УДК 616.36-004-073.43:004.8
https://doi.org/10.31612/2616-4868.6.2026.13

ІНТЕРПРЕТОВАНИЙ АНАЛІЗ ЕХОГРАМ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВІДБОРУ ДІТЕЙ НА ЕЛАСТОГРАФІЮ ПЕЧІНКИ: ПІЛОТНЕ ДОСЛІДЖЕННЯ

Володимир В. Солодущенко[1,2], Марина Б. Диба[1], Віталій О. Бабенко[1], Марина В. Глоба[2], Борис А. Тарасюк[1]

[1]Державна установа «Всеукраїнський центр материнства та дитинства Національної академії медичних наук України», м. Київ, Україна
[2]Національний університет охорони здоров'я України імені П. Л. Шупика, м. Київ, Україна

Резюме

Вступ. Стандартне двовимірне ультразвукове дослідження печінки у В-режимі є найбільш доступним методом первинної діагностики, проте його традиційна візуальна інтерпретація при хронічних дифузних захворюваннях у дітей обмежена суб'єктивністю та низькою чутливістю до ранніх стадій фіброзу. Еластографія зсувної хвилі потребує спеціалізованого обладнання і не завжди доступна на первинній ланці. Розробка інтерпретованих інструментів штучного інтелекту для об'єктивного комп'ютерного аналізу ехограм дозволяє здійснювати автоматизоване сортування (тріаж) пацієнтів та раціонально виділяти групи пріоритетного направлення на еластографію печінки.
Мета. Оцінити діагностичну ефективність та клінічну інтерпретованість розробленого комп'ютерного інструменту на основі штучного інтелекту для автоматизованого сортування дітей із хронічними дифузними захворюваннями печінки перед проведенням еластографічного дослідження.
Матеріали та методи. У ретроспективному одноцентровому дослідженні проаналізовано клінічний архів ехограм 105 дітей віком 5-18 років. Загальна вибірка охопила 661 область інтересу (337 двовимірних ультразвукових зображень у В-режимі), отриманих на ультразвукових системах експертного класу. Референтним стандартом слугували результати двовимірної еластографії зсувної хвилі (стратифікація на стадії F0 проти F1–F4 за верифікованими кПа-порогами). Математичну модель логістичної регресії навчали на основі локальних профілів яскравості сірої шкали (32 діапазони) з побудовою просторових карт локальної підтримки та розрахунком інтегрального бала пацієнта. Валідацію виконували за схемою 5-блокової перевірки з виключенням витоку даних. Додатково проведено аналіз селективної класифікації з виділенням зони невизначеного перегляду для сумнівних випадків.
Результати. У бінарній задачі диференціації норми (F0) та фібротичних змін (F1–F4) на рівні пацієнта модель продемонструвала площу під ROC-кривою (AUC) 0,91 [95% довірчий інтервал: 0,84-0,96], збалансовану точність 0,83 [0,76-0,90], чутливість 0,80 [0,68-0,91] та специфічність 0,86 [0,77-0,95] за стандартного операційного порогу 0,50. Застосування скринінгового порогу 0,32 дозволило підвищити чутливість виявлення змін до 0,91. Дослідницький аналіз селективної класифікації показав, що виділення категорії невизначеного перегляду дозволило перевести туди 15 із 17 (88%) помилок базової моделі. Після відсіву невизначених випадків збалансована точність на збереженій підмножині пацієнтів зросла до 0,90, а специфічність досягла 1,00 при чутливості 0,80.
Висновки. Розроблений інтелектуальний метод забезпечує високу точність сортування дітей для подальшої еластографічної стратифікації фіброзу. Просторові карти локальної підтримки роблять логіку моделі клінічно зрозумілою для лікаря, забезпечуючи візуальне пояснення розрахованого бала. Метод селективної класифікації ефективно виділяє сумнівні випадки у зону експертного перегляду, проте низьке покриття (31,4% пацієнтів поза невизначеною зоною) обмежує його застосування як автономного інструменту скринінгу.

Ключові слова: **педіатрична гепатологія, хронічні дифузні захворювання печінки, фіброз печінки, ультразвукова діагностика, ехографія, штучний інтелект, інтерпретованість, селективна класифікація, медичне сортування**

ВСТУП

Своєчасне визначення стадії фіброзу печінки є актуальною клінічною задачею при хронічних дифузних захворюваннях печінки (ХДЗП), зокрема з огляду на значну поширеність неалкогольної жирової хвороби печінки, включно з неалкогольним стеатогепатитом [1], а також хронічних вірусних гепатитів В і С [2]. Оцінювання вираженості фіброзу має важливе прогностичне значення, оскільки сучасні клінічні рекомендації [3] розглядають неінвазивні методи як інструменти визначення тяжкості захворювання печінки та прогнозу його перебігу. Як правило, в ролі морфологічного референтного методу виступає біопсія печінки [4]. Однак, наявність ризику ускладнень викликає необхідність у застосуванні неінвазивних підходів.

Неінвазивна оцінка фіброзу печінки може здійснюватися за допомогою еластографічних методів, зокрема магнітно-резонансної еластографії [5], транзієнтної еластографії [6] та ультразвукової еластографії зсувної хвилі, включно з двовимірною еластографією зсувної хвилі (2D-SWE) [7]. Діагностичну ефективність лабораторних тестів, ультразвукових методів та магнітно-резонансної еластографії для виявлення фіброзу при неалкогольній жировій хворобі печінки узагальнено в метааналізі [8]. Проте, дані методи потребують спеціалізованого обладнання, через що вони не завжди доступні на рівні районних або міських лікарень. Разом з тим, стандартне ультразвукове дослідження (УЗД) у В-режимі є значно доступнішим і, фактично, є рутинним методом у повсякденній клінічній практиці лікаря. Варто зазначити, що його традиційна візуальна інтерпретація при оцінці фіброзу обмежена суб'єктивністю та низькою чутливістю до ранніх структурних змін тканини. Таким чином, виникає практична задача отримання кількісного сигналу зі звичайного В-режиму, за допомогою якого з'являється можливість допоміжного скринінгу. Тобто, даний підхід не замінює еластографію, а лише допомагає визначити тих пацієнтів, яких доцільно скерувати на подальше обстеження.

