

<https://doi.org/10.31891/2219-9365-2026-85-38>

УДК 004.83

МЕЩЕРЯКОВ Олександр

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

<https://orcid.org/0009-0002-6485-9722>

e-mail: [ascellanova@gmail.com](mailto:ascellanova@gmail.com)

## МЕТОДИ БОРОТЬБИ З BIAS ТА DOMAIN SHIFT У МЕДИЧНИХ НЕЙРОННИХ МЕРЕЖАХ

У статті розглянуто проблему узагальнення медичних нейронних мереж в умовах domain shift, що виникає через гетерогенність апаратних засобів, протоколів збору даних і клінічних практик, та показано, що деградація якості моделей після деплою має системний характер і не зводиться ані до недостатньої складності архітектур, ані до обмеженості навчальних вибірок, а обумовлена зсувами розподілів між тренувальним і цільовим доменами; особливу увагу приділено аналізу архітектурних припущень стандартних компонентів нейронних мереж, зокрема механізмів нормалізації, в межах якого проведено формалізований аналіз Batch Normalization як механізму імпліцитного кодування статистики тренувального домену та показано, що використання running statistics формує внутрішні представлення, оптимальні лише за умови стаціонарності розподілу даних, роблячи моделі чутливими до covariate shift у реальних клінічних сценаріях; на цій основі розглянуто альтернативні підходи до нормалізації, зокрема Instance Normalization і Group Normalization, та проаналізовано їхню придатність для мультидоменних медичних даних, а також детально досліджено інженерні аспекти побудови й оцінки медичних моделей, обґрунтовано обмеженість стандартної internal validation і необхідність застосування протоколів оцінювання, що відтворюють реальні сценарії деплою, зокрема leave-one-hospital-out validation, на підставі чого сформульовано практичні рекомендації щодо вибору нормалізаційних механізмів, архітектурних рішень і протоколів валідації з метою підвищення стабільності та надійності медичних систем штучного інтелекту в продакшені.

Ключові слова: медичні нейронні мережі; domain shift; bias у даних; Batch Normalization; Group Normalization; domain generalization; медичні зображення; клінічний деплой; валідація моделей

MESHCHERIAKOV Oleksandr

National Technical University of Ukraine "Ihor Sikorskyi Kyiv Polytechnic Institute"

## METHODS FOR COMBATING BIAS AND DOMAIN SHIFT IN MEDICAL NEURAL NETWORKS

The article addresses the problem of generalization of medical neural networks under domain shift caused by heterogeneity in imaging hardware, acquisition protocols, and clinical practices. It is argued that the degradation of model performance after deployment is systemic and cannot be explained solely by insufficient architectural complexity or limited training data, but rather by distributional discrepancies between the source and target domains. The study focuses on the architectural assumptions embedded in standard neural network components, with particular emphasis on normalization mechanisms. A formal analysis of Batch Normalization is provided, demonstrating that its running statistics implicitly encode properties of the training domain and produce internal feature representations that remain optimal only under the assumption of distributional stationarity. Consequently, models relying on Batch Normalization become highly sensitive to covariate shift in real clinical environments where data distributions vary across institutions.

Based on this analysis, alternative normalization strategies are examined, including Instance Normalization and Group Normalization, which do not depend on global batch statistics and therefore exhibit greater robustness in multi-domain medical settings. The paper also investigates engineering aspects of model development and evaluation, highlighting the limitations of conventional internal validation procedures that fail to capture cross-site variability. To address this, the adoption of evaluation protocols that approximate real deployment conditions, particularly leave-one-hospital-out validation, is justified as a more reliable indicator of clinical performance. On this basis, practical recommendations are formulated regarding the selection of normalization mechanisms, architectural design choices, and validation methodologies aimed at improving the domain generalization and operational reliability of medical deep learning systems.

Keywords: medical neural networks; domain shift; data bias; Batch Normalization; Group Normalization; domain generalization; medical images; clinical deployment; model validation.

