штучного інтелекту безпосередньо в портативні ультразвукові сканери, реалізуючи концепцію Point-of-Care Ultrasound (POCUS).

Таким чином, перехід від класичних алгоритмів до глибокого навчання дозволяє реалізувати концепцію "End-to-End" аналізу, де система приймає на вхід "сире" зображення і видає ймовірність діагнозу, міняючи вразливі етапи ручної сегментації та фільтрації. Це забезпечує вищу об'єктивність та стійкість діагностики в умовах реальної клінічної практики.

Висновки за розділом 1

У першому розділі проведено комплексний аналіз проблеми автоматизації діагностики патологій печінки за даними ультразвукових досліджень. Встановлено, що головними перешкодами для комп'ютерного зору є специфічна фізична природа ультразвукового зображення, зокрема

наявність мультиплікативного спекл-шуму, ефектів затухання сигналу та низька контрастність меж патологічних утворень.

Проаналізовано клінічний контекст задачі та виявлено, що висока суб'єктивність візуальної інтерпретації, залежність від досвіду оператора та вразливість до когнітивних упереджень призводять до значної варіабельності діагнозів. Визначено критичну важливість балансу між чутливістю (для уникнення пропусків раку) та специфічністю (для запобігання гіпердіагностиці) при розробці алгоритмічних рішень.

Огляд існуючих методів показав, що класичні підходи, засновані на фільтрації та ручному проектуванні ознак (handcrafted features), вичерпали свій потенціал через нездатність охопити всю варіативність клінічних проявів в умовах зашумлених даних. Обґрунтовано перехід до методів глибокого навчання, зокрема згорткових нейронних мереж, як найбільш перспективного напрямку. Доведено, що здатність CNN до автоматичного навчання ієрархічних ознак та стійкість до шуму роблять їх оптимальним інструментом для створення ефективної системи підтримки прийняття лікарських рішень, розробці якої присвячено наступні розділи роботи.

РОЗДІЛ 2

МЕТОДОЛОГІЯ ПОБУДОВИ ТА НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ МЕДИЧНОЇ ДІАГНОСТИКИ

2.1 Парадигма глибокого навчання в ультразвуковій діагностиці

Традиційні підходи до автоматизованої діагностики базувалися на парадигмі ручного конструювання ознак (Feature Engineering), яка вимагала від дослідників формалізації візуальних симптомів хвороби у вигляді чітких математичних правил. Однак, як було показано в попередньому розділі, цей підхід має суттєві обмеження при роботі з ультразвуковими даними через їхню варіативність та зашумленість. Сучасним стандартом вирішення подібних задач стала парадигма глибокого навчання (Deep Learning), яка трансформує фундаментальний принцип взаємодії алгоритму з даними.

В основі глибокого навчання лежить *концепція автоматичного навчання представлень* (Representation Learning). На відміну від класичних методів, де на вхід класифікатора подається вектор розрахованих інженером характеристик (наприклад, ентропія текстури або кривизна контуру), моделі глибокого навчання, зокрема згорткові нейронні мережі (CNN), працюють безпосередньо з «сирими» даними – масивом пікселів зображення [7]. Головна перевага цього підходу полягає в тому, що оптимальний набір ознак для діагностики не задається апріорі, а формується моделлю автоматично в процесі навчання шляхом мінімізації функції втрат.

Цей процес реалізується через ієрархічну архітектуру CNN. На нижчих рівнях (перші згорткові шари) мережа навчається детектувати універсальні геометричні примітиви – локальні градієнти яскравості, лінії та кути, що функціонально нагадує роботу первинної зорової кори мозку. У міру просування вглиб мережі ці прості елементи комбінуються у складніші абстрактні структури: специфічні текстурні патерни, фрагменти меж органів та характерні акустичні артефакти. На найвищих рівнях формуються

семантичні представлення, що відповідають клінічно значущим об'єктам, таким як «гіпоехогенне утворення» або «неоднорідна паренхіма» [8].

Критичною перевагою глибокого навчання для ультразвукової діагностики є специфічний підхід до обробки шумів. У класичній парадигмі спекл-шум розглядався виключно як завада, яку необхідно усунути на етапі попередньої обробки, що часто призводило до втрати діагностично важливих деталей. Згорткові нейронні мережі пропонують інший підхід: замість боротьби з шумом, вони адаптуються до нього. Завдяки великій кількості параметрів та нелінійним функціям активації, CNN здатні моделювати складні залежності між сигналом і шумом, використовуючи статистичні характеристики розподілу спеклу як додаткову текстурну ознаку для диференціації тканин різної щільності.

Важливим аспектом застосування глибокого навчання в медицині є можливість використання *трансферного навчання* (Transfer Learning). Оскільки навчання глибоких мереж «з нуля» вимагає величезних масивів розмічених даних, які є дефіцитними в медичній сфері, ефективним рішенням є використання архітектур (таких як ResNet, VGG, DenseNet), попередньо навчених на великих базах даних загального призначення (наприклад, ImageNet). Здатність згорткових шарів виділяти універсальні візуальні примітиви дозволяє адаптувати ці моделі до завдань ультразвукової діагностики, "довчаючи" лише останні шари для розпізнавання специфічних патологій печінки.

Такий підхід дозволяє реалізувати наскрізний (End-to-End) процес діагностики, де вхідним сигналом є зображення, а вихідним – ймовірність патології [7]. Відмова від етапу ручної сегментації є особливо важливою при аналізі дифузних захворювань печінки або пухлин з нечіткими межами, де точне виділення контуру є неможливим. Мережа аналізує не лише локальну область підозрілого вогнища, а й глобальний контекст – стан навколишньої

паренхіми та деформацію судинного малюнка, що забезпечує вищу точність та об'єктивність діагностичного висновку.