У цій роботі увагу зосереджено на методах штучного інтелекту та комп'ютеризованого кількісного аналізу медичних зображень, що застосовуються для автоматизованого виявлення, класифікації та стадіювання фіброзу печінки. Зокрема, подібні методи використовували для автоматичного аналізу звичайних УЗ-зображень печінки [9, 10], оцінки неоднорідності її паренхіми на рутинному УЗД [11] та додаткового комп'ютерного аналізу зображень 2D-SWE для визначення стадії фіброзу [12]. Однак не все так однозначно, як здається на першу думку. Для медичного застосування недостатньо отримати лише числовий прогноз. Лікарю необхідно співвідносити результат із зображенням та клінічним маршрутом. Завдяки просторовій локалізації сигналу на області інтересу (ОІ) можна оцінити, наскільки впевненим є цей результат для висновку. Тому для практичного використання таких інструментів важливими є не лише показники діагностичної точності, а й прозоре подання принципів їх розроблення, оцінювання та клінічного застосування. Це узгоджується з рекомендаціями STARD 2015 [13] щодо повного й прозорого звітування досліджень діагностичної точності та TRIPOD+AI [14] щодо опису клінічних прогностичних моделей, зокрема моделей машинного навчання, для полегшення їх оцінювання, перевірки та потенційного впровадження. У задачах стратифікації фіброзу печінки додаткового значення набуває клінічно зрозуміле представлення результатів алгоритму, зокрема через аналіз ділянок інтересу та мультипараметричних ультразвукових ознак [15, 16].

МЕТА

Оцінка діагностичної ефективності та клінічної інтерпретованості (здатності до інтерпретації результатів лікарем) розробленого ШІ-інструменту. Для її досягнення було застосовано аналіз ОІ на зображеннях УЗД печінки, в результаті чого формується інтегральний бал підтримки рішення (що відображає ймовірність та вираженість фіброзу на основі локальних ехо-ознак) та будується просторова карта локального сигналу.

МАТЕРІАЛИ ТА МЕТОДИ

Дизайн дослідження. У цій роботі було проведено ретроспективне одноцентрове дослідження діагностичної ефективності на основі архівних ультразвукових зображень печінки. Слід зазначити, що дизайн та виклад результатів були орієнтовані на принципи STARD 2015 для досліджень діагностичної точності та TRIPOD+AI для опису моделей прогнозування за допомогою систем штучного інтелекту [13, 14].

Усі зображення були отримані в межах клінічного архіву одноцентрових ультразвукових обстежень печінки дітей. Інструментальні дослідження були виконані на ультразвукових системах експертного класу Aplio 500 / Aplio 800 за допомогою конвексного та лінійного датчиків. Варто зазначити, що вибір цієї лінійки апаратів зумовлений тим, що пороги жорсткості печінки 2D-SWE, які слугували референтом, валідовані саме для платформи Aplio. Оскільки алгоритми формування зсувної хвилі, калібрування та кПа-діапазони відрізняються між виробниками (Canon, Aixplorer, Philips, GE), отримані порогові значення не можна автоматично переносити на інші УЗ-системи.

Таким чином, основний аналіз було обмежено цільовим доменом конвексних і стандартних лінійних зображень. Зображення, що були отримані в технічно відмінному високочастотному режимі, не використовувались як окремий валідований домен для бінарного автоматичного висновку через виражений зсув умов їх отримання.

Когорта та референтна розмітка. Вихідний клінічний архів було сформовано з обстежень дітей віком 5-18 років із ХДЗП або без клінічних ознак ураження печінки. Варто зазначити, що до аналізу не були включені випадки таких станів, як:

– гострий гепатит або загострення хронічного процесу з активністю АЛТ/АСТ понад 5 верхніх меж норми;

– ознаки холестазу;

– порушення підготовки до УЗД;

– невалідні еластографічні вимірювання.

Після виключення ОІ з невизначеною або суперечливою еластографічною міткою до загальної вибірки увійшли 788 ОІ від 109 пацієнтів. Для основного оцінювання моделі залишали лише зображення цільового домену, отримані конвексним або стандартним лінійними датчиками (високочастотний/технічно відмінний режим не був включений до основного бінарного аналізу). Після цього доменного обмеження основна когорта становила 661 ОІ від 105 пацієнтів. Характеристика вибірки, доступність клінічних змінних та розподіл за METAVIR-еквівалентними категоріями наведені в таблиці 1.

Значну частину вибірки для оцінювання моделі становлять пацієнти, у яких були відсутні фіброзні зміни (стадія F0). Стадія F1 була представлена лише 4 пацієнтами, тому бінарний аналіз не слід трактувати як надійну оцінку мінімального раннього фіброзу.

Варто також звернути увагу на те, що наявність кількох ОІ на одного пацієнта (або на досліджувану ехограму) мала практичне значення, оскільки дифузні зміни паренхіми можуть бути неоднорідними, а окремі ділянки можуть містити локальні акустичні артефакти. Тому агрегація кількох ОІ дозволяла отримати більш стабільний бал підтримки на рівні пацієнта. Водночас широкий діапазон кількості ОІ на одного пацієнта відображає ретроспективний характер і нерівномірність реального клінічного архіву.

Таблиця 1

Клініко-технічна характеристика вибірки

Параметр	Значення
Вихідний архів	Педіатричні ультразвукові обстеження печінки
Віковий діапазон включених пацієнтів	5-18 років
Індивідуальний розподіл віку	Недоступний для фінальної підмножини ОІ
Стать	Недоступна для фінальної підмножини ОІ
ІМТ	Недоступні для фінальної підмножини ОІ
Лабораторні показники	Недоступні для фінальної підмножини ОІ
Етіологічні категорії	Недоступні для фінальної підмножини ОІ
УЗ-система	Aplio 500 / Aplio 800
Основні режими В-зображень	Конвексний та лінійний датчики
Основна когорта оцінювання	661 ОІ (337 зображень) від 105 пацієнтів
ROI на пацієнта	Медіана 5 [3; 8], діапазон 1-49
Зображень на пацієнта	Медіана 3 [2; 4], діапазон 1-19
F0	59 пацієнтів / 439 ОІ
F1	4 пацієнти / 31 ОІ
F2	16 пацієнтів / 82 ОІ
F3	18 пацієнтів / 87 ОІ
F4	8 пацієнтів / 22 ОІ

У цьому дослідженні як референтну розмітку для задачі класифікації використовували не гістологічні дані, а локально визначені METAVIR-еквівалентні категорії F0–F4, отримані шляхом перетворення значень жорсткості печінки за 2D-SWE у дискретні класи. Такий підхід було застосовано як операційне наближення для навчання та оцінювання алгоритму. Локальний протокол порогової класифікації було сформовано з урахуванням опублікованих рекомендацій щодо ультразвукової еластографії печінки, вимог до якості вимірювань, відомих обмежень інтерпретації кПа-порогів та особливостей застосування еластографії в педіатричних пацієнтів [7, 17-20].

Для перетворення значень жорсткості печінки у класи F0–F4 використовували такі пороги:

– F0: <7,0 кПа;
– F1: 7,0–<9,0 кПа;
– F2: 9,0–<11,5 кПа;
– F3: 11,5–<14,5 кПа;
– F4: ≥14,5 кПа.