Стаття надійшла до редакції / Received 05.01.2026

Прийнята до друку / Accepted 17.02.2026

Опубліковано / Published 05.03.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© Мещеряков Олександр

### ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Використання нейронних мереж у задачах медичної діагностики продемонструвало значний потенціал для автоматизації аналізу зображень, фізіологічних сигналів та клінічних даних. У численних дослідженнях моделі досягають високих показників точності на тестових вибірках, що створює передумови для їх впровадження у клінічну практику. Проте у реальних умовах експлуатації такі системи часто демонструють нестабільну поведінку та суттєву деградацію якості після деплою.

Ключовою проблемою є розбіжність статистичних властивостей даних між середовищем навчання та фактичним клінічним використанням. Медичні дані формуються в умовах значної гетерогенності апаратури,

протоколів збору, клінічних практик і популяцій пацієнтів. У результаті моделі, натреновані на даних одного центру або одного типу обладнання, виявляються чутливими до domain shift і не здатні стабільно узагальнювати на нові джерела даних.

На відміну від багатьох некритичних прикладних задач, у медицині помилки узагальнення мають безпосередні клінічні наслідки. Це робить проблему bias та domain shift не лише теоретичною, а й практично значущою з точки зору безпеки, регуляторного схвалення та масштабування систем штучного інтелекту в охороні здоров'я.

### АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Сучасні дослідження у сфері медичного машинного навчання зосереджені переважно на підвищенні точності моделей за рахунок архітектурних удосконалень, збільшення обсягу даних та використання методів попередньої обробки. Значна увага приділяється покращенню якості вхідних даних, зокрема шляхом нормалізації, фільтрації шуму, покращення контрасту та злиття ознак.

Окремі роботи аналізують ефективність ансамблевих та fusion-підходів, демонструючи зростання метрик на стандартних benchmark-наборах. Інші дослідження зосереджуються на боротьбі з аномаліями в медичних зображеннях, наприклад у задачах протомного аналізу або діагностики інфекційних захворювань.

Водночас більшість існуючих підходів оцінюються в умовах internal validation, де тренувальні та тестові вибірки походять з одного джерела або мають подібні статистичні властивості. Питання стійкості моделей до domain shift, а також вплив архітектурних компонентів зокрема нормалізаційних механізмів на узагальнювальну здатність часто залишаються поза фокусом аналізу або розглядаються фрагментарно.

### ВИДІЛЕННЯ НЕВИРІШЕНИХ РАНІШЕ ЧАСТИН ЗАГАЛЬНОЇ ПРОБЛЕМИ

Незважаючи на значну кількість робіт у галузі медичної класифікації, низка критичних аспектів залишається недостатньо дослідженою. Зокрема, відсутній системний аналіз того, як архітектурні припущення нейронних мереж взаємодіють з domain shift у реальних клінічних сценаріях.

Більшість моделей проектується з використанням стандартних компонентів, таких як Batch Normalization, без явного врахування того, що ці механізми можуть кодувати доменно-специфічні статистики. У результаті стабільність моделей оцінюється переважно в межах тренувального домену, тоді як їх поведінка при зміні джерела даних залишається непередбачуваною.

Крім того, у літературі бракує чітких інженерних рекомендацій щодо вибору нормалізаційних підходів і протоколів валідації для медичних систем, орієнтованих на реальний продакшен. Це ускладнює трансляцію наукових результатів у практичні клінічні рішення.

### ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою даної роботи є системний аналіз проблеми bias та domain shift у медичних нейронних мережах із фокусом на ролі нормалізаційних механізмів у формуванні узагальнювальної здатності моделей. Особливу увагу приділено Batch Normalization як архітектурному компоненту, що імпліцитно кодує статистику тренувального домену та може підсилувати крихкість моделей при деплої. У межах статті вирішуються такі завдання: формалізувати основні типи bias і domain shift, характерні для медичних даних та проаналізувати вплив Batch Normalization на поведінку моделей під час domain shift;

Отримані результати спрямовані на підвищення надійності та безпеки медичних систем штучного інтелекту в реальних клінічних умовах.

### ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

За останні роки нейронні мережі досягли вражаючих результатів у медичних задачах, демонструючи показники, що часто не поступаються або перевищують людський рівень на ретроспективних тестових вибірках. Однак у більшості практичних сценаріїв ці результати виявляються нестабільними після переходу від дослідницького середовища до реального клінічного використання. Типовою є ситуація, коли модель, натренована на даних однієї лікарні або одного типу обладнання, зазнає суттєвої деградації якості при застосуванні до даних з іншої установи. Ця проблема спостерігається в широкому спектрі медичних застосувань: від аналізу зображень (MRI, CT, рентгенографія) до обробки фізіологічних сигналів (ECG, EEG) та структурованих електронних медичних записів.[1,2] Важливо зазначити, що мова йде не про поодинокі випадки або погано зібрані датасети — навпаки, подібна поведінка моделей є системною і регулярно фіксується навіть у великих мультицентрових дослідженнях.

Ключовою причиною такої деградації є невідповідність статистичних властивостей даних між тренувальним і цільовим доменами. На відміну від багатьох некритичних прикладних задач, у медицині поняття «домен» не зводиться лише до формального визначення задачі. Воно включає комплекс взаємопов'язаних факторів: тип і налаштування апаратури, протоколи збору даних, популяційні характеристики пацієнтів, локальні клінічні практики, а також спосіб формування і верифікації міток.

Класичні підходи машинного навчання зазвичай ґрунтуються на припущенні, що тренувальні та тестові дані походять з одного і того ж розподілу. У медичних задачах це припущення майже завжди порушується. Більш того, саме медичні дані характеризуються особливою чутливістю до навіть незначних змін у розподілах, оскільки моделі часто оперують тонкими статистичними патернами, які не мають прямої фізіологічної інтерпретації.

На відміну від споживчих або маркетингових застосувань, помилки узагальнення в медицині мають безпосередні клінічні наслідки. Нестабільність поведінки моделі між доменами підриває довіру з боку лікарів, ускладнює процес регуляторної сертифікації та робить неможливим масштабування рішень на рівні систем охорони здоров'я. У цьому контексті робота з bias та domain shift перестає бути опціональним етапом оптимізації і має розглядатися як фундаментальний аспект проектування медичних нейронних мереж.

Зсуви розподілів у медичних даних є однією з ключових причин деградації якості нейронних мереж при переході від експериментального середовища до реального клінічного використання. На відміну від багатьох інших прикладних галузей, у медицині доменні відмінності виникають не лише через різні джерела даних, але й через складну взаємодію технологічних, організаційних та людських факторів. У результаті дані, що використовуються для навчання моделі, часто суттєво відрізняються від даних, з якими модель стикається під час деплою.

Для коректного аналізу проблеми domain shift та вибору ефективних стратегій його компенсації необхідно чітко розрізнити основні типи bias, які можуть виникати на різних етапах життєвого циклу даних. У цьому розділі розглядаються найбільш поширені та критичні для медичних застосувань форми зсувів розподілів, а також їхній вплив на поведінку нейронних мереж.

Covariate shift є одним із найбільш поширених типів domain shift у медичному машинному навчанні. Він виникає у ситуаціях, коли статистичні властивості вхідних даних змінюються між тренувальним і цільовим доменами, тоді як фундаментальні залежності між станом пацієнта та клінічним результатом залишаються незмінними або змінюються незначно. У медичному контексті це означає, що фізіологічні закономірності зберігаються, однак спосіб їх представлення в даних зазнає змін.

Основними джерелами covariate shift у медичних даних є використання різних виробників і моделей медичного обладнання, відмінності в параметрах сканування, а також варіації в preprocessing pipeline. Навіть у межах однієї модальності, наприклад магнітно-резонансної томографії, зміни товщини зрізів, reconstruction kernels або контрастних налаштувань можуть призводити до суттєвих відмінностей у статистичних властивостях зображень.