2.2 Теоретичні основи побудови згорткових нейронних мереж

Архітектура згорткових нейронних мереж (CNN), обрана як базис для розробки системи діагностики, є спеціалізованим класом моделей глибокого навчання, оптимізованим для обробки даних із сітковою топологією. На відміну від повнозв'язних мереж (Multilayer Perceptron), які ігнорують просторову структуру вхідного сигналу, CNN базуються на математичних операціях, що дозволяють зберігати та аналізувати локальні кореляції між пікселями. У контексті аналізу ультразвукових зображень це є критично важливим, оскільки діагностична інформація (наприклад, наявність гіпоехогенного обідка навколо пухлини) міститься не в значенні окремого пікселя, а у взаємному розташуванні групи пікселів.

2.2.1 Математична модель згорткового шару

Основним будівельним блоком мережі є згортковий шар, функція якого полягає у виділенні локальних ознак. Замість матричного множення вхідного вектора на матрицю ваг, як це відбувається у класичних мережах, тут застосовується операція дискретної згортки. Нехай вхідне зображення (або карта ознак з попереднього шару) представлено як двовимірна матриця I , а навчальний фільтр (ядро) – як матриця K розмірністю $k_1 \times k_2$. Тоді значення вихідної карти ознак S у точці (i, j) обчислюється за формулою [9]:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - n). \quad (2.1)$$

У практичній реалізації в середовищі TensorFlow, для оптимізації обчислень, операція часто реалізується як крос-кореляція (без «перевертання» ядра), проте в термінології глибокого навчання вона традиційно називається згорткою. Ядро K ковзає по вхідному зображенню з певним кроком (stride), обчислюючи скалярний добуток між своїми вагами та локальною ділянкою зображення (рецептивним полем).

Використання операції згортки забезпечує три фундаментальні властивості, що роблять CNN ефективними для аналізу медичних зображень:

- *Розрідженість взаємодій (Sparse Interactions)*. У повнозв'язних шарах кожен вихідний нейрон пов'язаний з усіма вхідними. У CNN, завдяки використанню ядра розміром, наприклад, 3×3 , кожен нейрон карти ознак залежить лише від 9 пікселів вхідного зображення [9]. Це дозволяє детектувати дрібні деталі, такі як мікрокальцинати або тонкі межі кіст, використовуючи значно меншу кількість параметрів.

- *Спільне використання параметрів (Parameter Sharing)*. Ваги ядра K залишаються незмінними для всіх позицій (i, j) на зображенні. Це означає, що якщо мережа навчилася розпізнавати вертикальну межу (наприклад, стінку судини) у верхньому лівому куті зображення, вона зможе розпізнати таку ж структуру в будь-якій іншій частині знімка [9]. Ця властивість критично важлива для УЗД, де положення органу може варіюватися.

- *Еквіваріантність до трансляції (Equivariance)*. Якщо вхідне зображення зміщується, то й карта ознак зміщується на таку ж величину. Це дозволяє моделі коректно локалізувати патологію незалежно від її розташування в кадрі [9].

У розробленій базовій моделі (Custom CNN) та в адаптованих архітектурах (VGG16, ResNet50) використано фільтри малого розміру – 3×3 пікселі. Такий вибір обумовлений тим, що послідовне застосування кількох шарів з малими ядрами дозволяє збільшити ефективне рецептивне поле мережі, зберігаючи при цьому кількість параметрів на низькому рівні. Наприклад, стек з двох шарів 3×3 має рецептивне поле 5×5 , але містить менше ваг і забезпечує більшу нелінійність, ніж один шар 5×5 .

Важливим гіперпараметром згортки є *падинг (Padding)*. Оскільки операція згортки з ядром $k > 1$ зменшує розмірність вихідної карти, у глибоких мережах це може призвести до швидкої втрати просторової інформації. Для запобігання цьому у досліджуваних моделях застосовано

техніку «same padding» (нульове доповнення). Вона полягає у додаванні рамки з нульових значень навколо вхідного зображення, ширина якої P розраховується так, щоб вихідний розмір карти ознак збігався з вхідним [10]. Це дозволяє зберегти інформацію про крайові ефекти, що важливо для аналізу контурів печінки, які часто знаходяться на периферії скану.

Сама по собі операція згортки є лінійною. Послідовність будь-якої кількості лінійних шарів математично еквівалентна одному лінійному шару, що обмежує можливості моделі лише найпростішими перетвореннями. Для того, щоб мережа могла апроксимувати складну нелінійну функцію, що відображає залежність між пікселями зображення та діагнозом (наприклад, розрізнити складну текстуру гепатоцелюлярної карциноми від цирозу), після кожного згорткового шару застосовується нелінійна функція активації.

У всіх прихованих шарах розроблених моделей використано функцію *ReLU* (*Rectified Linear Unit*). Математично вона визначається як [11]:

$$f(x) = \max(0, x). \quad (2.2)$$

Вибір ReLU замість класичних сигмоїдальних функцій обумовлений двома ключовими факторами, що впливають на ефективність навчання глибоких мереж:

1. *Проблема зникаючого градієнта*. Сигмоїдальні функції мають похідну, яка прямує до нуля при великих за модулем значеннях аргументу (насичення). У глибоких мережах (таких як ResNet50 з 50 шарами) при зворотному поширенні помилки (backpropagation) ці малі градієнти перемножуються, внаслідок чого сигнал помилки швидко згасає, і перші шари мережі перестають навчатися. Похідна ReLU дорівнює 1 для всіх $x > 0$, що забезпечує безперешкодне проходження градієнта крізь глибокі шари.

2. *Обчислювальна ефективність*. Обчислення максимуму є значно швидшою операцією, ніж розрахунок експонент, що прискорює процес навчання на великих масивах даних. Крім того, ReLU сприяє розрідженості

активацій (sparsity), оскільки для від'ємних входів вихід дорівнює нулю, що імітує біологічний принцип енергоефективності нейронів.

Єдиним винятком є модель EfficientNetB0, де замість ReLU використано функцію активації Swish ($f(x) = x \cdot \sigma(x)$) [11]. Ця функція є гладкою та немонотонною, що, за даними останніх досліджень, дозволяє краще оптимізувати глибокі мережі в умовах складного ландшафту функції втрат.

Отже, комбінація лінійної операції згортки (для виділення просторових патернів) та нелінійної активації (для моделювання складних залежностей) формує той фундамент, на якому базується здатність мережі до навчання ознак. Наступним етапом обробки сигналу в CNN є агрегація отриманих ознак та зниження розмірності, що реалізується через шари підвибірки (Pooling), які будуть розглянуті далі.