Дані пороги використовувались у ролі локальних METAVIR-еквівалентних категорій для протоколу

та систем Aplio 500 / Aplio 800. Їх переносимість на інші УЗ-системи, виробників або еластографічні алгоритми у даній роботі не оцінювалась. Обстеження були виконані після щонайменше 4-годинного голодування. Для еластографічної оцінки проводили не менше 10 валідних вимірювань жорсткості печінки. Результат вважали технічно надійним, якщо відносна варіабельність вимірювань не перевищувала 30%, тобто міжквартильний розмах, поділений на медіану значень жорсткості, становив ≤ 30%. Таким чином, розроблена модель фактично оцінювала здатність зображення УЗД наближати 2D-SWE-похідну METAVIR-еквівалентну стратифікацію, а не гістологічну верифікацію (яка є «золотим стандартом»). Слід зазначити, що верифікація через біопсію в даній когорті не проводилась.

Для фінальної відфільтрованої підмножини ОІ індивідуально верифіковані демографічні, антропометричні, лабораторні та етіологічні змінні не були доступні у погодженому для публікації вигляді, через що вони не були включені до моделі, стратифікованого аналізу або описової статистики. Цей факт є суттєвим клінічним обмеженням, оскільки такі змінні, як вік, стать, ІМТ (індекс маси тіла), етіологія, стеатоз, лабораторна активність запалення та холестаз, найімовірніше, можуть впливати як на ехогенність паренхіми у В-режимі, так і на 2D-SWE показники.

Індекс-тест та ШІ-інтерпретація. Індекс-тестом слугував автоматизований аналіз попередньо виділених деперсоналізованих ОІ печінкової паренхіми на зображеннях УЗД. Дані ОІ були розміщені в ділянках без видимих великих судин, жовчних проток, капсули, реберної тіні та грубих акустичних артефактів.

Логіка аналізу була побудована не як пошук окремої анатомічної ознаки фіброзу, а як порівняння ехогенності ОІ між двома групами. У межах кожного навчального блоку було окремо побудовано два профілі сірих тонів: для ОІ пацієнтів із 2D-SWE-похідною категорією F0 та для ОІ пацієнтів із категоріями F1–F4. Для кожної ОІ аналізувались не окремі пікселі по одному, а їхній розподіл за яскравістю. Спочатку уся шкала сірого була поділена на 32 діапазони. Потім для кожного діапазону було розраховано, яка частина пікселів ОІ має відповідну яскравість (наприклад, якщо 8% пікселів потрапляли в певний діапазон сірого, то частка для даного діапазону становила 0,08). Ці профілі яскравості були окремо усереднені для групи F0 та для групи F1–F4. Після цього кожному діапазону яскравості було присвоєно локальний знак і вагу:

– якщо такий рівень сірого частіше траплявся у навчальній групі F1–F4, він отримував позитивний сигнал;

– якщо частіше траплявся у F0 – негативний сигнал;

– якщо різниця між групами була малою – сигнал був близьким до нуля.

Локальний бал для кожного інтервалу яскравості розраховувався як логарифмічне відношення шансів:

$$S_i = \ln\left(\frac{p_{F1-F4,i} + \varepsilon}{p_{F0,i} + \varepsilon}\right)$$

де $p_{F1-F4,i}$ і $p_{F0,i}$ – згладжені частки пікселів відповідного інтервалу у двох навчальних профілях, а ε – мале стабілізуюче значення, що є необхідним для уникнення ділення на нуль. Таким чином, цей локальний бал фактично виступає в ролі безрозмірної величини в логарифмічних одиницях: він не вимірюється у кПа, відсотках або абсолютній ймовірності.

Надалі цю шкалу яскравості було перенесено назад на ОІ, в результаті чого формувалась карта локальної підтримки. Дана карта демонструє не «місце фіброзу» і не піксельну ймовірність захворювання, а просторовий розподіл інтенсивностей, які в навчальних даних були більш подібними до F0 або до F1–F4. Вона є необхідною для візуальної перевірки правдоподібності сигналу. Тобто, за допомогою неї лікар має можливість оцінити, чи розташований сигнал дифузно в паренхімі, чи виглядає він як локальний артефакт або технічно неоднорідна ділянка. Принцип формування карти локальної підтримки та бала на рівні пацієнта більш наочно продемонстровано на рис. 1.

З рисунка вище можна побачити, що перехід від розподілу сірих тонів до підсумкового висновку здійснюється поетапно (від локального сигналу до загального бала). Для числового висновку було використано не саму «помаранчеву площу», а кількісні ознаки даної карти. До них належали такі показники: середній і медіанний рівень підтримки, розкид сигналу, частка позитивного сигналу та його розподіл у підділянках ОІ. Дані ознаки подавались до моделі логістичної регресії, за допомогою якої формувався бал на рівні ОІ. Далі ці бали усереднювались до рівня пацієнта. Таким чином, підсумковий бал підтримки відображає сукупний інтенсивнісний і просторовий сигнал ОІ, а не просто частку зафарбованої площі на рисунку.

Для клінічної інтерпретації результат було подано у вигляді трьох складових: бал підтримки на рівні пацієнта, карта локального сигналу на ОІ та категорія перегляду (низька підтримка, невизначений перегляд або підвищена підтримка). Дана форма подання відповідає дослідницькому тріажному сценарію. Тобто за допомогою інструменту не встановлюється фінальна стадія фіброзу, а здійснюється допомога у визначенні того, чи потребує даний пацієнт пріоритетного подальшого еластографічного обстеження або експертного перегляду.

Технічні параметри розробленого індекс-тесту та загальний дизайн валідації узагальнені в табл. 2.