Особливу складність становить той факт, що нейронні мережі здатні експлуатувати тонкі статистичні відмінності у даних, які не мають прямої клінічної інтерпретації. У присутності covariate shift модель може навчитися використовувати нефізіологічні ознаки — наприклад, характерний рівень шуму або специфічні текстурні патерни — як проксі для прогнозування. Це призводить до штучно завищених результатів у межах тренувального домену та різкої втрати узагальнювальної здатності при застосуванні до даних з іншого джерела.

На відміну від багатьох інших задач машинного навчання, у медицині мітки рідко є об'єктивним відображенням істинного стану системи. Вони формуються в результаті клінічної інтерпретації, яка залежить від досвіду спеціаліста, локальних діагностичних протоколів та контексту, в якому здійснюється анотація. Це призводить до виникнення різних форм label bias, які можуть суттєво впливати на поведінку моделей.

Однією з найбільш поширених форм є inter-observer variability, коли різні лікарі по-різному трактують одні й ті самі клінічні випадки. Навіть за наявності формалізованих критеріїв діагностики, суб'єктивні фактори та рівень досвіду анотатора можуть призводити до систематичних розбіжностей у мітках. Додатково, protocol-dependent labeling виникає внаслідок використання різних діагностичних стандартів або порогових значень у різних медичних установах.

Ще одним важливим аспектом є temporal bias, пов'язаний зі змінами клінічних рекомендацій або стандартів лікування з часом. Дані, зібрані в різні періоди, можуть відображати різні уявлення про одну й ту саму патологію, що створює додатковий рівень невизначеності.

У результаті модель навчається не лише прогнозувати клінічний результат, але й відтворювати специфічний стиль анотації, притаманний певному центру або часовому проміжку. При перенесенні на інший домен це може проявлятися у вигляді систематичних помилок, зокрема для окремих підгруп пацієнтів, що робить таку модель небезпечною для практичного використання без додаткової калібровки.

Selection bias виникає тоді, коли процес формування датасету не є репрезентативним щодо цільової популяції, на яку планується застосування моделі. У медичних задачах це явище є особливо поширеним через особливості організації клінічних процесів та доступності даних.

Одним із типових прикладів є referral bias, коли в датасет потрапляють переважно складні або нетипові випадки, що направляються до спеціалізованих центрів. Крім того, демографічні перекося за віком, статтю або етнічною належністю можуть виникати внаслідок географічних або соціально-економічних факторів. Спеціалізація медичних закладів, таких як онкологічні або кардіологічні центри, також призводить до формування вибірок, які не відображають загальну клінічну популяцію.

Модель, натренована на такому датасеті, може демонструвати високі показники на внутрішніх тестових вибірках, але виявляється непридатною для застосування в ширшому клінічному контексті. Це створює ілюзію надійності та ускладнює виявлення потенційних ризиків до моменту реального деплою.

Окрему категорію проблем становлять спуріозні кореляції, характерні для медичних даних. У таких випадках модель використовує ознаки, які статистично корелюють з цільовою змінною в межах конкретного домену, але не мають причинно-наслідкового зв'язку з клінічним станом пацієнта.

Типовими прикладами таких ознак є маркери апарата або медичного закладу, специфічні артефакти зображень, а також особливості формату або структури даних. За наявності стабільних кореляцій у тренувальному датасеті модель неминуче починає використовувати ці сигнали для прийняття рішень, оскільки вони знижують емпіричний ризик.

Проблема полягає в тому, що подібні кореляції часто залишаються непоміченими при стандартній валідації, особливо якщо train- та validation-вибірки походять з одного джерела.[1] Лише при переході до нового домену, де ці сигнали відсутні або мають інше значення, виявляється крихіткість моделі та різка деградація її якості.