2.2.2 Агрегація ознак та зниження розмірності

Після виділення локальних ознак згортковими шарами виникає необхідність у зменшенні просторової розмірності карт ознак. Це завдання вирішують шари підвибірки (Pooling Layers). Їх функція є двоякою: по-перше, вони зменшують кількість розрахункових параметрів та обчислювальне навантаження на наступних етапах; по-друге, вони забезпечують інваріантність моделі до незначних геометричних спотворень та зсувів об'єкта. Якщо згортковий шар каже «тут є край», то шар пулінгу узагальнює це до «в цій області є край», роблячи систему менш чутливою до точного положення пікселів.

У розробленій архітектурі Custom CNN та класичних моделях (VGG16) використано операцію Max Pooling. Вона працює шляхом переміщення вікна фіксованого розміру (зазвичай 2×2) по карті ознак і вибору максимального значення з кожного вікна. Математично для вікна W операція записується як $s_j = \max_{i \in W} x_i$. Вибір саме максимуму, а не середнього значення (Average Pooling), обумовлений природою візуальних ознак: максимальне значення відповідає найбільш вираженій присутності шуканого патерну (наприклад,

найяскравішій точці межі). Крім того, Max Pooling діє як своєрідний фільтр шуму, відкидаючи слабкі активації фону.

У сучасних архітектурах, що використовуються для трансферного навчання (ResNet50, DenseNet121), застосовується інший підхід – Global Average Pooling (GAP). Цей шар розміщується перед фінальним класифікатором і обчислює середнє значення для кожної карти ознак в цілому, перетворюючи тривимірний тензор розміром $H \times W \times C$ у одновимірний вектор довжиною C . Цей підхід має критичну перевагу над традиційним «вирівнюванням» (Flatten): він радикально зменшує кількість ваг у повнозв'язній частині мережі, що мінімізує ризик перенавчання. Наприклад, для карти ознак 7×7 з 512 каналами Flatten створив би вхідний вектор розміром 25088, тоді як GAP – лише 512. Для нашого дослідження, де обсяг навчальних даних обмежений, використання GAP у трансферних моделях стало ключовим фактором стабілізації навчання [12].

Фінальним етапом обробки інформації в CNN є класифікація, яку виконують повнозв'язні шари (Dense Layers). На вхід цієї частини мережі подається вектор високорівневих ознак, сформований згортковою базою. Кожен нейрон повнозв'язного шару пов'язаний з усіма елементами попереднього шару, що дозволяє моделювати глобальні залежності між різними частинами зображення. У всіх досліджуваних моделях (як власних, так і трансферних) було використано уніфіковану архітектуру класифікатора: один прихований шар з 256 нейронів (активація ReLU) та вихідний шар.

Головною загрозою при навчанні глибоких мереж на медичних даних є перенавчання (overfitting), коли модель просто «запам'ятовує» тренувальні приклади замість вивчення загальних закономірностей. Для протидії цьому в архітектуру класифікатора інтегровано шар *Dropout*. Під час кожної ітерації навчання цей метод випадковим чином «вимкне» (обнуляє вихід) задану частку нейронів (у нашому випадку коефіцієнт $p = 0.5$) [13]. Це змушує мережу не покладатися на окремі, специфічні ознаки (які можуть бути

шумом), а формувати розподілене, робастне представлення інформації. На етапі тестування (inference) всі нейрони працюють, але їхні виходи масштабуються на коефіцієнт $(1 - p)$.

Вихідний шар мережі містить кількість нейронів, що дорівнює кількості класів діагностики ($K = 3$: Normal, Benign, Malignant). Для перетворення виходів цих нейронів (логітів z_i) у розподіл ймовірностей використовується функція *Softmax*:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}. \quad (2.3)$$

Результатом є вектор, елементи якого знаходяться в діапазоні $[0, 1]$ і в сумі дають 1, що дозволяє інтерпретувати їх як ймовірність приналежності вхідного зображення до відповідного класу.

2.2.3 Математична формалізація процесу навчання

Навчання нейронної мережі – це процес ітеративної оптимізації її вагових коефіцієнтів з метою мінімізації помилки передбачення. Мірою цієї помилки виступає функція втрат (Loss Function). Для задачі багатокласової класифікації стандартом є категоріальна перехресна ентропія (Categorical Cross-Entropy). Для одного зображення вона визначається як [7]:

$$L = - \sum_{c=1}^K y_c \log(p_c), \quad (2.4)$$

де y_c – бінарний індикатор (0 або 1), чи належить зображення до класу c (істинна мітка у форматі one-hot encoding), а p_c – прогнозована моделлю ймовірність для цього класу. Функція штрафує модель логарифмічно сильніше за впевнені, але неправильні передбачення.

Мінімізація функції втрат здійснюється методом градієнтного спуску. У даній роботі використано адаптивний алгоритм оптимізації Adam (Adaptive Moment Estimation). На відміну від класичного стохастичного градієнтного спуску (SGD), який використовує єдину швидкість навчання (learning rate) для всіх ваг, Adam обчислює індивідуальні адаптивні швидкості для кожного параметра на основі оцінок першого (середнє) та другого (незмінена

дисперсія) моментів градієнтів. Це дозволяє алгоритму швидко сходитися на початкових етапах і акуратно наближатися до мінімуму на фінальних, ефективно долаючи "плато" та локальні мінімуми функції втрат, характерні для складних ландшафтів медичних даних.

Додатково, для підвищення стабільності навчання, впроваджено динамічне планування швидкості навчання (Learning Rate Scheduling). Стратегія експоненційного затухання (Exponential Decay) поступово зменшує крок навчання з кожною епохою [7]. Це дозволяє робити великі кроки на старті, швидко навчаючись загальним ознакам, і переходити до мікро-кроків у кінці для точного налаштування (fine-tuning) ваг під специфічні текстурні нюанси патологій печінки [14].

