

СЕКЦІЯ «ПРИКЛАДНА МАТЕМАТИКА»

УДК 004.93'1:004.852:004.932

DOI 10.31651/2076-5886-2025-1-4-18

PACS 87.57.N-, 07.05.Mh

ЮРЦУН Світлана Олександрівна
студентка спеціальності «Прикладна
математика» Черкаського національного
університету імені Богдана Хмельницького
e-mail: yurtsun.svitlana1621@vu.cdu.edu.ua

БОГАТИРЬОВ Олександр Олегович,
кандидат фізико-математичних наук,
доцент, доцент кафедри прикладної
математики та інформатики Черкаського
національного університету імені Богдана
Хмельницького
e-mail: a.o.bogatyrev@gmail.com
ORCID 0000-0002-2590-6023

**ДОСЛІДЖЕННЯ СТІЙКОСТІ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ
ДО ДАНИХ З РІЗНИХ ДОМЕНІВ
У ЗАДАЧІ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МЕЛАНОМИ
НА ДЕРМАТОСКОПІЧНИХ ЗОБРАЖЕННЯХ**

У статті розглянуто проблему автоматизованого виявлення меланому на дерматоскопічних зображеннях із використанням методів глибокого навчання в умовах обмеженої кількості даних, що може сприяти ранній діагностиці небезпечного захворювання. Основною метою є пошук шляхів підвищення узагальнюючої здатності моделей глибокого навчання при класифікації даних з різних джерел. В роботі використано два набори дерматоскопічних зображень з ISIC Archive в різних поєднаннях. Проведено серію експериментів із застосуванням трансферного навчання CNN та трансформерної архітектури ViT, власної згорткової моделі, ансамблів моделей. Особливу увагу приділено впливу на якість класифікації попередньої обробки зображень (видалення чорних контурів, аугментації), використання доповнення даних з інших джерел тільки для класу меншин та для всіх класів, а також узагальненню моделей на міждоменних даних. Результати показали перевагу трансферного навчання, тоді як власна модель поступалася за точністю. Серед згорткових мереж найбільшу якість класифікації продемонструвала архітектура EfficientNet, також перспективними є моделі на основі Vision Transformer, які показали кращий результат порівняно з більшістю CNN архітектур. Дослідження виявило, що поєднання даних із різних джерел може знижувати ефективність класифікації через навчання на відмінностях доменів замість медичних ознак, особливо їх вплив спостерігається при доповненні тільки класу меншин з інших наборів даних з метою зменшення наслідків дисбалансу класів при навчанні. Завдяки видаленню шумів на зображеннях у вигляді чорного фону, характерного для одного з наборів даних, вдалося підвищити якість розпізнавання захворювання. Результати дослідження свідчать про необхідність ретельної підготовки даних та застосування спеціальних стратегій до міждоменного навчання для практичного застосування.

Ключові слова: глибоке навчання, трансферне навчання, комп'ютерний зір, згорткові нейронні мережі, автоматичне розпізнавання меланому, домен даних, дисбаланс класів, класифікація медичних зображень.

Вступ

Останні роки відзначаються стрімким розвитком комп'ютерного зору в медичній діагностиці, зокрема у виявленні меланоми на дерматоскопічних зображеннях. Завдяки глибокому навчанню, особливо CNN, з'явилася можливість автоматизованого аналізу, що сприяє ранньому виявленню захворювання. Хоча класичні CNN-архітектури продемонстрували високу точність, вони страждали від перенавчання, слабкої узагальнюваності та проблеми дисбалансу класів. Значним проривом став Vision Transformer (ViT), який у сценаріях трансферного навчання показав конкурентні результати й кращу генералізацію на міждомених даних, що критично для практичного застосування [1, 2]. Водночас залишаються виклики, пов'язані з дисбалансом класів, варіативністю даних та обмеженою інтерпретованістю моделей, що ускладнює їхнє впровадження у клінічну практику [3, 4].

Мета статті – дослідження впливу поєднання міждомених даних та попередньої обробки зображень на якість класифікації зображень методами глибокого навчання на прикладі задачі про раннє виявлення меланоми.

Виклад основного матеріалу

Підбір датасетів дерматоскопічних зображень, доповнення даних. Для навчання моделі було використано International Skin Imaging Collaboration (ISIC) Archive [5, 6], зокрема ISIC Challenge Datasets (2016–2024) – набори даних, що містять декілька класів шкірних утворень, підготовлені для щорічних змагань з автоматичного аналізу медичних зображень шкірних уражень, з яких для розробки моделі відібрано набір 2020 року, який містить 33126 екземплярів доброякісних та злоякісних утворень шкіри від більш ніж 2000 пацієнтів, а також розглянуто набір 2019 року для доповнення даних. Обраний розмір зображень – 256x256 пікселів, всі дані були поділені на два класи: «меланома» та «інші».

У даних за 2020 рік присутній значний дисбаланс класів (1,78% клас меланоми та 98,22% інші утворення), що суттєво перешкоджає продуктивності машинного навчання, впливаючи на процеси прийняття рішень. За таких умов доцільним є пошук та використання додаткових даних, що зменшать незбалансованість та сприятимуть підвищенню якості класифікації.

У деяких випадках доповнення даних застосовується для збільшення вибірок у класі меншин для покращення продуктивності класифікації. З експериментального аналізу застосування доповнення даних лише до класу меншин в наборі даних про рак молочної залози було виявлено, що таке доповнення виявляється ефективним для незбалансованих наборів даних [7].

Зважаючи на дисбаланс класів, було вирішено доповнити набір 2020 року зображеннями позитивного класу з набору 2019 року, який складається з зображень класу меланоми (17,85%) та інших (82,15%).

Набори зображень, на яких навчались та випробувались моделі, можна розділити на два види:

- набір 1, що включає всі дані 2020 року та зображення тільки позитивного класу 2019 року;
- набір 2, що об'єднує всі дані за обидва роки (рис. 1, 2).

Під час випробування моделей з цих наборів використовувались вибірки різного розміру та структури, з обробкою та перетвореннями зображень.

Розробка моделі глибокого навчання проводилась тільки з використанням зображень без метаданих. Для валідації використано 4,4-6% навчальних даних із збалансованими класами.

Для зменшення дисбалансу класів на рівні даних виконувалась аугментація зображень зі стандартними перетвореннями (повороти, відображення, масштабування, обрізка, зсув з заповненням, зміна освітлення, контрасту, відтінку). Перетворення в освітленні та кольорах використані мінімальні, щоб помірно змінити зображення, не спотворюючи, та зберігаючи реалістичність, адже модель повинна працювати з реальними даними. Обрізка темних країв зображень як на тренувальних, так і на тестових даних, у більшості випадків сприяла покращенню розпізнавання зображень з меланомою.

На рівні моделі використано додавання ваг класів та розрахунок початкового зміщення (initial bias).