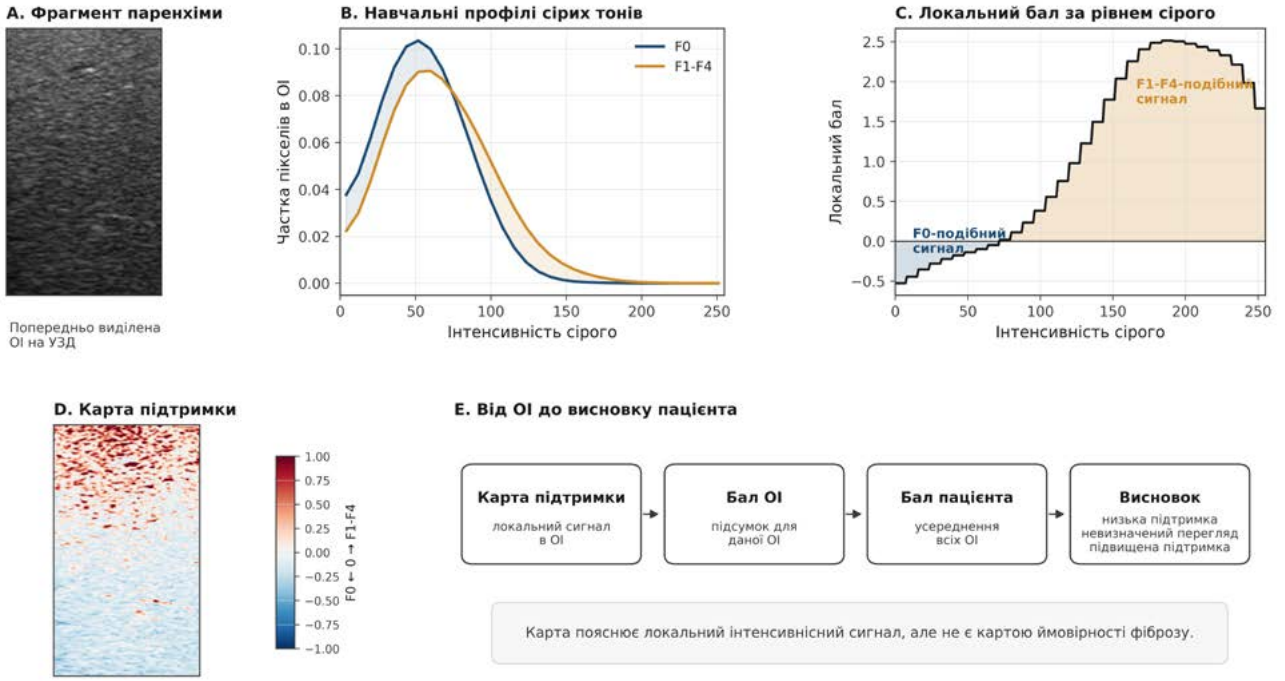

Рисунок 1. Принцип формування карти локальної підтримки та бала на рівні пацієнта.

Таблиця 2

Етапи формування результату індекс-тесту

Етап	Що виконували	Для чого це потрібно
1. Вхідні дані	Аналізували попередньо виділені OI печінкової паренхіми на зображеннях УЗД	Оцінити локальний ехографічний патерн без персональних даних пацієнта
2. Референт для порівняння	Кожного пацієнта відносили до 2D-SWE-похідної METAVIR-еквівалентної категорії за кПа-порогами	Мати єдину еластографічну мітку для навчання та оцінювання моделі
3. Порівняння сірих тонів	У навчальних даних порівнювали розподіл яскравості OI у групах F0 та F1–F4	Визначити, які рівні ехогенності частіше траплялися у кожній групі
4. Карта локальної підтримки	На OI позначали ділянки, інтенсивність яких більше нагадувала F0 або F1–F4	Дати лікарю візуальне пояснення, звідки походить сигнал моделі
5. Бал підтримки	Ознаки карти перетворювали на числовий бал і усереднювали кілька OI до рівня пацієнта	Отримати один підсумковий показник для пацієнта
6. Категорія перегляду	За балом пацієнта відносили до низької підтримки, невизначеного перегляду або підвищеної підтримки	Подати результат у вигляді тріажної підказки, а не як самостійний діагноз
7. Перевірка моделі	Оцінювання виконували в 5 незалежних блоках так, щоб OI одного пацієнта не потрапляли одночасно в навчання і перевірку	Зменшити ризик завищення точності через повторне використання даних одного пацієнта

З табл. 2 видно, що процес розробки та валідації був суворо регламентований. Усі технічні етапи попередньої обробки та навчання моделі були виконані виключно в межах відповідного навчального блоку. Дані тестового блоку не були використані для побудови гістограм, налаштування ознак або навчання моделі, завдяки чому було досягнуто зменшення ризику витоку даних.

Статистичний аналіз та оцінювання моделі. Оцінювання розробленої моделі було виконано на рівні пацієнта. Основною задачею виступало відділення пацієнтів із 2D-SWE-похідною категорією F0 від пацієнтів із категоріями F1–F4. Однак оскільки раніше була зазначена мала кількість F1-випадків, даний аналіз не слід розглядати як надійну оцінку мінімального раннього фіброзу. Тому результати були інтерпретовані більш обережно. Тобто клінічно це розглядалось як відділення F0 від переважно більш виражених 2D-SWE-похідних змін, які у більшості випадків представлені категоріями F2–F4.

Слід зазначити, що для перевірки модель не навчалась і не оцінювалась на одних і тих самих пацієнтах одночасно. Дану вибірку було поділено на 5 блоків із збереженням співвідношення класів F0 та F1–F4. Усі OI одного пацієнта завжди потрапляли виключно в один блок. Це було необхідно зробити для того, щоб під час навчання модель фактично не мала змоги «бачити» інші OI даного пацієнта. Таким чином, на кожному кроці 4 блоки були використані для навчання, а 1 блок – в якості перевірочного. Завдяки

цьому було отримано прогнози для пацієнтів, які не брали участі у відповідному етапі навчання.

З метою уникнення витоку даних, усі операції, що могли залежати від структури вибірки, були виконані лише всередині навчальної частини кожного блоку. Зокрема, це стосувалося таких кроків, як побудова профілів сірих тонів, заміна відсутніх числових значень медіаною навчального блоку, приведення ознак до єдиного масштабу та безпосереднє навчання моделі логістичної регресії. Дані перевірочного блоку використовувались виключно для фінального розрахунку прогнозу.

Діагностична здатність розробленої моделі оцінювалась за допомогою таких метрик: площа під ROC-кривою (AUC), збалансована точність, чутливість, специфічність, а також позитивна та негативна прогностична цінність. Для стандартного порогу були застосовані такі співвідношення: чутливість = TP/(TP+FN), специфічність = TN/(TN+FP), збалансована точність (ЗТ) = (чутливість + специфічність)/2, позитивна прогностична цінність = TP/(TP+FP), негативна прогностична цінність = TN/(TN+FN), де TP, TN, FP і FN відповідали істиннопозитивним, істиннонегативним, хибнопозитивним і хибнонегативним результатам відповідно.

Крім того, додатково було оцінено якість числового бала за допомогою індексу Браєра (Brier score). Даний показник розраховувався як середня квадратична різниця між прогнозованим балом і фактичною бінарною міткою:

$$\text{Brier} = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - y_i)^2,$$

де $\hat{p}_i$ – бал підтримки для і-го пацієнта, а y_i – його фактична бінарна мітка (0 для F0, 1 для F1–F4). Також була розрахована очікувана похибка калібрування (expected calibration error, ECE), за допомогою якої оцінювалась розбіжність між середнім прогнозованим балом і фактичною часткою позитивних випадків у групах пацієнтів із подібними значеннями бала. Ці показники були використані для технічної оцінки стабільності числового бала. Варто зазначити, що сам бал не трактувався як індивідуальна клінічна ймовірність фіброзу. Для розрахунку довірчих інтервалів ключових метрик було використано метод повторного випадкового відбору пацієнтів із поверненням (1000 повторів).