Таким чином, описаний теоретичний базис – від операції згортки до алгоритму Adam – формує цілісну методологію, яка дозволяє створити модель, здатну автоматично виділяти значущі ознаки з зашумлених ультразвукових зображень та виконувати їх високоточну класифікацію.

2.3 Формування та підготовка набору даних для навчання

Ефективність навчання глибоких нейронних мереж у задачах медичної діагностики критично залежить від якості, обсягу та репрезентативності навчальної вибірки. Специфікою ультразвукової діагностики є висока варіабельність зображень, зумовлена різними типами сканерів, налаштуваннями оператора та конституційними особливостями пацієнтів. Для забезпечення робастності розроблюваних моделей до цих факторів було сформовано консолідований набір даних шляхом інтеграції двох відкритих джерел: «Annotated Ultrasound Liver Images Dataset» та «Liver Histopathology Fibrosis Ultrasound Images».

Такий підхід дозволив вирішити проблему недостатньої кількості прикладів для рідкісних патологій та збагатити вибірку проміжними станами

захворювань (стадіями фіброзу), що є необхідним для навчання моделі розпізнаванню тонких текстурних відмінностей.

2.3.1 Структура та класифікація даних

Процес консолідації даних вимагав розробки уніфікованої схеми маппінгу (відображення) діагнозів, оскільки вихідні набори мали різну номенклатуру. Виходячи з клінічної мети системи – диференціації станів, що вимагають негайного втручання, від станів, що підлягають моніторингу, – було застосовано наступну логіку групування класів:

1. Клас «Normal» (Норма). До цієї категорії віднесено зображення здорової паренхіми печінки без ознак жирової інфільтрації або вогнищевих утворень. Також сюди включено зображення зі стадією фіброзу F0 (відсутність фіброзу) з додаткового датасету, оскільки візуально та клінічно вони відповідають здоровій тканині. Це дозволило сформувати еталонний клас текстури для базового рівня порівняння.

2. Клас «Benign» (Доброякісні зміни). Цей клас об'єднав патології, які не несуть безпосередньої онкологічної загрози. Окрім класичних вогнищевих утворень (прості кісти, гемангіоми), до цього класу було інтегровано зображення з клінічно значущими стадіями фіброзу: F1 (портальний фіброз) та F2 (перипортальний фіброз). Таке рішення обґрунтоване тим, що тактика ведення пацієнтів з початковим фіброзом та доброякісними утвореннями є схожою (динамічне спостереження), і це дозволяє навчити нейромережу відрізняти дифузні зміни тканини від агресивних злоякісних процесів.

3. Клас «Malignant» (Злоякісні зміни). Це цільовий клас, що складається виключно з верифікованих випадків гепатоцелюлярної карциноми (ГЦК) та метастатичних уражень. Жорсткі критерії включення гарантують фокусування моделі на вивченні найнебезпечніших патернів: інфільтративного росту, нечітких меж та специфічної «мозаїчної» неоднорідності.

У результаті фільтрації пошкоджених файлів та дублікатів було сформовано фінальний датасет, що містить понад 4300 зображень.

Статистичний аналіз розподілу даних у навчальній вибірці (Train set) виявив характерну для медичних задач диспропорцію класів (рис. 2.1).

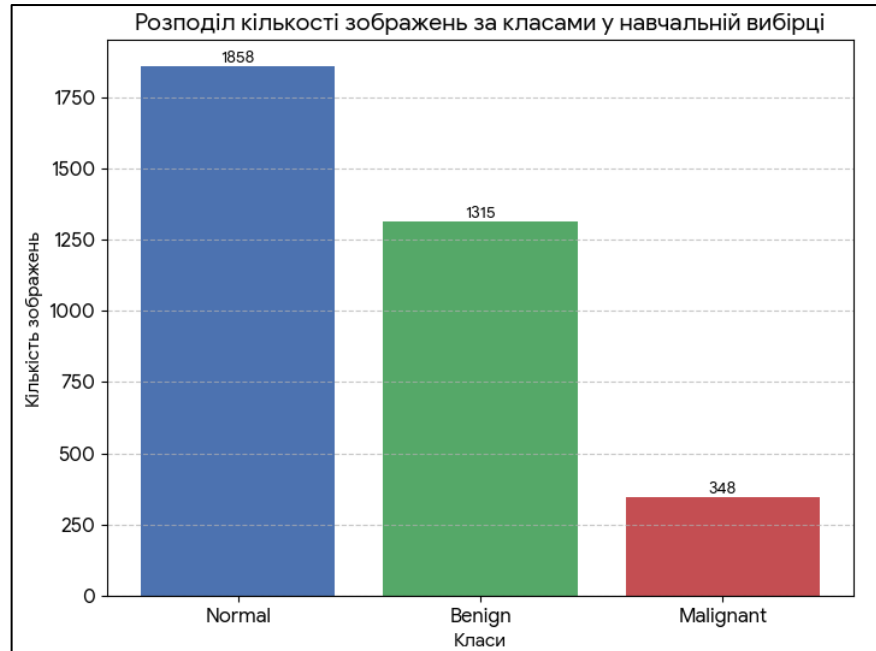


Рисунок 2.1 – Гістограма розподілу зображень за класами у навчальній вибірці

Як видно з діаграми, клас «Normal» налічує 1858 зразків, «Benign» – 1315, тоді як клас «Malignant» представлений лише 348 зображеннями. Співвідношення «норма/патологія» становить приблизно 5.3 до 1. Такий дисбаланс створює ризик упередженості моделі (bias), яка може досягати високої загальної точності шляхом ігнорування мінорного класу. Для компенсації цього ефекту на етапі навчання було застосовано методи зважування класів, опис яких наведено в наступних розділах.

Для об'єктивної оцінки здатності моделі до узагальнення було виділено незалежний валідаційний набір (Validation set), що становить 20% від загального обсягу даних (~800 зображень). Враховуючи сильний дисбаланс класів, просте випадкове перемішування (random shuffle) було неприпустимим, оскільки існував ризик потрапляння критично малої кількості прикладів раку у валідаційну вибірку. Тому було застосовано алгоритм

стратифікованого розбиття (Stratified Split), який гарантує збереження ідентичного відсоткового співвідношення класів у тренувальній та валідаційній підмножинах [16].