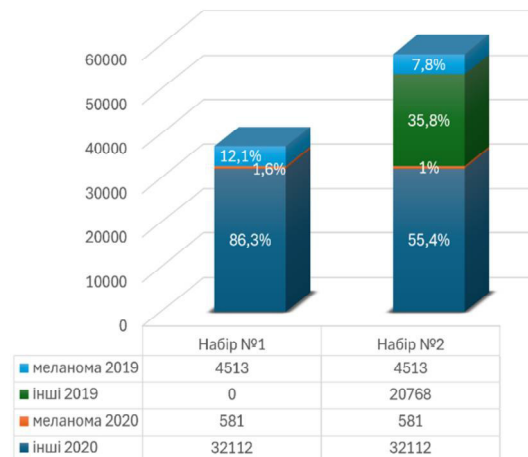


Рис. 1. Структура датасетів, використаних для навчання моделей

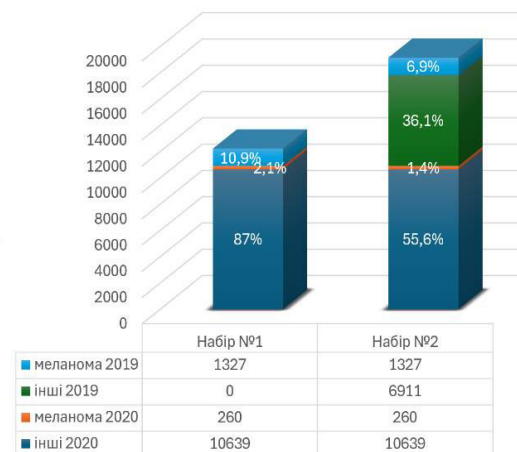


Рис. 2. Структура датасетів, використаних для оцінки моделей

Побудова моделі нейронної мережі з використанням трансферного навчання та аналіз помилок класифікації. Розробка моделей машинного навчання може йти двома основними шляхами: навчання з нуля та донавчання готових моделей.

Оскільки задача побудови моделі для діагностики меланом на дерматоскопічних зображеннях є специфічною з одного боку та супроводжується обмеженою кількістю наявних розмічених даних, особливо для позитивних випадків, з іншого, з метою порівняння в роботі використано два підходи до розробки моделі – навчання з нуля та донавчання стандартних моделей з API бібліотеки Keras з різними архітектурами згорткових мереж. В обох підходах використовувалась як зміна параметрів навчання, архітектур, так і зміна набору даних.

Після навчання на вибірці та тестування 8 моделей комп'ютерного зору, представлених в бібліотеці Keras [8], найкращу за метриками та оптимальну за кількістю параметрів модель EfficientNetV2B0 було обрано для навчання на повному наборі даних № 1. За допомогою перетворення зображень менший клас розширено з 4075 до 20375 екземплярів. Після донавчання з розмороженням спочатку 121 шару базової моделі (1 блок), а потім збільшенням до 130 шарів вдалося отримати метрики оцінки класифікації: precision – 0.915, recall – 0.849, AUC – 0.9692, AUC-PRC – 0.9303 (рис. 3).

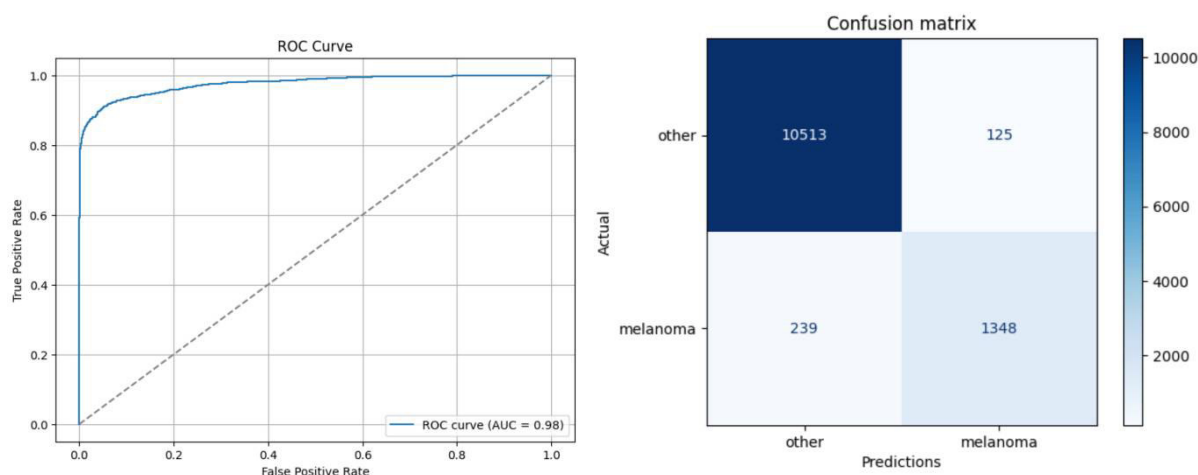


Рис.3. Матриця розбіжностей та ROC-AUC крива EfficientNetV2B0 після навчання на даних з набору №1

Аналіз помилок класифікації (false positives/negatives). Для аналізу помилок моделей була застосована візуалізація передбачень моделей Xception та EfficientNetV2B0 (рис. 4). Виявлено, що серед правильно класифікованих (True Positive) прикладів більшість мають вигляд зображення в колі з темним оточенням, хибно негативні (False Negative) екземпляри таких ознак не мають.

Дерматоскопічні зображення отримують за допомогою дермоскопії – дослідження з використанням оптичного інструменту спостереження, що застосовується для оцінки дрібних деталей шкірних захворювань [9]. Різні шуми, артефакти на зображеннях захворювань шкіри, що отримані з різнорідних джерел, створюють проблеми для надійності моделей у завданні класифікації раку шкіри. При навчанні на високоякісних наборах даних про ураження шкіри модель глибокого навчання може досягти високого рівня діагностики. Однак, оскільки модель класифікації раку шкіри чутлива до зображень, отриманих за допомогою різних пристроїв, налаштувань освітлення та фонів, вона часто не демонструє задовільних результатів класифікації при тестуванні з різними зображеннями.

Серед зображень позитивного класу, більшість яких взята з датасету 2019 року в наборі № 1, тобто іншого домену даних, досить багато знімків, що виглядають як ділянка шкіри з утворенням в колі на чорному фоні, зображення набору 2020 року в більшості такого чорного фону від апарату не мають, таким чином розподіл зображень з чорним фоном та без нього у різних класах суттєво відрізнявся, адже джерелом даних для негативного класу був тільки набір 2020 року. Модель досить впевнено відрізняла зображення з набору 2019 року як позитивний клас (рис. 4), що дає підстави вважати, що такий тип зображень може хибно класифікуватись як ознака меланоми та свідчить про наявність значних відмінностей в наборах даних, а класифікація зміщена на пошук доменної різниці, а не захворювання.

Для подальшої перевірки передбачень моделі відібрано 2963 зображення негативного класу (інші) з набору 2019 року з чорним оточенням із дерматоскопа, на яких модель не навчалася. Обрані зображення були класифіковані як меланома з впевненістю до 100% (рис. 5), тільки 4 знімки були правильно класифіковані як негативний клас. Після перевірки на всьому тестовому наборі 2019 року було підтверджено, що модель погано класифікує саме негативний клас з цього набору, у той час як з набору 2020 року негативний клас визначався більш впевнено.