Одним із ключових, але часто недооцінених факторів, що підсилюють вплив domain shift у медичних нейронних мережах, є використання Batch Normalization (BN).[4] Хоча цей механізм традиційно розглядається як технічний засіб стабілізації та прискорення навчання, у контексті мультидомених медичних даних він відіграє значно глибшу роль. Зокрема, Batch Normalization імпліцитно кодує статистичні властивості тренувального домену, формуючи внутрішні представлення, оптимальні лише за умови стаціонарності розподілу вхідних даних.[6]

У поєднанні з формами covariate shift, selection bias та спуріозних кореляцій, описаними в попередньому розділі, BN може виступати не просто як пасивний компонент архітектури, а як активний механізм

Під час тренування нейронної мережі механізм Batch Normalization накопичує статистичні оцінки активацій, які є згладженими усередненнями значень, спостережуваних у тренувальних батчах. За умови достатньо великої та репрезентативної тренувальної вибірки ці оцінки поступово наближаються до типових статистичних характеристик активацій, зумовлених тренувальним доменом. Надалі саме ці накопичені статистики використовуються під час інференсу як фіксована основа для нормалізації вхідних сигналів. За умови достатньо великої та репрезентативної тренувальної вибірки ці величини асимптотично наближаються до моментів розподілу активацій, згенерованих тренувальними даними: Таким чином, під час inference Batch Normalization реалізує детерміноване перетворення, жорстко прив'язане до статистики тренувального домену. Це неявно передбачає, що дані, з якими модель стикається після деплою, походять з того ж самого розподілу. Як було показано в розділі 2, у медичних застосуваннях це припущення систематично порушується.

У сценарії, коли після деплою модель стикається з даними, статистичні характеристики яких відрізняються від тренувальних, нормалізація, заснована на зафіксованих статистиках, стає систематично зміщеною. У таких умовах активації зсуваються відносно тих значень, на які була оптимізована модель під час навчання. Цей зсув не компенсується навчуваними параметрами нормалізаційного шару та має тенденцію накопичуватися у глибоких шарах мережі, що призводить до спотворення геометрії латентного простору та деградації фінальних передбачень.

Для медичних моделей цей ефект є особливо критичним. Дані часто характеризуються низьким співвідношенням сигналу до шуму та високою залежністю від апаратних і preprocessing-параметрів. У таких умовах навіть незначні зсуви статистик можуть призводити до непропорційно великих змін у внутрішніх представленнях і, як наслідок, до деградації фінальних передбачень.

З точки зору навчання представлень, Batch Normalization є невід'ємною частиною обчислювального графа моделі, а не лише допоміжною операцією. Параметри нейронної мережі оптимізуються у просторі, де всі внутрішні активації вже нормалізовані відповідно до статистики тренувального домену. У результаті внутрішні представлення стають умовно оптимальними лише за умови збереження цих статистичних властивостей. Хоча модель явно не навчається розпізнавати домен, її параметри та нормалізаційні шари імпліцитно кодують доменно-специфічну інформацію, що дозволяє трактувати Batch Normalization як механізм прихованого memorization домену.

У такій постановці параметри моделі оптимізуються не лише відносно функції втрат, але й відносно конкретної нормалізації. У результаті внутрішні представлення мережі стають умовно оптимальними лише за збереження статистичних властивостей тренувального домену. Хоча модель явно не навчається розпізнавати домен, її параметри та нормалізаційні шари кодують доменно-специфічну інформацію, що дозволяє інтерпретувати Batch Normalization як механізм імпліцитного domain memorization.

Типові протоколи оцінки моделей у машинному навчанні ґрунтуються на випадковому розбитті даних на train та validation вибірки. За таких умов зазвичай виконується і Batch Normalization використовує ті самі running statistics як під час навчання, так і під час оцінки.

У результаті ефект domain memorization залишається прихованим: covariate shift відсутній, а модель демонструє стабільні та відтворювані метрики. Лише при переході до справді незалежного домену —

наприклад, іншої лікарні або іншого типу обладнання — порушується припущення стаціонарності, і крихкість моделі стає очевидною

### Стратегії роботи з нормалізацією під domain shift у медичних нейронних мережах

Як було показано в попередньому розділі, Batch Normalization імпліцитно кодує статистику тренувального домену та формує внутрішні представлення, чутливі до зсувів розподілу вхідних даних. У медичних застосуваннях, де припущення стаціонарності систематично порушується, це призводить до деградації узагальнювальної здатності моделей після деплою. Відповідно, ефективна робота з domain shift вимагає явного перегляду ролі нормалізаційних механізмів у архітектурі нейронної мережі.