Окремо було виконано дослідницький аналіз із трьома категоріями перегляду: низька підтримка, невизначений перегляд і підвищена підтримка. Метою даного аналізу було не надання автоматичного висновку для всіх пацієнтів, а зменшення кількості надмірно впевнених бінарних рішень у випадках із проміжним або неоднорідним сигналом. Пороги для даних категорій були визначені ретроспективно на основі прогнозів для пацієнтів, які не входили до відповідного навчального блоку. Тому результати селективної класифікації слід розглядати як гіпотезоутворюючі. Для практичного клінічного використання ці пороги потребують подальшої проспективної перевірки.

Етичне схвалення. Дане дослідження було схвалено Комітетом з медичної етики ДУ «Інститут ядерної медицини та променевої діагностики НАМН України» (Протокол № 2 від 9 квітня 2020 р.). На момент проведення дослідження дана установа функціонувала у складі ДУ «Всеукраїнський центр материнства та дитинства НАМН України». Оскільки архітектура роботи мала ретроспективний дизайн, у якому були використані виключно деперсоналізовані дані, отримані під час планових клінічних обстежень, повторне отримання інформованої згоди від пацієнтів не вимагалося за рішенням комітету з етики. Слід зазначити, що дослідження було проведено у повній відповідності до принципів Гельсінської декларації.

РЕЗУЛЬТАТИ

Діагностична ефективність моделі. У технічній бінарній задачі відділення F0 від F1-F4, для розробленої гістограмної ШІ-моделі на рівні пацієнта було досягнуто таких показників, як:

1) AUC 0,91 [95% ДІ 0,84-0,96];

2) ЗТ 0,83 [0,76-0,90];

3) чутливості 0,80 [0,68-0,91];

4) специфічності 0,86 [0,77-0,95].

При цьому, індекс Браєра становив 0,14 [0,11-0,16], а очікувана похибка калібрування склала 0,13[0,10-0,20].

З метою практичної клінічної інтерпретації було розглянуто три операційні точки даного балу підтримки моделі. Вони відображають різні способи перетворення числового балу моделі у бінарний висновок. Слід зазначити, що стандартний поріг 0,50 відповідав базовому рішенню логістичної моделі. Своєю чергою, нижчий (скринінговий) поріг 0,32 був орієнтований на досягнення вищої чутливості. Тобто за рахунок цього відбувається зменшення кількості пропущених F1–F4-подібних випадків. Натомість вищий (специфічний) поріг 0,56 був орієнтований на вищу специфічність, завдяки чому досягається зменшення кількості хибнопозитивних висновків.

Абсолютні числа TP, FP, TN і FN для цих трьох операційних точок наведені в табл. 3, а їхнє просторове розташування на ROC-кривій продемонстровано на рис. 2.

Таблиця 3

Операційні точки за балом підтримки моделі

Операційна точка	Поріг балу підтримки	TP	FP	TN	FN	Чутливість	Специфічність	ЗТ
Скринінг	0,32	42	27	32	4	0,91	0,54	0,73
Стандарт	0,50	37	8	51	9	0,80	0,86	0,83
Специф.	0,56	33	4	55	13	0,72	0,93	0,82

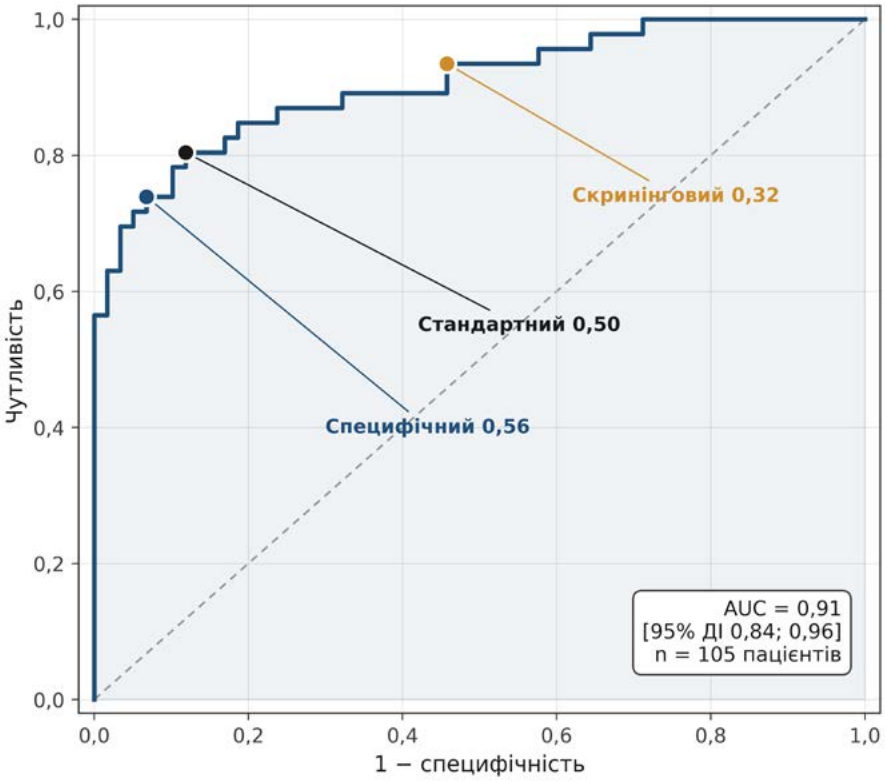

Рисунок 2. ROC-крива гістограмної ШІ-моделі на рівні пацієнта. Показано технічну задачу F0 проти F1–F4. На кривій позначено три операційні точки балу підтримки моделі: скринінговий поріг 0,32, стандартний поріг 0,50 та специфічний поріг 0,56.

З табл. 3 та рис. 2 видно, що при стандартному порозі 0,50 для моделі було досягнуто чутливості 0,80, специфічності 0,86 і ЗТ 0,83. За рахунок застосування скринінгового порогу 0,32 чутливість підвищувалась до 0,91, проте це супроводжувалося зниженням специфічності до 0,54. З іншого боку, специфічний поріг 0,56 забезпечував специфічність на рівні 0,93 при чутливості 0,72. Таким чином, залежно від обраного клінічного сценарію, розроблена модель може бути налаштована або на більш чутливий тріаж, або на обережніше підтвердження підвищеної підтримки.