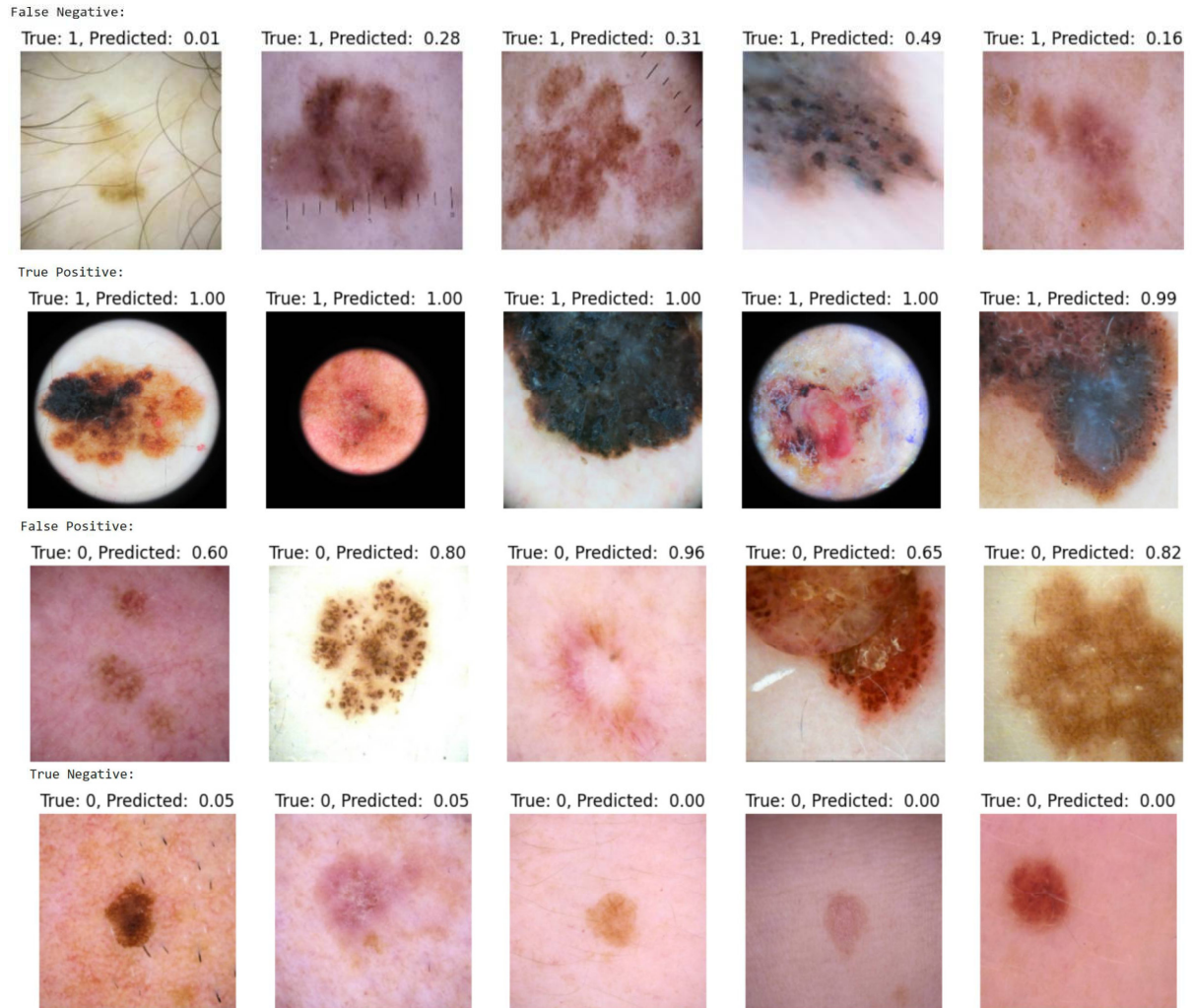


Рис. 4. Передбачення моделі Xception після донавчання на вибірці з набору даних №1

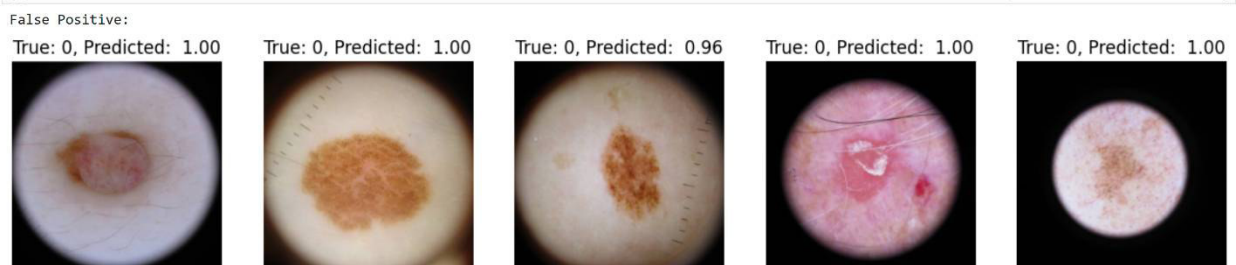


Рис.5. Зображення негативного класу з набору 2019, неправильно класифіковані моделлю EfficientNetV2B0 як меланома

Оцінка впливу аугментації та модифікації тренувального та тестового наборів на продуктивність моделі. Для підвищення стійкості моделі до виявлених артефактів випробувано два підходи:

- проаналізувати присутність зображень з темним оточенням в обох наборах даних та вирівняти їх розподіл для обох класів частковою зміною набору даних, додавши до негативного класу екземпляри саме з темним фоном з набору 2019 року, та застосувати до частини зображень цього класу штучний чорний фон, як аугментацію;
- змінити набір даних, додаванням всіх зображень за 2019 рік, що збільшить дисбаланс класів, але внесе різноманіття в негативний клас, що дозволить узагальнити модель.

Для першого підходу проведено експеримент з маскуванням. За допомогою бібліотеки OpenCV для python розроблено та додано до зображень маски у вигляді темних контурів різної прозорості.

У навчальному наборі даних № 1 негативний клас представлений 32 112 зображеннями, серед яких 106 були попередньо класифіковані як такі, що мають темний фон. Проте в результаті візуальної перевірки з'ясовано, що більшість із них містять значну кількість волосся або мають недостатнє освітлення. Натомість 21,4% зображень меланом (1 089 прикладів) дійсно характеризуються наявністю темного контуру. З метою вирівнювання представленості зображень із темним фоном у обох класах, з негативної підмножини датасету 2019 року було вилучено 2 963 зображення з темним фоном, а до 3 337 зображень застосовано аугментацію, що передбачає штучне додавання темного контуру. Цими даними був доповнений набір № 1.

З метою оцінки впливу модифікацій у навчальному наборі було застосовано підхід трансферного навчання моделі MobileNet на випадковій вибірці з доповненого набору, що дозволило порівняти зміну результатів. Точність та повнота класифікації знизилась з 0.695 до 0.5828 та з 0.833 до 0.491 відповідно. Нові дані та додатковий шум негативно вплинули на продуктивність моделі, вона адаптувалась до відмінностей у даних з різних доменів.

Для того, щоб визначити, як впливає зменшення шуму в даних за рахунок обрізки чорних країв на роботу моделі, проведено декілька експериментів.