Фізична організація даних реалізована у вигляді ієрархічної структури каталогів, де назва папки відповідає семантичній мітці класу (наприклад, `dataset/train/Malignant/image_01.png`). Такий підхід забезпечує сумісність зі стандартними генераторами даних бібліотеки TensorFlow (`image_dataset_from_directory`), які автоматично виконують зчитування файлів та конвертацію міток у векторний формат One-Hot Encoding (наприклад, мітка «Malignant» кодується вектором $[0, 1, 0]$).

2.3.2 Попередня обробка даних

Оскільки консолідований набір даних містив зображення з різною роздільною здатністю та співвідношенням сторін, безпосередня подача їх на вхід нейронної мережі була неможливою. Для уніфікації вхідного потоку даних було розроблено та імплементовано алгоритм препроцесингу, що включає три критичні етапи.

Першим етапом стала геометрична стандартизація. Усі зображення приводилися до єдиного розміру вхідного тензора – 256×256 пікселів. Вибір саме такої роздільної здатності є результатом компромісу між збереженням діагностичної деталізації спекл-шуму та доступними обчислювальними ресурсами (VRAM). Для зміни розміру використано метод білінійної інтерполяції, який забезпечує згладжування пікселів без внесення високочастотних артефактів [17].

Другим етапом стала адаптація колірних каналів. Ультразвукові зображення є одноканальними (відтінки сірого). Однак більшість сучасних архітектур (ResNet, VGG, DenseNet), обраних для трансферного навчання, були попередньо натреновані на кольорових зображеннях ImageNet (формат RGB, 3 канали). Для забезпечення сумісності без модифікації першого згорткового шару (що призвело б до втрати ваг), одноканальні знімки

Extractor) та класифікатора. На вхід мережі подається тензор розміром $256 \times 256 \times 3$. Екстрактор ознак побудований за класичним пірамідальним принципом: зі збільшенням глибини просторова роздільна здатність карт ознак зменшується, а кількість каналів (фільтрів) зростає. Це дозволяє мережі переходити від детекції простих геометричних примітивів на великому зображенні до аналізу складних абстрактних патернів на стиснутому представленні.