Варто зазначити, що під час контрольного порівняння з радіомічною моделлю на 93 IBSI-ознаках було продемонстровано зіставну дискримінаційну ефективність. Зокрема, AUC радіоміки становила 0,90 [0,83-0,95], гістограмної моделі – 0,91 [0,84-0,96], а комбінованої моделі – 0,92 [0,86-0,96]. Оскільки спостерігається перекриття довірчих інтервалів, цей факт не дозволяє стверджувати про перевагу однієї моделі над іншою. Фактично, практична відмінність даного гістограмного підходу полягає не у досягненні вищої AUC, а в тому, що в результаті його роботи формується просторова карта локального сигналу. Тобто, ця карта слугує візуальним поясненням, яке можна продемонструвати лікарю разом із розрахованим числовим балом.

Більш складні алгоритми для побудови моделей ШІ, такі як випадковий ліс [21, 22], не застосувались в даному дослідженні з огляду на високий ризик перенавчання на обмеженій вибірці пацієнтів. Крім того, логістична регресія є золотою серединою між ефективністю прогнозування та інтерпретабільностю результатів. Другий фактор є вкрай важливим для клініцистів та представників сфери охорони здоров'я.

Селективна класифікація та невизначений перегляд. Окрім стандартного бінарного рішення, додатково було виконано дослідницький аналіз селективної класифікації. Слід зазначити, що у даному варіанті розроблену модель не примушували видавати автоматичний висновок для кожного пацієнта. Натомість пацієнтів було розподілено на три категорії за балом підтримки, а саме: низька підтримка,

невизначений перегляд та підвищена підтримка. Метою даного підходу виступала необхідність оцінити, чи потрапляють сумнівні та помилкові випадки переважно в проміжну зону, де виникає потреба в експертному перегляді або подальшій еластографічній оцінці.

У загальній когорті 24/105 (22,9%) пацієнтів було віднесено до низької підтримки, 72/105 (68,6%) – до невизначеного перегляду, а 9/105 (8,6%) – до підвищеної підтримки. Пороги для даних категорій були визначені ретроспективно на прогнозах тієї самої когорти, які були отримані поза відповідним навчальним блоком. Тому результати цього аналітичного блоку слід розглядати як гіпотезоутворюючі, а не у вигляді готової клінічної політики. Узагальнення ефективності до та після відсіву невизначених випадків наведені в табл. 4.

Таблиця 4

Ретроспективна селективна класифікація: ефективність стандартного бінарного рішення в усій когорті та після відсіву невизначених випадків

Метрика	Уся когорта	Після відсіву
Пацієнтів	105	33
Покриття	100.0%	31.4%
ЗТ	0,83	0,90
Чутливість	0,80	0,80
Специфічність	0,86	1,00
Помилок	17	2
Помилок, переведених у невизначений перегляд	—	15/17 (88%)

З табл. 4 видно, що ефективність класифікації суттєво змінюється залежно від відсіву. Більш детально рух пацієнтів між даними трьома категоріями перегляду продемонстровано на рис. 3.

Рисунок 3. STARD-подібна схема дослідницької ретроспективної селективної класифікації з категорією невизначеного перегляду.

З рис. 3 можна побачити, що категорія невизначеного перегляду фактично концентрувала в собі більшість помилок стандартного бінарного рішення. Зокрема, 15 із 17 помилок (88,2%) були переведені саме в дану категорію. Після відсіву невизначених випадків залишалась підмножина з 33 пацієнтів, у якій ЗТ становила 0,90, чутливість – 0,80, а специфічність – 1,00.

Разом з тим, покриття даного підходу виявилося досить низьким. Тобто автоматичний бінарний висновок залишався лише для 33/105 пацієнтів (31,4%), тоді як 72/105 пацієнтів (68,6%) потребували невизначеного перегляду. Таким чином, даний аналіз демонструє потенціал розробленого підходу для виявлення сумнівних випадків, однак він не може трактуватися як клінічно готовий скринінг. Висока частка невизначених результатів є критичним обмеженням щодо практичної придатності поточної версії моделі.

Візуальна інтерпретація. Просторова карта локальної підтримки є ключовим елементом клінічної інтерпретованості. Ця карта демонструє, які саме

ділянки OI мають інтенсивнісний патерн, що є більш подібним до F0, а які – до F1–F4. Тобто вона не репрезентує пряму карту піксельної ймовірності наявності захворювання. Репрезентативні приклади даних просторових карт локальної підтримки наведені на рис. 4.