Навчальний набір №1 було доповнено даними негативного класу 2019 року, на яких присутнє чорне оточення (2963 екземпляри, збільшено на 9,5%). В тестовий набір включено всі дані для тестування з наборів обох років (додано 6911 прикладів негативного класу, збільшено на 64%). За допомогою бібліотеки OpenCV проведена детекція наявності та ширини чорного контуру та його обрізка для всіх таких зображень у тренувальному наборі.

На випадковій вибірці з отриманого набору №1 проведено навчання моделі EfficientNetV2B0 для вилучення ознак (feature extraction), гіперпараметри навчання, окрім корекції ваг класів, не змінювались. Набір для оцінки моделі також перевірено на наявність артефактів у вигляді чорного оточення, в тестовій вибірці були присутні 11,5% таких прикладів для класу «інші» та 33,2% в класі «меланома». Оцінки отриманої моделі замірялись до та після обрізки контуру на тестовому наборі (табл.1).

Модель стала частіше помилково класифікувати негативні зображення як меланому (позитивний клас), адже нові дані з іншого набору відносяться до таких, які модель більше бачила саме в позитивному класі. Значно зросла кількість хибно позитивних діагнозів, але одночасне зниження повноти класифікації (recall) говорить про те, що модель стала менш впевненою у своїх передбаченнях позитивного класу та менше переоцінює наявність меланом. Обрізка зображень тестового набору позитивно вплинула на якість класифікації.

Для перевірки зміни якості моделі при навчанні на тренувальному наборі, що пройшов обрізку темних країв, було використано вибірку з набору № 1 без доповнення

новими даними. Після навчання EfficientNetV2B0 спостерігалось покращення точності класифікації після обрізки чорних країв зображень навчального набору: precision зріс з 0.9133 до 0.9294, але recall впав з 0.8162 до 0.8072. Також ця модель була перевірена на тестовому наборі 2019 року, в якому саме негативний клас відносився до іншого домену даних. В результаті, тільки 24% зображень, що не є меланомою, були класифіковані як негативні, 98% помилок моделі – віднесення до позитивного класу.

Таблиця №1

Результати оцінки моделі EfficientNetV2B0 після feature extraction на вибірці з набору даних №1

	Повний тестовий набір до застосування обрізки контуру	Повний тестовий набір після застосування обрізки контуру	Тестовий набір до доповнення негативного класу
precision	0.1985	0.2156	0.7715
recall	0.5315	0.6486	0.5324
AUC	0.7779	0.8201	0.8936
AUC-PRC	0.2416	0.2794	0.7113

Таким чином, обрізка, як видалення шумів та артефактів, покращує точність передбачень, але при змішуванні даних модель погано розпізнає приклади з іншого домену.

Аналіз якості класифікації при навчанні на міждомених даних. Узагальнююча здатність моделі значною мірою визначається як обсягом, так і різноманіттям навчальних даних. З цією метою було сформовано набір № 2, який об'єднує всі зображення з наборів 2019 та 2020 років, що належать до різних доменів. До зображень було застосовано обрізання країв з чорними пікселями для видалення зайвого шуму, аугментацію, причому інтенсивність та типи, кількість перетворень варіювались з метою досягнення оптимального балансу між класами, уникнення створення нових артефактів, що значно спотворюють зображення. Такий підхід спрямований на підвищення уваги моделі до представників позитивного класу (меланома) без введення надмірного перекосу. Крім того, розширення обсягу навчальних прикладів дозволяє ефективно використовувати більш складні архітектури, здатні вилучати високорівневі ознаки, що важливо для точної класифікації, але потребує значної кількості даних.

Трансферне навчання моделі EfficientNetV2B0. Для донавчання найбільш результативної моделі EfficientNetV2B0 на датасеті № 2 було спочатку застосовано аугментацію до 50% зображень негативного класу та всіх зображень позитивного класу, співвідношення класів 3:1 відповідно, параметри навчання становили: learning rate = $1e-4$ на етапі transfer learning та $1e-5$ із поступовим зменшенням на етапі fine-tuning, batch size = 32, розморожено 1 блок моделі. Проте, вже на початкових етапах fine-tuning спостерігалось швидке перенавчання: валідаційна помилка почала зростати, а розрив між тренувальними та валідаційними метриками суттєво збільшився, що вказує на недостатню здатність моделі до узагальнення в поточних умовах навчання. Була зроблена спроба підвищення здатності моделі розрізняти класи з розморожуванням двох блоків моделі EfficientNetV2B0 (195 шарів) та збільшенням кількості даних за рахунок аугментації всього набору даних №2, а не тільки позитивного класу, що змінило співвідношення негативного та позитивного класів 4:1. Отримана модель помітно підвищила повноту розпізнавання меланоми – з 0.4921 до 0.6106, але за

рахунок збільшення позитивних передбачень, в тому числі і для негативного класу – точність впала з 0.5713 до 0.4872 (табл. 2).

Для того, щоб змістити фокус уваги моделі до складних екземплярів, проаналізовано передбачення найкращої з отриманих моделей, серед них обрано всі екземпляри, на яких модель помилялась, а також такі, що були класифіковані недостатньо впевнено (з більшою помилкою), для чого відібрані правильно негативні відповіді з передбаченою ймовірністю більше 0.25 та правильно позитивні передбачення з ймовірністю менше 0.75. Кількість таких екземплярів склала близько 23% з 88551 прикладів, з урахуванням аугментованих зображень, на яких проводилось передбачення. До отриманих екземплярів додано 30% зображень, на яких модель була впевнена. Отриманий набір використано для трансферного навчання EfficientNetV2B0, швидкість навчання, починаючи з $1e-5$, поступово знижувалась, розмір пакету зменшено до 16, функція втрат змінена з binary cross entropy на binary focal cross entropy.

Таблиця №2
Результати оцінки моделей після fine-tuning на наборі даних № 2

	EfficientNetV2B0				ViT		Власна модель	
	Після розморожування 1 блоку		Після розморожування 2х блоків та збільшення кількості перетворених екземплярів	Після fine-tuning на вибірці зі складними екземплярами	До обрізки контуру на тестовому наборі	Після обрізки контуру на тестовому наборі	До обрізки контуру на тестовому наборі	Після обрізки контуру на тестовому наборі
	до обрізки контуру на тестовому наборі	після обрізки контуру на тестовому наборі						
precision	0.5755	0.5713	0.4872	0.5977	0.5635	0.5493	0.4407	0.4807
recall	0.4033	0.4921	0.6106	0.4203	0.5035	0.5406	0.4707	0.4386
AUC	0.8705	0.9036	0.8989	0.8654	0.8614	0.8720	0.8611	0.8433
AUC-PRC	0.5041	0.5537	0.5641	0.5183	0.5532	0.5654	0.4651	0.4640

Оцінка моделі проводилась до та після обробки тестового набору обрізкою чорного оточення. На тестовому наборі № 2 після навчання EfficientNetV2B0 отримано метрики, представлені в таблиці 2.