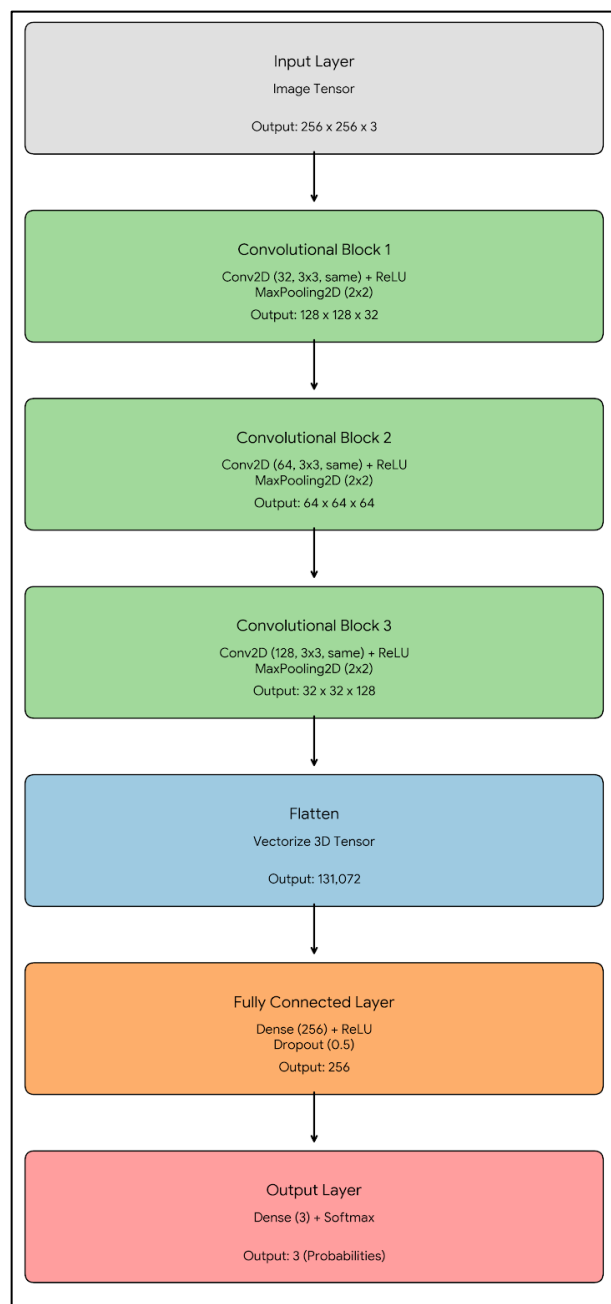


Рисунок 2.2 – Структурна схема архітек

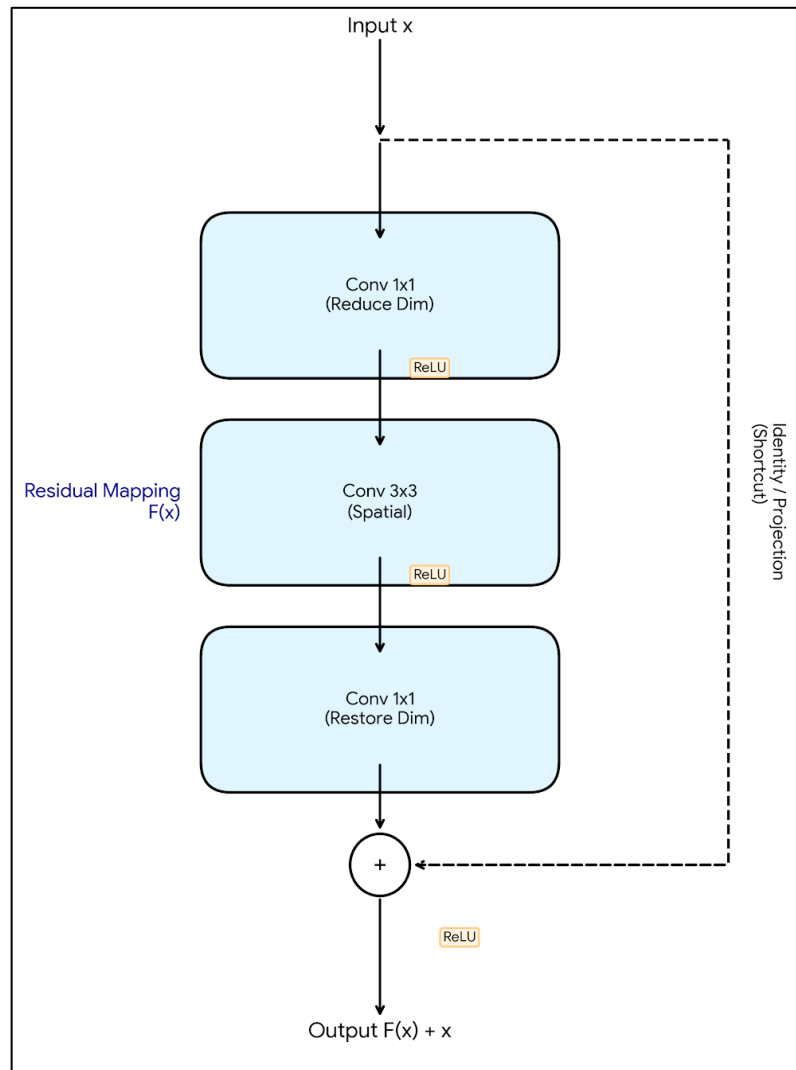


Рисунок 2.3 – Схема залишкового блоку (Residual Block) в архітектурі ResNet

Перевага ResNet50 для ультразвукової діагностики полягає в здатності будувати надзвичайно глибокі ієрархії ознак. Якщо перші шари детектують спекл-шум, то 50-й шар оперує абстракціями високого рівня, такими як «гетерогенна структура» або «інфільтративний край». Це робить ResNet50 потужним інструментом для диференціації складних випадків, наприклад, розрізнення атипових гемангіом від гепатоцелюлярної карциноми, де вирішальну роль відіграють саме високорівневі семантичні відмінності, а не проста яскравість пікселів. Крім того, завдяки Batch Normalization після кожного шару, мережа є менш чутливою до ініціалізації ваг та дозволяє використовувати вищі швидкості навчання.

Для запобігання перенавчанню та економії обчислювальних ресурсів було впроваджено систему зворотних викликів (Callbacks) [16]. Механізм ранньої зупинки (Early Stopping) моніторив метрику точності на валідаційному наборі даних. Якщо точність не покращувалася протягом 5 епох (параметр терпіння), навчання автоматично припинялося, а ваги моделі поверталися до стану, в якому було зафіксовано найкращий результат. Паралельно працював механізм контрольних точок (Model Checkpoint), який зберігав на диск файл моделі лише у випадку покращення валідаційних метрик. Це гарантувало, що для фінального тестування буде використана найкраща версія нейромережі, а не остання, яка могла вже почати перенавчатися.

Загальна тривалість циклу навчання для кожної моделі була обмежена 40 епохами, при розмірі пакету (batch size) 32 зображення. Цей розмір пакету був обраний як компроміс між стабільністю оцінки стохастичного градієнта та обмеженнями відеопам'яті GPU при роботі з глибокими мережами. Таким чином, розроблена методика навчання об'єднує передові підходи трансферного навчання з тонким налаштуванням процесу оптимізації під специфіку незбалансованих медичних даних, створюючи надійну основу для отримання високоточних діагностичних моделей.

Висновки за розділом 2

У другому розділі роботи розроблено комплексну методологію створення системи автоматизованої діагностики патологій печінки. Обґрунтовано перехід від класичних методів інженерії ознак до парадигми глибокого навчання, що дозволяє ефективно працювати в умовах високого рівня шукл-шуму. Сформовано репрезентативний набір даних шляхом консолідації різномірних джерел та застосовано методи попередньої обробки для уніфікації вхідних тензорів.

Детально проаналізовано та адаптовано вісім архітектур нейронних мереж, що охоплюють як власні розробки (Custom CNN), так і сучасні трансферні моделі (ResNet, DenseNet, MobileNet). Розроблено уніфікований протокол навчання, який включає використання адаптивного оптимізатора Adam, динамічного планування швидкості навчання та стратегії зважування класів для подолання дисбалансу даних. Впровадження механізмів регуляризації та ранньої зупинки забезпечує стійкість моделей до перенавчання. Описаний методичний апарат є повністю готовим до проведення обчислювальних експериментів, результати яких будуть проаналізовані в наступному розділі.

сильного дисбалансу класів (де норма переважає над патологією у співвідношенні 5:1), модель може досягти високої Accuracy, просто ігноруючи клас хворих, що робить її клінічно непридатною.

Тому для комплексного та об'єктивного аналізу розроблених моделей було використано розширену систему метрик:

1. Точність (Accuracy) – базовий індикатор, що показує частку правильних відповідей серед усіх прогнозів моделі. Розраховується як відношення суми істинно-позитивних (TP) та істинно-негативних (TN) результатів до загальної кількості зразків [26]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \quad (3.1)$$

Ця метрика використовувалася для загального порівняння стабільності моделей, але не як вирішальний критерій вибору.

2. Чутливість (Recall / Sensitivity): показує здатність моделі виявляти позитивний клас [26]. Для класу «Malignant» (Злоякісні) ця метрика є індикатором діагностичної безпеки: чим вищий Recall, тим менше шансів пропустити рак.

$$Recall = \frac{TP}{TP+FN}. \quad (3.2)$$

Високий показник Recall є критичною вимогою для скринінгових систем, оскільки ціна пропуску (FN) є максимальною.

3. Прогностична цінність позитивного результату (Precision): характеризує точність спрацювань моделі, показуючи, яка частка об'єктів, класифікованих як «патологія», дійсно є патологією. Високий Precision означає низький рівень гіпердіагностики (False Positives) [26].

$$Precision = \frac{TP}{TP+FP}. \quad (3.3)$$

Ця метрика важлива для мінімізації ятрогенії – шкоди, завданої пацієнту внаслідок непотрібних інвазивних втручань (біопсій), призначених через помилку алгоритму.