У цьому розділі розглядаються основні стратегії модифікації або заміни Batch Normalization, які застосовуються для зменшення чутливості моделей до доменних зсувів. Особлива увага приділяється їхнім властивостям з точки зору domain sensitivity, стабільності навчання та придатності для медичних даних.

Попри описані обмеження, Batch Normalization залишається поширеним вибором у медичних нейронних мережах. Це пояснюється її здатністю стабілізувати оптимізацію, пришвидшувати збіжність та зменшувати залежність навчання від ініціалізації параметрів. У межах одного домену BN часто демонструє високу стабільність і відтворюваність результатів.

Однак ці переваги досягаються за рахунок сильного архітектурного припущення про стаціонарність розподілу даних. Як наслідок, Batch Normalization характеризується високою domain sensitivity: навіть помірний covariate shift може призводити до систематичного зсуву активацій і деградації якості. У медичних сценаріях, де джерела даних майже завжди гетерогенні, це робить BN малопридатною без додаткових механізмів компенсації.

Instance Normalization ґрунтується на нормалізації активацій окремо для кожного зразка, без використання статистик батчу або накопичених усереднень. Завдяки цьому нормалізація стає інваріантною до глобальних змін статистики даних і значно менш чутливою до доменних зсувів. У задачах медичного аналізу зображень це часто призводить до стабільнішої поведінки моделей при зміні апаратури або параметрів сканування.[8] Водночас усунення глобальної статистичної інформації може призводити до втрати клінічно значущих сигналів, пов'язаних з абсолютними інтенсивностями або контрастом.

Завдяки цьому IN є інваріантною до batch-level та domain-level статистик, що суттєво знижує її чутливість до covariate shift. У контексті медичних зображень це часто призводить до більш стабільної поведінки моделей при зміні апаратури або параметрів сканування. Водночас Instance Normalization має і суттєві обмеження. Усунення глобальної статистики призводить до втрати інформації про абсолютні інтенсивності та контраст, які можуть бути клінічно релевантними. Тому IN зазвичай демонструє середню стабільність і є найбільш придатною для задач, де форма та локальні патерни важливіші за глобальні інтенсивнісні характеристики, наприклад у сегментації або структурному аналізі зображень.

Group Normalization (GN) була запропонована як компроміс між Batch Normalization та Instance Normalization. Group Normalization реалізує компромісний підхід, при якому канали активацій розподіляються на групи, а нормалізація виконується незалежно в межах кожної з них. На відміну від Batch Normalization, цей механізм не залежить від розміру батчу і не використовує накопичені статистики, що знижує його чутливість до domain shift.[5] Водночас, на відміну від Instance Normalization, Group Normalization зберігає частину глобальної статистичної інформації, що дозволяє уникнути надмірної втрати корисних ознак.

З практичної точки зору GN характеризується низькою domain sensitivity, високою стабільністю навчання та доброю придатністю для медичних задач, особливо за умов малих batch size та мультидоменних датасетів. Саме тому Group Normalization часто розглядається як дефолтна альтернатива Batch Normalization у сучасних медичних моделях комп'ютерного зору.

Узагальнюючи властивості розглянутих методів, можна виділити такі характеристики:

- Batch Normalization: висока стабільність у межах одного домену, але висока чутливість до domain shift і низька придатність для медичних даних без додаткових механізмів адаптації.
- Instance Normalization: низька domain sensitivity і добра інваріантність, але обмежена стабільність та можливі втрати клінічно релевантної інформації.
- Group Normalization: низька domain sensitivity, висока стабільність і найкращий компроміс між інваріантністю та збереженням корисних статистичних властивостей.

Ці відмінності мають принципове значення при виборі архітектурних компонентів для медичних нейронних мереж, оскільки вони безпосередньо впливають на узагальнювальну здатність моделей у реальних клінічних умовах.