Після обробки (обрізка) тестових даних модель продемонструвала незначне зниження точності (precision) на фоні суттєвого підвищення повноти (recall), що свідчить про покращення здатності моделі виявляти приклади позитивного класу. Зросла кількість дійсно позитивних передбачень (true positive) та зменшилася кількість пропущених випадків меланом (false negative). Водночас незначно збільшилася кількість хибно позитивних передбачень, що негативно вплинуло на точність. Підвищення значення AUC вказує на загальне покращення якості класифікації незалежно від обраного порогу. У більшості протестованих моделей видалення обрізкою чорного контуру зі зображень сприяло помітному покращенню показника recall, тобто кращому виявленню меланом, однак це часто супроводжувалося невеликим зниженням precision через збільшення кількості хибно позитивних відповідей (рис. 6).

Донавчання на складних екземплярах не підвищило продуктивність, розпізнавання меланом погіршилось, хоча точність моделі зросла.

Трансферне навчання ViT моделі. Для реалізації ViT моделі було використано бібліотеку Transformers від Hugging Face, яка дає можливість працювати з моделями трансформерів для різних завдань глибокого навчання, в тому числі для завдань комп'ютерного зору.

Для подальшого донавчання базовою була обрана модель медичного спрямування з бінарної класифікації, яка дозволяє розрізнити, чи виявляють зображення МРТ пухлину головного мозку [10].

Зважаючи на те, що трансформери потребують досить великої кількості даних для якісного навчання, для всіх наявних даних міждоменого набору № 2 була проведена аугментація, для позитивного класу використано два різновиди комбінацій перетворень, загальна кількість екземплярів склала 102562 для негативного класу та 24450 для меланому. Для вилучення ознак, що супроводжується заморожуванням шарів базової моделі, використано повнозв'язні шари з 128 та 1 нейроном, регуляризація у вигляді Dropout шару, обрано швидкість навчання $2e-4$, розмір пакета – 32, модель навчалась протягом 5 епох. Fine-tuning проводився протягом 3-х епох з використанням learning rate schedule exponential decay з початковим значенням $3e-5$ зі зменшення швидкості навчання після кожної епохи на 5%, кількість параметрів, що навчалися, склала близько 86,5 млн. Результати моделі представлено в таблиці 2.

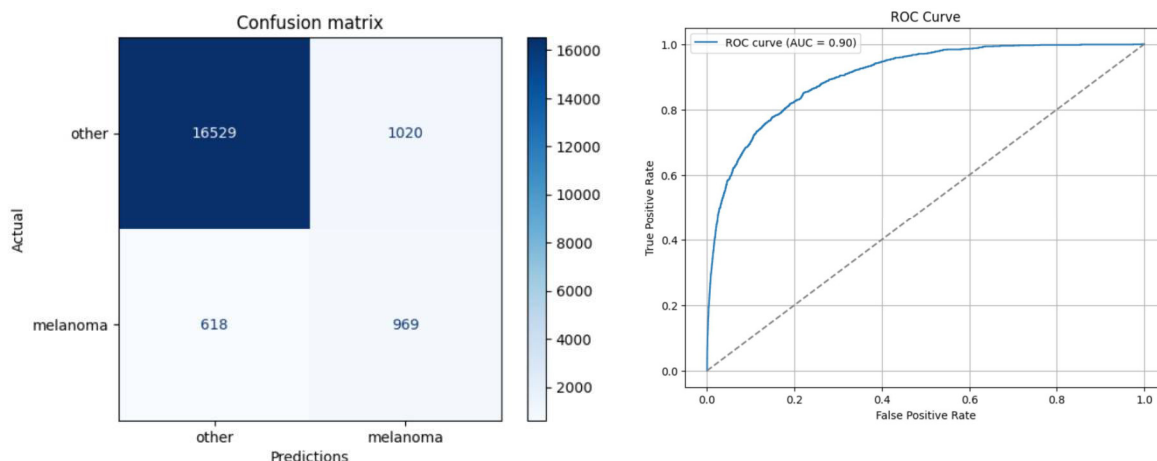


Рис.6. Матриця розбіжностей та ROC-AUC EfficientNetV2B0

Розробка моделі нейронної мережі з нуля. Базові моделі, використані для донавчання на зображеннях шкірних утворень, спочатку навчалися на наборі даних ImageNet, який не містить спеціалізованих медичних зображень. Через це їхня здатність аналізувати дерматологічні дані може бути обмеженою. Враховуючи цю особливість, доцільним рішенням може бути розробка моделі з нуля, яка буде оптимізована для роботи саме з медичними зображеннями, забезпечуючи більшу точність та адаптованість до специфіки даних.

З цією метою було розроблено власну згорткову модель для класифікації меланому, яка навчалася виключно на датасеті № 2 з використанням аугментованих даних. Архітектура поєднувала блоки ConvDepthConv із bottleneck-структурою, механізми уваги СВМ [7], skip-connections, dropout шари та L2-регуляризацію, використовувалась функція активації 'silu'. Попри застосування сучасних технік, результати моделі виявилися гіршими порівняно з більшістю випробуваних переднавчених мереж. Після видалення чорного контуру з зображень на тестових даних власна модель погіршила показники передбачень на відміну від інших моделей (табл. 2).

Отримані результати демонструють перевагу переднавчаних мереж для подібних задач.

Використання ансамблів моделей для отримання передбачень. Ансамблеві методи машинного навчання – це техніка, яка об'єднує кілька моделей для покращення точності передбачень. Основна ідея ансамблювання полягає в тому, що група слабких моделей може створити сильну модель, якщо їхні передбачення правильно комбінувати. Ансамблі допомагають зменшити перенавчання, підвищити стійкість до шуму та покращити загальну продуктивність моделі.

У роботі для створення ансамблю моделей застосовано техніку стекинг (stacking), в якій кілька моделей генерують передбачення, які потім використовуються як вхідні дані для метамоделі, що робить остаточне передбачення. В ансамблях використано різні поєднання отриманих раніше моделей, що навчалися на наборах даних № 1 та об'єднаному № 2. У рамках експериментів було досліджено продуктивність різних мета-класифікаторів для задачі виявлення меланоми на дерматоскопічних зображеннях. В якості вхідних ознак використовувались прогнози моделей EfficientNetV2B0, DenseNet121, Xception, ResNet50V2 та розробленої CNN. В якості метамоделі випробувана логістична регресія, XGBoost, GradientBoostingClassifier та багатошаровий перцептрон (MLP), які навчалися на передбаченнях базових моделей для валідаційного набору. Моделі оцінювались на повному тестовому наборі № 2, що містить зображення 2020 та 2019 років з видаленим чорним контуром для тих зображень, в яких він був виявлений (рис. 7).