4. F1-міра (F1-score): гармонічне середнє між Precision та Recall. Ця метрика є інтегральним показником якості, особливо важливим для незбалансованих даних, оскільки вона «карає» модель як за пропуски, так і за хибні спрацювання, не дозволяючи досягти високого результату за рахунок лише одного з показників [26].

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3.4)$$

При розрахунку підсумкових метрик для багатокласової класифікації використовувалося зважене усереднення (weighted average), яке враховує кількість прикладів у кожному класі (support), що дозволяє нівелювати вплив дисбалансу вибірки.

5. Матриця невідповідностей (Confusion Matrix): це таблиця розміром $N \times N$ (де $N = 3$ – кількість класів), яка дозволяє візуалізувати структуру помилок [26]. Вона дає можливість зрозуміти не просто скільки разів помилилася модель, а як саме вона помилилася: чи плутає вона рак з доброякісною кістою (менш небезпечно), чи рак з нормою (критично небезпечно). Інтерпретація елементів матриці для задач онкології наведена в таблиці 3.1.

Таблиця 3.1

Інтерпретація елементів матриці помилок для медичної діагностики

Прогноз моделі	Реальний стан: Malignant	Реальний стан: Benign/Normal
Malignant	True Positive (TP): Правильне виявлення раку.	False Positive (FP): Гіпердіагностика. Призначення зайвої біопсії.
Benign/Normal	False Negative (FN): Пропуск раку. Критична помилка.	True Negative (TN): Правильне підтвердження здоров'я.

Використання описаного програмного інструментарію та системи метрик дозволило провести коректний порівняльний аналіз восьми архітектур нейронних мереж, мінімізуючи суб'єктивність оцінювання та фокусуючись на клінічній значущості результатів. Отримані кількісні показники стали основою для висновків, викладених у наступних підрозділах.

3.2 Порівняльний аналіз результатів навчання моделей

Етап навчання нейронних мереж є критичною фазою експериментального дослідження, під час якої відбувається оптимізація мільйонів вагових коефіцієнтів для мінімізації функції втрат. Аналіз динаміки цього процесу дозволяє не лише констатувати кінцеву точність, але й зрозуміти «поведінку» моделі: наскільки швидко вона збігається до оптимуму, чи схильна вона до механічного запам'ятовування даних (перенавчання) та наскільки стабільно вона узагальнює знання на нові приклади [26]. У цьому підрозділі проведено детальний розбір історії навчання (learning history) для ключових архітектур та представлено зведені результати тестування на незалежній вибірці.

3.2.1 Аналіз динаміки навчання та збіжності алгоритмів

Усі досліджувані моделі навчалися протягом 40 епох з використанням ідентичних гіперпараметрів (оптимізатор Adam, динамічне зменшення Learning Rate, розмір пакету 32), що забезпечило рівні умови для порівняння. Однак характер кривих навчання суттєво відрізнявся залежно від типу архітектури (навчання з нуля проти трансферного навчання).

Динаміка базової моделі (Custom CNN): найбільший інтерес становив аналіз поведінки моделі Custom CNN, яка ініціалізувалася випадковими вагами. Графіки залежності точності та функції втрат від епохи демонструють класичну картину успішного навчання глибокої мережі (рис. 3.1).

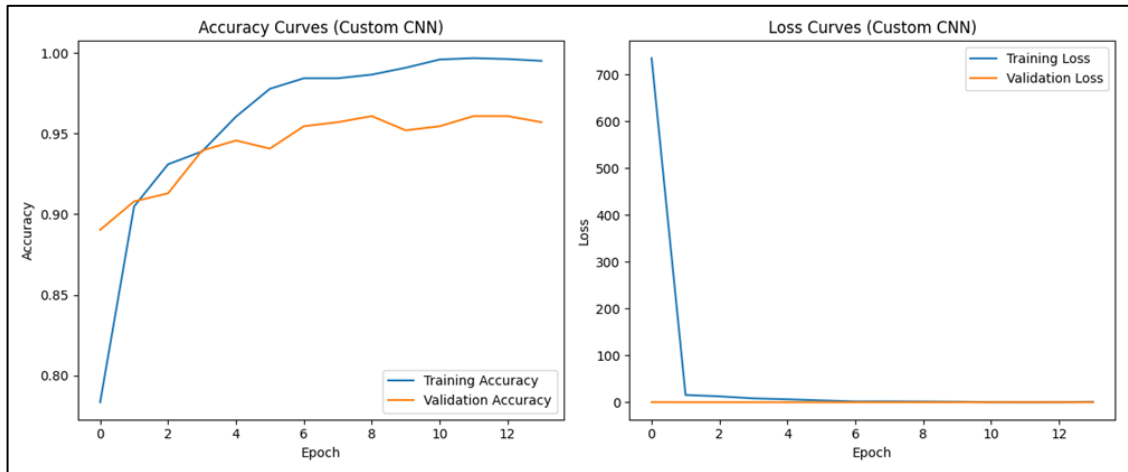


Рисунок 3.1 – Динаміка навчання моделі Custom CNN: (а) графік точності, (б) графік функції втрат

На початковому етапі (1–15 епохи) спостерігається фаза активного навчання: функція втрат стрімко падає, а точність зростає майже лінійно. Це свідчить про те, що спроектована архітектура з трьох згорткових блоків успішно захоплює низькорівневі візуальні примітиви (градієнти, плями) безпосередньо з «сирих» пікселів. Починаючи з 16-ї епохи, криві виходять на плато стабілізації. Критично важливим є те, що розрив (gap) між кривою навчання (Training Accuracy) та кривою валідації (Validation Accuracy) залишається мінімальним протягом усього процесу. Це є прямим наслідком ефективної роботи шару Dropout (0.5), який запобігає ко-адаптації нейронів. Відсутність різких осциляцій функції втрат на валідації підтверджує, що модель не перенавчається, а формує робастні ознаки, стійкі до спекл-шуму. Фінальна точність на валідації досягла 96.97%, що є найвищим показником серед усіх учасників експерименту.