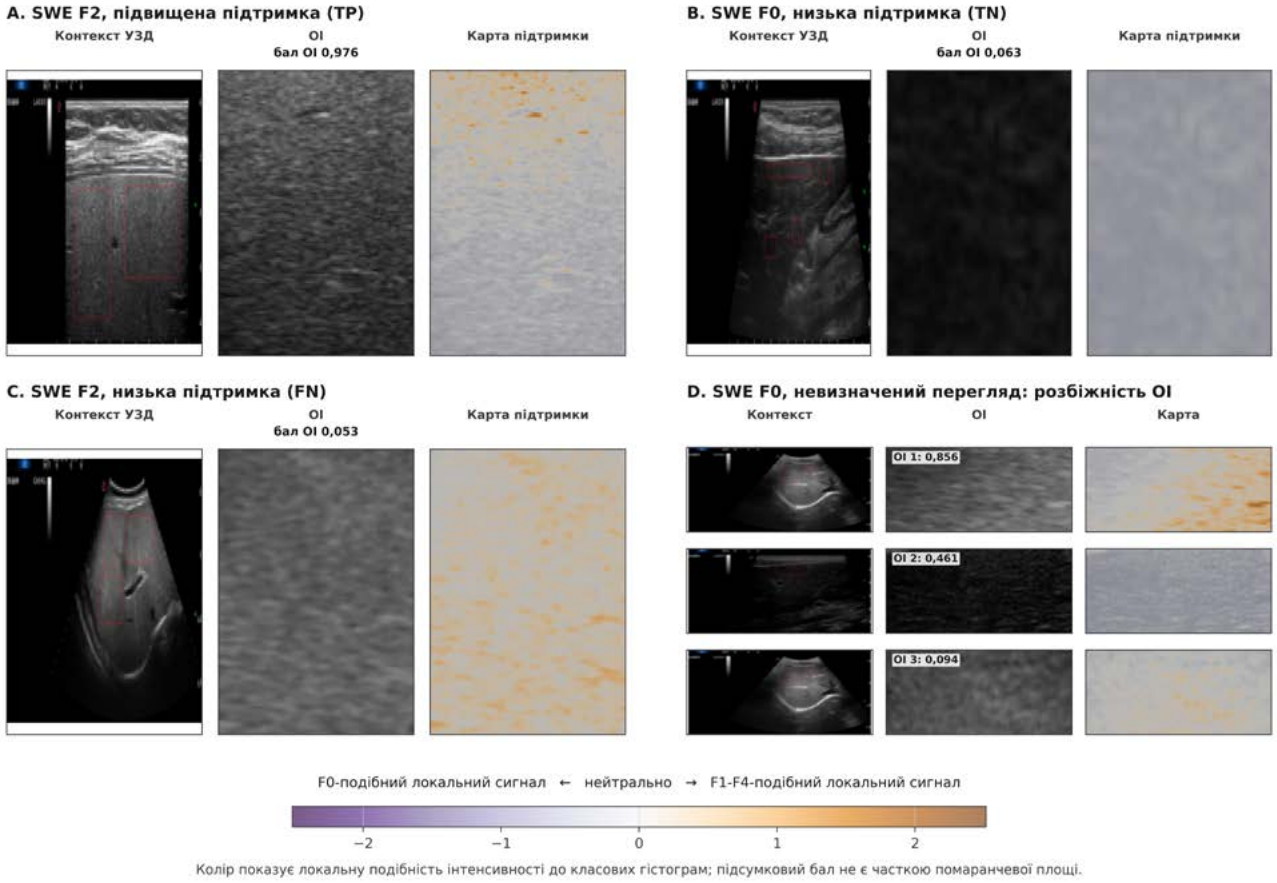

Рисунок 4. Репрезентативні приклади просторових карт локальної підтримки. Для кожного випадку наведено УЗ-контекст із рамкою OI, відповідну OI та карту локальної підтримки. A – SWE F2, підвищена підтримка (TP); B – SWE F0, низька підтримка (TN); C – SWE F2, низька підтримка (FN); D – SWE F0, невизначений перегляд із розбіжністю між OI одного пацієнта.

Із рис. 4 можна побачити, що панель C наочно демонструє обмеження щодо візуальної інтерпретації – наявність окремих F1–F4-подібних локальних ділянок на карті не гарантує досягнення високого підсумкового балу. Оскільки фінальне рішення формується моделлю лише після вилучення просторових ознак та агрегації кількох OI, даний локальний сигнал може усереднюватися або нівелюватися. Слід зазначити, що це подання не доводить патоморфологічний механізм походження сигналу. Проте за допомогою нього у лікаря з'являється можливість візуально оцінити, чи виглядає отриманий результат просторово переконливим або ж, навпаки, сумнівним.

ДИСКУСІЯ

У даному одноцентровому пілотному дослідженні за допомогою інтерпретованого ШІ-аналізу OI на зображеннях УЗД печінки дітей було досягнуто AUC 0,91 та ЗТ 0,83 у технічній задачі F0 проти F1–F4. Варто зазначити, що найважливішим результатом є не сама дискримінаційна метрика, а безпосередньо спосіб подання висновку. Тобто лікар отримує бал підтримки, просторову карту локального сигналу та окрему категорію невизначеного перегляду.

З клінічної точки зору, даний інструмент найдоцільніше розглядати як проміжну ланку між рутинним УЗД у В-режимі та направленням на еластографію. У закладах, де еластографія недоступна на місці, але в наявності є стандартний УЗ-апарат, за допомогою програмного аналізу OI потенційно з'являється можливість пріоритизувати пацієнтів для їх подальшого обстеження у спеціалізованому центрі. Такий сценарій не потребує трактування В-режиму як повноцінної заміни еластографії – навпаки, еластографія і надалі залишається методом уточнення.

Важливою особливістю є те, що система ШІ в даному дослідженні не виступає у вигляді «чорного ящика», який видає лише фінальний клас. Насправді логіка розробленого алгоритму

побудована таким чином: спочатку вивчається, які саме діапазони інтенсивності ОІ частіше пов'язані з SWE-похідними категоріями F0 та F1–F4, а потім ОІ перетворюється на карту локальної підтримки, після чого формується підсумковий бал на рівні пацієнта. Завдяки цьому лікар має змогу бачити не тільки числовий результат, а й просторовий розподіл проміжного сигналу. Водночас цю карту слід читати виключно як пояснення джерела сигналу, а не як самостійний діагностичний шар. Тобто помаранчеві ділянки означають локальну F1–F4-подібність за інтенсивністю, проте фінальний бал залежить також від розподілу, неоднорідності та узгодженості сигналу між кількома ОІ. Саме це і відрізняє запропонований підхід від класичних моделей, які подають лише фінальну ймовірність без візуального пояснювального зображення.

Крім того, другим клінічно важливим компонентом є дослідницька селективна класифікація. На практиці виникає чимало нюансів, адже у реальній діагностиці невизначеність ні в якому разі не повинна маскуватися під впевнений результат. У проведеному аналізі категорія невизначеного перегляду охопила більшість помилок стандартного бінарного рішення, але це було досягнуто ціною дуже низького покриття. Тобто автоматичний висновок залишався тільки для 31% пацієнтів. Тому даний блок фактично не демонструє готову клінічну політику, а, найімовірніше, виявляє межі практичної придатності моделі в поточній когорті. Оскільки пороги були визначені ретроспективно на відкладених прогнозах тієї самої когорти, цей результат слід розглядати виключно як гіпотезоутворюючий.

Контрольне порівняння з радіомікою продемонструвало, що розроблена гістограмна модель не поступалася їй за метрикою AUC у межах даної когорти. Однак це не слід трактувати як доказ загальної переваги гістограмного підходу. Варто зазначити, що радіомічна модель має самостійну методологічну цінність, оскільки використовує стандартизовані кількісні характеристики медичних зображень, розрахунок яких уніфіковано міжнародною ініціативою IBSI [23]. Це робить результати такого аналізу більш відтворюваними та порівнюваними між різними дослідженнями та програмними інструментами [24]. Натомість основною перевагою гістограмного підходу в контексті даної роботи виступає клінічно зрозуміла карта локальної підтримки, яка може бути використана лікарем УЗД для візуального пояснення отриманого результату.

Дане дослідження має низку суттєвих обмежень. До основних факторів належать, зокрема:

1. Відсутність погоджених індивідуальних демографічних, антропометричних, лабораторних та етіологічних даних. Це є важливим обмеженням зовнішньої валідності, оскільки не дозволяє оцінити репрезентативність когорти та вплив таких параметрів, як вік, стать, ІМТ, стеатоз, етіологія захворювання, а також вплив лабораторної активності запалення й холестазу на ехогенність паренхіми і 2D-SWE-показники.

2. Через малу кількість F1-випадків у даній роботі неможливо надійно оцінити виявлення мінімального раннього фіброзу, тому отримані результати необхідно трактувати переважно як відділення F0 від більш виражених (F2–F4) 2D-SWE-похідних змін.