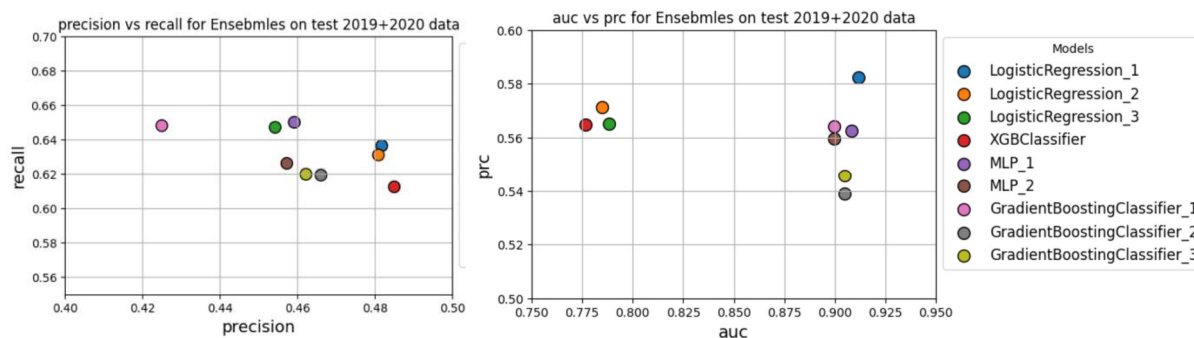


Рис.7. Порівняння ансамблевих моделей за precision, recall та AUC, AUC-PRC метриками

Найкращу дискримінативну здатність серед ансамблів продемонструвала мета-модель логістична регресія, яка поєднувала моделі тільки архітектури EfficientNetV2B0, що навчалися на повному наборі даних, за метриками $AUC = 0.9117$ та $AUC-PRC = 0.5823$ вона випереджає інші. Всі ансамблі підвищили повноту розпізнавання меланоми порівняно з окремими моделями, проте точність залишилась низькою. Прямої залежності від використання моделей з різних наборів даних не спостерігається, так само як і від підвищення кількості слабких моделей. Включення різних архітектур в ансамбль не завжди супроводжувалось підвищенням якості класифікації. Мета-модель MLP показала кращий результат з меншою кількістю найсильніших моделей та нейронів у повнозв'язних шарах, ніж при додаванні більш слабких моделей та збільшенні нейронів.

Загалом, результати демонструють більшу ефективність простих ансамблевих стратегій з невеликою кількістю більш потужних моделей, зокрема EfficientNetV2B0 у поєднанні з лінійними мета-класифікаторами або MLP. У той же час, використання різномірних джерел або складних ансамблів потребує додаткового аналізу та налаштування для уникнення втрати точності, також треба врахувати, що моделі, що

навчались на наборі № 1, мають низьку точність розпізнавання негативного класу на змішаних даних.

Порівняння продуктивності моделей на різних наборах даних. Для порівняння продуктивності моделей на різних наборах даних було обрано найкращі згорткові моделі з архітектурами EfficientNetV2B0, DenseNet121, Xception, а також розроблену власну CNN, натреновані на наборі № 2, та модель EfficientNetV2B0-04-12-10, що навчалась на наборі № 1. Оцінювання проводилось на тестових наборах 2019 та 2020 років, а також на змішаному наборі – до та після попередньої обробки зображень (обрізання чорних рамок).

Згідно з результатами (таблиці 3-4), якість класифікації на тестовому наборі 2020 року, де частка зображень меланоми становить лише 2,39%, була помітно нижчою порівняно з набором 2019 року, що містить більше прикладів позитивного класу. Водночас усі згорткові моделі продемонстрували вищі значення AUC на даних 2020 року, що пов'язано з особливостями розбалансованості класів – ця метрика менш чутлива до дисбалансу, ніж інші.

Таблиця №3

Оцінка на тестовому наборі 2020 року

Model	precision	recall	auc	prc	f1_beta
EfficientNetV2B0_04-29-19	0.3	0.4385	0.8854	0.3455	0.4014
EfficientNetV2B0_04-23-10	0.5128	0.2308	0.8855	0.2965	0.2593
Custom CNN_05-11-17	0.3149	0.2192	0.8377	0.198	0.2334
DenseNet121_05-05-17	0.35	0.2692	0.8903	0.2617	0.2823
Xception_05-20-16	0.4	0.2538	0.8866	0.2812	0.2739
EfficientNetV2B0_04-12-10_last	0.3454	0.2577	0.8288	0.2372	0.2715

Таблиця №4

Оцінка на тестовому наборі 2019 року

Model	до обрізки чорних контурів зображень					після обрізки чорних контурів зображень				
	precision	recall	auc	prc	f1_beta	precision	recall	auc	prc	f1_beta
EfficientNetV2B0_04-29-19	0.5457	0.5539	0.81	0.5733	0.5522	0.5314	0.6443	0.8629	0.6127	0.618
EfficientNetV2B0_04-23-10	0.5829	0.4371	0.8004	0.5449	0.4601	0.5682	0.5245	0.8455	0.5805	0.5327
Custom CNN_05-11-17	0.4554	0.52	0.7949	0.5153	0.5056	0.5043	0.4815	0.7836	0.5191	0.4859
DenseNet121_05-05-17	0.474	0.4333	0.7756	0.4714	0.4409	0.4584	0.5275	0.7978	0.4856	0.5121
Xception_05-20-16	0.4616	0.3444	0.6751	0.3965	0.3628	0.4464	0.4582	0.7465	0.4076	0.4558
EfficientNetV2B0_04-12-10_last	0.1867	0.9653	0.6245	0.2056	0.5263	0.186	0.9623	0.621	0.2041	0.5246

Попередня обробка зображень (видалення чорного контуру) позитивно вплинула на всі моделі – на даних 2019 року показники покращились, окрім власної CNN. Для набору 2020 року зміни були незначними, оскільки лише 24 зображення містили чорний контур. Після обрізання зменшився precision, однак метрики recall, AUPRC, AUC та f1 здебільшого зросли.

Загалом моделі демонструють вищу ефективність на збалансованішому наборі 2019 року. Утім, навчання на міждомених даних не дозволило досягти стабільно високої якості класифікації на окремих доменах – моделі не змогли ефективно узагальнити ознаки. Винятком стала модель EfficientNetV2B0-04-12-10, навчена лише на наборі №1, яка досягла високої повноти при дуже низькій точності (табл. 5). Це можна пояснити тим, що модель не бачила негативного класу з 2019 року, тому мала

значну кількість хибно позитивних прогнозів. Водночас, на наборі №1 (який містить позитивні приклади з 2019 року та негативні з 2020 року) вона досягла precision = 0.915 і recall = 0.849, що підкреслює суттєвий вплив доменної специфіки на класифікацію.

На змішаних даних значення AUPRC для всіх моделей було нижчим, ніж на наборі 2019 року, при цьому AUC залишалось вищим. Ці набори мають різні баланси класів, а AUPRC краще відображає якість моделі на незбалансованих класах, які характерні для практичних застосувань задачі виявлення меланому.