Динаміка трансферних моделей (на прикладі VGG16): процес навчання моделей, побудованих на основі попередньо навчених мереж (VGG16, ResNet50), мав принципово інший характер. Оскільки згорткова база була «заморожена», оптимізації

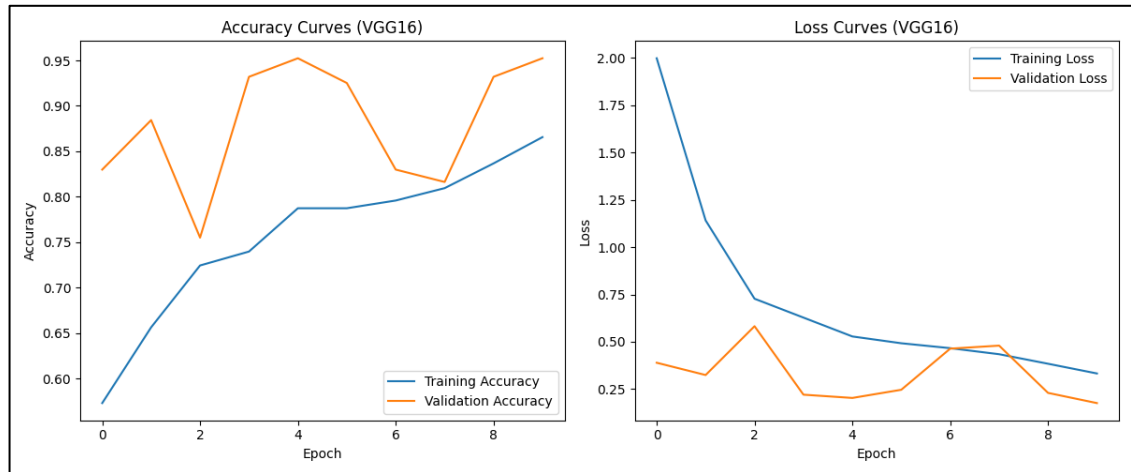


Рисунок 3.2 – Динаміка навчання моделі VGG16

Як видно з рисунка 3.2, трансферні моделі демонструють ефект «швидкого старту»: вже на перших 3–5 епохах валідаційна точність сягає рівня 85–90%. Це пояснюється тим, що мережа вже володіє сформованими детекторами

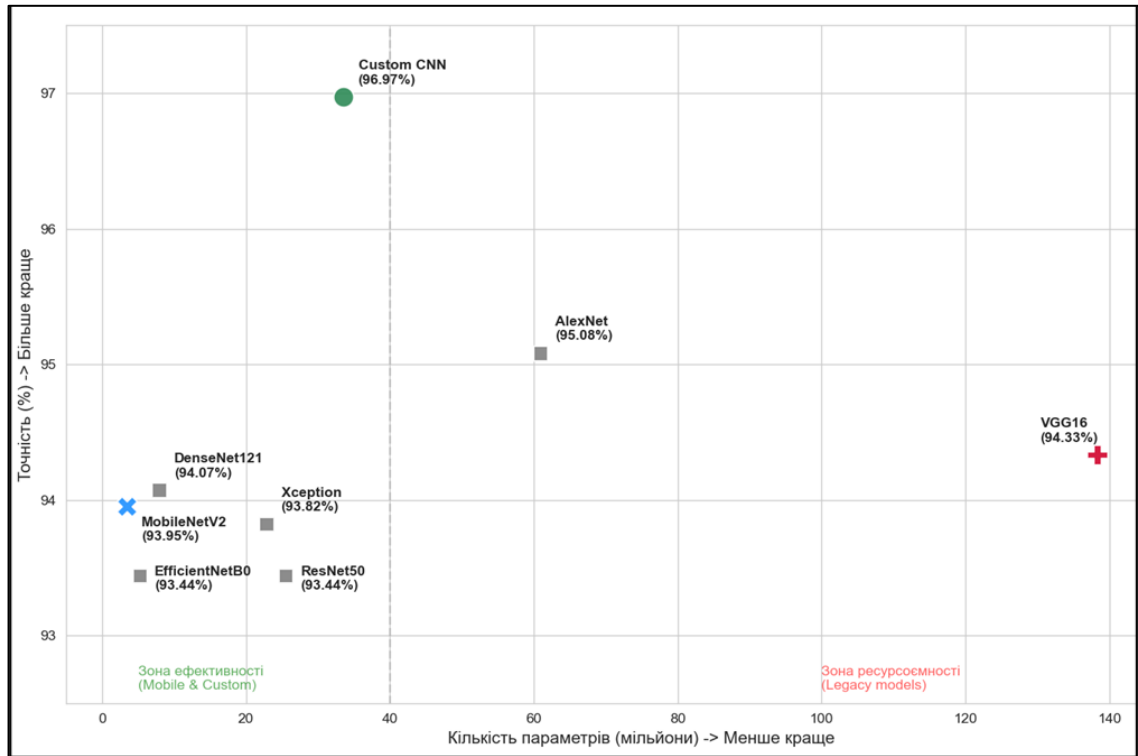


Рисунок 3.4 – Співвідношення точності та розміру моделей (Accuracy vs Model Size)

Аналіз діаграми дозволяє чітко виділити кластери моделей за їхньою ефективністю:

1. *Лідер ефективності (Efficiency Leader): Custom CNN.* Ця модель знаходиться у лівому верхньому куті графіка, що є ідеальною позицією. Вона забезпечує найвищу точність (96.97%) при відносно невеликій кількості параметрів (~30 млн). Вона швидко навчається і швидко працює (низький Inference Time), що робить її найкращим вибором для стаціонарних робочих станцій лікарів-діагностів.

2. *Лідер мобільності (Mobility Leader): MobileNetV2.* Ця модель має найменший розмір (лівий край графіка, < 4 млн параметрів), втрачаючи лише ~3% точності порівняно з лідером. Мінімальний обсяг займаної пам'яті та висока швидкість обробки роблять її безальтернативним кандидатом для мобільних додатків та портативних сканерів.