У практичних системах часто виявляється, що жоден з розглянутих підходів не є оптимальним у чистому вигляді. Це призводить до використання комбінованих або гібридних стратегій нормалізації, які дозволяють розподілити domain-invariant та domain-adaptive компоненти по різних рівнях мережі.

Поширеними прикладами таких підходів є:

- використання Batch Normalization у ранніх шарах мережі, де доменні зсуви можуть бути менш критичними, та Group Normalization у глибоких шарах, відповідальних за семантичні представлення;

- поєднання Domain-Specific Batch Normalization з adversarial domain training для явного розділення доменних і клінічно релевантних ознак;
- застосування Adaptive Batch Normalization лише на окремих рівнях мережі з обмеженим впливом на глобальну геометрію латентного простору.

Domain shift у медичних нейронних мережах має системний характер і значною мірою підсилюється архітектурними припущеннями, закладеними в механізмах нормалізації. Зокрема, Batch Normalization імпліцитно кодує статистику тренувального домену, що робить внутрішні представлення моделей чутливими до змін джерела даних.[3,7] У реальних клінічних умовах, де гетерогенність апаратури, протоколів і популяцій є нормою, це призводить до крихкості моделей, яка часто залишається непоміченою на етапі розробки. Відповідно, робота з domain shift повинна розглядатися не як опціональне покращення, а як фундаментальний аспект інженерного дизайну медичних систем штучного інтелекту.

Попри ці обмеження, Batch Normalization не є априорі непридатною для медичних застосувань. Її використання може бути виправданим у строго контрольованих сценаріях, де домен деплою чітко визначений і практично збігається з тренувальним доменом. До таких випадків належать системи, призначені для використання в одному медичному центрі з фіксованою апаратурою, стабільними протоколами збору даних і незмінним preprocessing pipeline. У подібних умовах BN забезпечує високу стабільність оптимізації та ефективну роботу з великими batch size. Водночас навіть у цих сценаріях модель слід розглядати як жорстко прив'язану до конкретного домену, а будь-яке розширення області застосування без повторного аналізу доменних відмінностей створює суттєві ризики.

Таблиця 1.

**Порівняння нормалізаційних підходів у нейронних мережах з медичними даними**

Метод	Domain sensitivity	Стабільність навчання	Придатність для медичних даних
Batch Normalization	Висока	Висока	Низька при domain shift
Instance Normalization	Низька	Середня	Висока для imaging
Group Normalization	Низька	Висока	Найкращий компроміс

У більшості практичних медичних застосувань припущення стаціонарності розподілу даних не виконується, що робить використання Batch Normalization небажаним. Мультицентрові датасети, малі batch size, гетерогенність апаратури та поступовий domain drift є типовими, а не винятковими умовами. У таких сценаріях Batch Normalization підсилює залежність моделей від доменно-специфічних статистик і може призводити до нестабільної поведінки після деплою, навіть якщо internal validation демонструє високі метрики. На цьому тлі Group Normalization постає як найбільш збалансоване рішення: вона не використовує batch-level статистики, зберігає стабільність навчання та демонструє низьку чутливість до domain shift. Саме тому GN доцільно розглядати як дефолтний вибір для медичних моделей, якщо немає переконливих причин використовувати Batch Normalization.

Окремої уваги потребує питання валідації моделей. Стандартна internal validation, заснована на випадковому розбитті даних, майже завжди маскує проблеми domain shift, оскільки зберігає доменні кореляції та дозволяє нормалізаційним механізмам працювати в знайомих умовах. Для медичного продакшену критично важливо використовувати протоколи, що явно моделюють реальні сценарії деплою, зокрема leave-one-hospital-out validation. Такий підхід дозволяє оцінити чутливість моделей до доменних зсувів, порівняти архітектурні рішення в умовах реального domain shift і отримати значно більш надійні оцінки узагальнювальної здатності. У підсумку, побудова стабільних і масштабованих медичних систем III вимагає поєднання доменно-інваріантних архітектурних рішень і коректних протоколів оцінки, а не опори виключно на високі internal метрики.

**ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ  
І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ**

У цій роботі було показано, що проблеми узагальнення медичних нейронних мереж не можна пояснити лише складністю моделей або обмеженістю доступних даних. Основна причина полягає у системних зсувах розподілів між тренувальними та цільовими даними. На відміну від багатьох інших прикладних галузей, у медицині domain shift є типовим сценарієм, зумовленим різноманіттям апаратури, відмінностями в протоколах збору даних, клінічними практиками та характеристиками пацієнтських популяцій. У таких умовах класичне припущення про стаціонарність даних виявляється непридатним для більшості реальних клінічних застосувань.

Центральним результатом роботи є детальний аналіз ролі Batch Normalization як механізму, що імпліцитно закріплює статистику тренувального домену в моделі. Було показано, що BN не є нейтральною операцією нормалізації: вона безпосередньо впливає на формування внутрішніх представлень, роблячи їх оптимальними лише за умови збереження статистичних властивостей тренувальних даних. За наявності

covariate shift це призводить до систематичних зсувів активацій і накопичення помилок у глибоких шарах мережі, що добре пояснює нестабільну поведінку багатьох медичних моделей після деплою.

Спираючись на цей аналіз, у роботі було розглянуто альтернативні підходи до нормалізації та сформульовано практичні рекомендації для інженерного використання медичних моделей. Зокрема, Group Normalization була визначена як найбільш збалансоване рішення з точки зору стабільності навчання та стійкості до доменних зсувів, особливо в умовах малих batch size і мультидомених датасетів. Також було наголошено, що вибір нормалізаційної схеми має супроводжуватися коректними протоколами оцінки, такими як leave-one-hospital-out validation, оскільки стандартна internal validation часто не дозволяє виявити проблеми узагальнення.

У цілому результати цієї роботи підкреслюють необхідність переходу від локальної оптимізації метрик до системного інженерного підходу при розробці медичних систем штучного інтелекту. Архітектурні припущення, механізми нормалізації та протоколи валідації мають розглядатися як взаємопов'язані елементи єдиної системи, а не як незалежні технічні рішення. Лише за такого підходу можливе масштабування медичних моделей за межі лабораторних експериментів і їх безпечне та надійне впровадження в реальну клінічну практику.

### References

1. Zech J. R., Badgeley M. A., Liu M., Costa A. B., Titano J. J., Oermann E. K. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Medicine, 2018. <https://doi.org/10.1371/journal.pmed.1002683>
2. Kelly C. J., Karthikesalingam A., Suleyman M., Corrado G., King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine, 2019. <https://doi.org/10.1186/s12916-019-1426-2>
3. Recht B., Roelofs R., Schmidt L., Shankar V. Do ImageNet classifiers generalize to ImageNet? ICML, 2019. <https://doi.org/10.48550/arXiv.1902.10811>
4. Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML, 2015. <https://doi.org/10.48550/arXiv.1502.03167>
5. Wu Y., He K. Group Normalization. ECCV, 2018. [https://openaccess.thecvf.com/content/ECCV\\_2018/papers/Yuxin\\_Wu\\_Group\\_Normalization\\_ECCV\\_2018\\_paper.pdf](https://openaccess.thecvf.com/content/ECCV_2018/papers/Yuxin_Wu_Group_Normalization_ECCV_2018_paper.pdf)
6. Li Y., Wang N., Shi J., Hou X., Liu J. Adaptive Batch Normalization for practical domain adaptation. Pattern Recognition, 2018. <https://doi.org/10.1016/j.patcog.2018.03.005>
7. Ganin Y., Ustinova E., Ajakan H., et al. Domain-adversarial training of neural networks. JMLR, 2016. <https://doi.org/10.48550/arXiv.1505.07818>
8. Castles S., et al. Domain generalization in medical imaging: A systematic review. IEEE Transactions on Medical Imaging, 2020. <https://doi.org/10.1109/JPROC.2024.3507831>