Таблиця №5

Порівняння результатів оцінки моделей, що навчались на міждоменному наборі даних 2019 та 2020 рр., після обробки зображень на змішаному тестовому наборі №2

Model	loss	accuracy	precision	recall	tp	fp	tn	fn	auc	prc	f1_beta
Ensemble	n/a	0.913000	0.481600	0.636400	891.0	938.0	16612.0	695.0	0.911700	0.582300	0.598000
ft-EfficientNetV2B0_04-23-10	0.19322115182876587	0.927258	0.571324	0.492124	781.0	586.0	16963.0	806.0	0.903627	0.553670	0.506157
tl-EfficientNetV2B0_04-29-09	0.2242589294910431	0.914402	0.487179	0.610586	969.0	1020.0	16529.0	618.0	0.898946	0.564072	0.581144
ft-DenseNet121_05-05-17	0.23544709384441376	0.907295	0.445860	0.485192	770.0	957.0	16592.0	817.0	0.875133	0.451088	0.476780
ViT-05-04-07	0.24359740316867828	0.925115	0.549296	0.540643	858.0	704.0	16845.0	729.0	0.872017	0.565482	0.542351
CNN_05-11-07	0.2394837588071823	0.906564	0.440708	0.470699	747.0	948.0	16601.0	840.0	0.861108	0.465147	0.464379
ft-Xception_05-20-16	0.2570240795612335	0.907713	0.441388	0.424701	674.0	853.0	16696.0	913.0	0.840367	0.380810	0.427937
EfficientNetV2B0_04-12-10_last	3.2422122955322266	0.688702	0.190422	0.846881	1344.0	5714.0	11835.0	243.0	0.802486	0.199254	0.501268

Зважаючи на прикладне значення задачі – раннє виявлення меланому – найбільш важливою метрикою класифікації виступає recall, що характеризує здатність моделі виявляти якомога більше позитивних випадків. При порівнянні якості всіх моделей за цією метрикою, а також за AUC та AUPRC, на міждоменному наборі даних найкращі результати продемонстрував ансамбль моделей на базі EfficientNetV2B0 (рис. 8), ViT має один з найвищих балансів між точністю та повнотою та площу під кривою AUPRC.

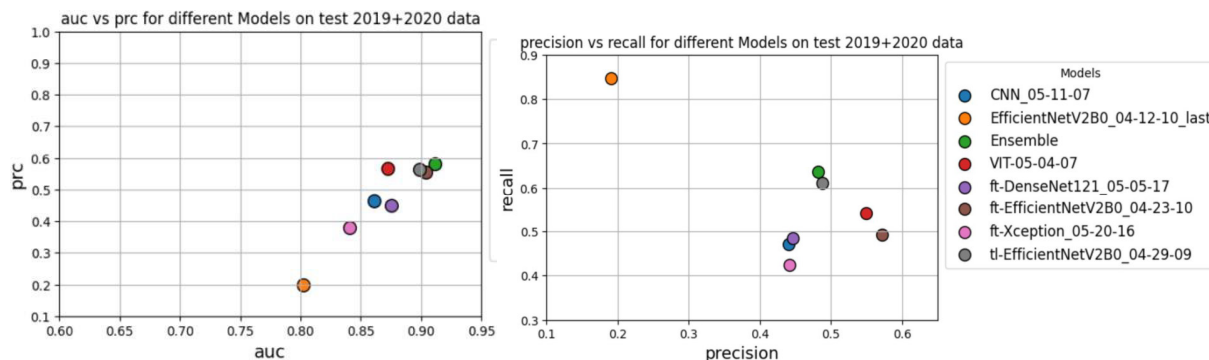


Рис. 8. Порівняння моделей за метриками AUC та AUPRC, precision та recall при тестуванні на міждоменному наборі даних 2019 та 2020 рр.

Отже, ансамблеві методи та архітектури EfficientNet і ViT є найбільш перспективними для класифікації меланому на міждоменних даних.

Висновки

У статті досліджено, що поєднання зображень із різних джерел у деяких випадках призводить не до покращення, а до погіршення результатів. Змішування міждоменних даних під час навчання моделі може спричинити систематичне зміщення класифікації, коли алгоритм починає асоціювати особливості зображень не з ознаками захворювання, а з характеристиками домену, з якого отримані дані. Тобто модель навчається розрізняти

джерела зображень замість фактичних медичних ознак патології, що негативно позначається на її узагальнюючій здатності та діагностичній точності. Попри збільшення загального обсягу даних, різниця у статистичних характеристиках між доменами (типи зйомки, умови освітлення, формат зображень, баланс класів) ускладнює навчання моделі й може знизити її здатність розрізняти класи в окремих доменах. Це проявляється у зниженні точності та повноти, навіть при високому значенні AUC, та свідчить про те, що модель не здатна ефективно розрізняти класи при реальних порогах, що є критичним для клінічного застосування.

З практичної точки зору, результати роботи вказують на важливість ретельної підготовки даних та необхідність розробки адаптивних стратегій для міждоменного навчання. Незважаючи на складності, побудова моделей, які здатні працювати на різномірних даних, залишається перспективним напрямом, адже саме такий підхід є найближчим до умов реального застосування в медичній практиці.

Список використаної літератури:

1. Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale / A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al. // *International Conference on Learning Representations*. - 2020. – Режим доступу: <https://arxiv.org/pdf/2010.11929>.
2. Alijani S. Vision transformers in domain adaptation and domain generalization: A study of robustness / J. Fayyad, H. Najjaran // *Neural Computing and Applications*. – Springer. – 2024. – Режим доступу: <https://doi.org/10.1007/s00521-024-10353-5>.
3. Flosdorf C. Skin Cancer Detection utilizing Deep Learning: Classification of Skin Lesion Images using a Vision Transformer / C. Flosdorf, J. Engelker, I. Keller, N. Mohr. – 2024. - [Електронний ресурс] – Режим доступу: <https://doi.org/10.48550/arXiv.2407.18554>.
4. Wu Y. Skin Cancer Classification With Deep Learning: A Systematic Review / Y. Wu, B. Chen, A. Zeng, D. Pan, R. Wang, S. Zhao // *Frontiers in Oncology*. – 2022. – Vol.12. - Режим доступу: <https://doi.org/10.3389/fonc.2022.893972>.
5. ISIC Challenge Datasets [Електронний ресурс] – Режим доступу: <https://doi.org/10.34970/2020-ds01>.
6. SIIM-ISIC 2020 Challenge Dataset // *International Skin Imaging Collaboration* [Електронний ресурс] – Режим доступу: <https://challenge.isic-archive.com/data/#2020>.
7. M. Saini. Tackling class imbalance in computer vision: a contemporary review / M. Saini, S. Susan // *Artificial Intelligence Review*. – 2023. – Vol. 56. – P. 1279–1335.
8. Keras Applications. – [Електронний ресурс] – Режим доступу: <https://keras.io/2/api/applications/>.
9. Hosny K. Skin Cancer Classification using Deep Learning and Transfer Learning / K. Hosny, M. Kassem, M. Fouad // *9th Cairo International Biomedical Engineering Conference (CIBEC)*. – 2018. – P.90-93.
10. Hugging Face. Brian Dunn / vit-base-patch16-224-in21k_brain_tumor_diagnosis [Електронний ресурс] – Режим доступу: https://huggingface.co/DunnBC22/vit-base-patch16-224-in21k_brain_tumor_diagnosis.
11. Woo S. CBAM: Convolutional Block Attention Module / S. Woo, J. Park, J.-Y. Lee, I. Kweon // *European Conference on Computer Vision*. – 2018. – P. 3-19.

References:

1. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*. <https://arxiv.org/pdf/2010.11929>.
2. Alijani, S., Fayyad, J., & Najjaran, H. (2024). *Vision transformers in domain adaptation and domain generalization: A study of robustness*. *Neural Computing and Applications*. Springer. <https://doi.org/10.1007/s00521-024-10353-5>.
3. Flosdorf, C., Engelker, J., Keller, I., & Mohr, N. (2024). Skin cancer detection utilizing deep learning: Classification of skin lesion images using a vision transformer. *arXiv*. <https://doi.org/10.48550/arXiv.2407.18554>.
4. Wu, Y., Chen, B., Zeng, A., Pan, D., Wang, R., & Zhao, S. (2022). Skin cancer classification with deep learning: A systematic review. *Frontiers in Oncology*, 12. <https://doi.org/10.3389/fonc.2022.893972>.
5. ISIC Challenge Datasets. (2020). *International Skin Imaging Collaboration*. <https://doi.org/10.34970/2020-ds01>.

6. International Skin Imaging Collaboration. (2020). *SIIM-ISIC 2020 challenge dataset*. <https://challenge.isic-archive.com/data/#2020>.
7. Saini, M., & Susan, S. (2023). Tackling class imbalance in computer vision: A contemporary review. *Artificial Intelligence Review*, 56, 1279–1335. <https://doi.org/10.1007/s10462-023-10557-6>.
8. Keras Applications. <https://keras.io/2/api/applications/>.
9. Hosny, K., Kassem, M., & Fouad, M. (2018). Skin cancer classification using deep learning and transfer learning. In *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*, 90–93. <https://doi.org/10.1109/CIBEC.2018.8641762>.
10. Hugging Face. (n.d.). *Brian Dunn: vit-base-patch16-224-in21k_brain_tumor_diagnosis*. https://huggingface.co/DunnBC22/vit-base-patch16-224-in21k_brain_tumor_diagnosis.
11. Woo, S., Park, J., Lee, J.-Y., & Kweon, I. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.

YURTSUN Svitlana,

Student, Department of Applied Mathematics and Informatics, The Bohdan Khmelnytsky National University of Cherkasy, Ukraine

BOGATYREV Oleksandr,

Candidate of Physical and Mathematical Sciences, Associate Professor, Department of Applied Mathematics and Informatics, The Bohdan Khmelnytsky National University of Cherkasy, Ukraine

ROBUSTNESS OF DEEP LEARNING MODELS TO DOMAIN SHIFT IN AUTOMATIC MELANOMA DETECTION ON DERMATOSCOPIC IMAGES

Summary. Introduction. *The article addresses the problem of automated melanoma detection in dermoscopic images using deep learning methods under conditions of limited data availability, which may contribute to the early diagnosis of this life-threatening disease. The main objective is to explore ways of improving the generalization ability of deep learning models in classifying data from different sources. The study utilised two sets of dermoscopic images from the ISIC Archive in various combinations. A series of experiments was conducted applying transfer learning with CNNs and the Vision Transformer (ViT) architecture, a custom convolutional model, and model ensembles. Particular attention is given to the impact of image preprocessing (removal of black borders, augmentations) on classification quality, the use of data from other sources only for the minority class and for all classes, as well as the generalization capability of models on cross-domain data. The results demonstrated the advantage of transfer learning, while the custom model showed lower accuracy. Among convolutional neural networks, the EfficientNet architecture achieved the highest classification performance, while Vision Transformer models demonstrated strong potential, outperforming most CNN architectures. The study revealed that combining data from different sources may not improve, but rather decrease classification performance due to the model learning domain differences instead of medical features. This effect was especially evident when supplementing only the minority class from external datasets to address class imbalance during training. By removing noise in the form of black borders characteristic of one dataset, the accuracy of disease recognition was improved. The results of the study indicate the necessity of thorough data preparation and the application of specialized strategies for cross-domain training to ensure practical applicability.*

Purpose. *The aim of the article is to investigate the impact of combining cross-domain data and image preprocessing on the quality of melanoma classification using deep learning methods.*

Results. *The study reveals that combining images from different sources often leads not to improvement but to a decline in performance. Mixing cross-domain data during model training can introduce systematic classification bias, where the algorithm begins to associate image features not with disease characteristics but with the domain from which the data originate. As a result, the model learns to distinguish between data sources rather than actual medical indicators of pathology, which negatively affects its generalization ability and diagnostic accuracy. Despite the increased overall volume of data, differences in statistical characteristics between domains (such as imaging types, lighting conditions, image formats, and class balance) complicate model training and may reduce its ability to discriminate classes within individual domains. This is reflected in a decline in accuracy and recall, even when AUC value remains high, indicating that the model cannot effectively separate classes at realistic thresholds—a critical limitation for clinical use.*

Conclusion. From a practical perspective, the results of this work highlight the importance of thorough data preparation and the development of adaptive strategies for cross-domain training. Despite these challenges, developing models capable of reliably operating on heterogeneous data remains a promising direction, as this approach most closely reflects the conditions of real-world medical practice.

Keywords: deep learning, transfer learning, computer vision, convolutional neural networks, automated melanoma detection, data domain, class imbalance, medical image classification.

Одержано редакцією 22.07.2025 р.
Прийнято до публікації 24.09.2025 р.

УДК 004.75:004.052.42

DOI 10.31651/2076-5886-2025-1-18-32

PACS 07.05.Bx, 07.05.Mh

КИРПА Григорій Анатолійович
студент спеціальності «Прикладна
математика» Черкаського національного
університету імені Богдана
Хмельницького

ДЗЮБА Вікторія Анатоліївна
кандидат технічних наук, викладач
кафедри прикладної математики та
інформатики Черкаського національного
університету імені Богдана
Хмельницького
e-mail: viktoriya.dzyuba15@gmail.com
ORCID 0000-0003-1655-0333

ГЛАДКА Людмила Іванівна
к.ф.-м.н., доцент кафедри автоматизації та
комп'ютерно-інтегрованих технологій,
Черкаський національний університет
імені Б. Хмельницького
e-mail: l_i_gladka@vu.cdu.edu.ua
ORCID 0000-0002-7030-9666

ПОРІВНЯЛЬНИЙ АНАЛІЗ АЛГОРИТМІВ ПЛАНУВАННЯ РЕСУРСІВ У ХМАРНИХ ОБЧИСЛЮВАЛЬНИХ СЕРЕДОВИЩАХ

У статті досліджуються алгоритми планування обчислювальних ресурсів у хмарних середовищах. Проведено порівняльний аналіз класичних та евристичних методів розподілу завдань: *First Come First Served (FCFS)*, *Round Robin (RR)*, *Longest Processing Time (LPT)*, *Shortest Processing Time (SPT)* та *Weighted Round Robin (WRR)*. Задача розподілу ресурсів формалізована як задача мінімізації показника *Makespan* у гетерогенному обчислювальному середовищі. Для проведення експериментів розроблено модульний програмний комплекс на мові *Python*, який реалізує імітаційне моделювання хмарних навантажень при трьох рівнях інтенсивності вхідного потоку: низькому (10 завдань), середньому (20 завдань) та високому (40 завдань). Оцінювання якості алгоритмів здійснюється за допомогою системи метрик, що включає *Makespan*, стандартне відхилення завантаженості вузлів, коефіцієнт варіації (*CV*) та інтегральний показник ефективності *Score*, який поєднує якість розкладу і обчислювальні витрати на планування. Встановлено, що алгоритм *LPT* демонструє найкращі показники *Makespan* за умов середнього та високого навантаження, тоді як *WRR* забезпечує найвищі значення інтегрального показника в сценаріях з гетерогенними вузлами. Результати підтверджують відсутність універсального оптимального алгоритму та обґрунтовують