3. Референтом виступала 2D-SWE з перекодуванням кПа-порогів у METAVIR-еквівалентні категорії F0–F4, а не біопсія. Отже, розроблена модель наближала еластографічну, а не гістологічну стратифікацію.

4. Ретроспективний та одноцентровий дизайн дослідження, виконаний на зображеннях систем однієї лінійки Aplio 500 / Aplio 800, через що локальні порогові значення та отриманий В-режимний сигнал не можна автоматично переносити на інші сканери, виробників або популяції.

5. Низьке покриття після відсіву. Оскільки автоматичний бінарний висновок залишався лише для 33/105 пацієнтів, поточну версію не слід розглядати як автономний скринінговий інструмент.

6. Неповна відтворюваність точних параметрів кожної ОІ з ретроспективного архіву (зокрема сегмент, глибина, налаштування gain/TGC та засліплення оператора до SWE-категорії). Цей аспект необхідно стандартизувати в наступному проспективному протоколі.

7. Ретроспективне оцінювання порогів селективної класифікації. Оскільки вони були визначені на відкладених прогнозах тієї самої когорти, вони потребують обов'язкової проспективної перевірки.

Таким чином, наступним етапом має стати проспективна зовнішня валідація з наперед зафіксованими порогами та стандартизованим протоколом отримання ОІ. Оптимальний дизайн майбутнього дослідження повинен включати такі складові, як принаймні кілька клінічних центрів, різні УЗ-системи, синхронні лабораторні показники для порівняння з індексами FIB-4 та APRI, а також біопсійну верифікацію в тих випадках, де вона є клінічно показаною.

ВИСНОВКИ

За результатами аналізу ретроспективної одноцентрової когорти зі 105 педіатричних пацієнтів та 661 областей інтересу, для розробленої інтерпретованої гістограмної моделі штучного

інтелекту на основі ультразвукового дослідження в B-режимі було досягнуто збалансованої точності 0,83 у технічній задачі F0 проти F1–F4. Оскільки F1 = 4, дану задачу слід клінічно трактувати переважно як відділення F0 від F2–F4-домінантних 2D-SWE-похідних змін.

Було встановлено, що у межах дослідницького аналізу селективної класифікації 15 із 17 помилок стандартного бінарного рішення були переведені у категорію невизначеного перегляду. Проте на збереженій підмножині з 33 пацієнтів збалансована точність становила 0,90, що фактично було досягнуто ціною виключення 72/105 пацієнтів із автоматичного висновку. Даний результат потребує обов'язкової проспективної перевірки з наперед зафіксованими порогами.

Практична цінність даного підходу полягає у формуванні зрозумілого для лікаря бала підтримки та просторової карти локального сигналу, а не інструменту для автономного встановлення діагнозу. Таким чином, запропонований інструмент слід розглядати виключно як дослідницьку підтримку штучного інтелекту в тріажі щодо подальшого еластографічного обстеження.

Перспективи подальших досліджень. Подальший розвиток запропонованого методу передбачає проведення проспективної зовнішньої валідації на базі кількох клінічних центрів із використанням ультразвукових діагностичних систем різних виробників та стандартизованим протоколом отримання ехограм. Окремим перспективним напрямком є порівняльний аналіз діагностичної ефективності розробленого комп'ютерного методу із сироватковими індексами фібротичних змін (зокрема FIB-4 та APRI), а також зіставлення результатів із даними гістоморфологічного дослідження біоптатів печінки у клінічно показаних випадках.

ДОТРИМАННЯ ЕТИЧНИХ НОРМ

Етичне схвалення. Дане дослідження було схвалено Комітетом з медичної етики ДУ «Інститут ядерної медицини та променевої діагностики НАМН України» (Протокол № 2 від 9 квітня 2020 р.). Дослідження було проведено у повній відповідності до принципів Гельсінської декларації.

Доступність даних та коду. Вихідний код розробленого інструменту є відкритим і доступний за адресою: https://github.com/vbabenko97/LiverRight. Первинні ультразвукові зображення не можуть бути оприлюднені у відкритому доступі через чинні етичні обмеження та захист персональних даних пацієнтів. Анонімізовані похідні дані моделі можуть бути надані авторами за обґрунтованим запитом до автора-кореспондента.

Використання генеративного штучного інтелекту. Генеративний штучний інтелект (ChatGPT 5.5 Thinking, OpenAI) було використано для лінгвістичного та стилістичного редагування тексту рукопису без зміни наукового змісту. Інструменти генеративного ШІ не були використані для статистичного аналізу, зміни результатів або клінічної інтерпретації. Усі редагування були переглянуті та затверджені авторами.

ФІНАНСУВАННЯ ТА КОНФЛІКТ ІНТЕРЕСІВ

Дане дослідження виконано без залучення зовнішнього фінансування. Робота виконана в межах науково-дослідної роботи (НДР) «Розробка технології цифрової обробки зображень променевої діагностики на базі елементів штучного інтелекту при дифузній патології печінки у пацієнтів різних вікових груп» (державний реєстраційний номер 0120U000437).

Автори заявляють про відсутність будь-якого конфлікту інтересів.

ВНЕСОК АВТОРІВ

Солодущенко В. В.[B, C, D, F]

Диба М. Б.[B, F]

Бабенко В. О.[A, B, C, D, F]

Глоба М. В.[A, E, F]

Тарасюк Б. А.[A, D, E, F]

REFERENCES

1. Younossi, Z. M., Golabi, P., Paik, J. M., Henry, A., Van Dongen, C., & Henry, L. (2023). The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): A systematic review. Hepatology, 77(4), 1335-1347. https://doi.org/10.1002/hep.32723

2. Hofstraat, S. H. I., Falla, A. M., Duffell, E. F., Hahné, S. J. M., Amato-Gauci, A. J., Veldhuijzen, I. K., & Tavoschi, L. (2017). Current prevalence of chronic hepatitis B and C virus infection in the general population, blood donors and pregnant women in the EU/EEA: A systematic review.

Epidemiology and Infection, 145(14), 2873-2885. https://doi.org/10.1017/S0950268817001947
3. European Association for the Study of the Liver. (2021). EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis. Journal of Hepatology, 75(3), 659-689. https://doi.org/10.1016/j.jhep.2021.05.025
4. Rockey, D. C., Caldwell, S. H., Goodman, Z. D., Nelson, R. C., & Smith, A. D. (2009). Liver biopsy. Hepatology, 49(3), 1017-1044. https://doi.org/10.1002/hep.22742
5. Singh, S., Venkatesh, S. K., Wang, Z., Miller, F. H., Motosugi, U., Low, R. N., Hassanein, T., Asbach, P., Godfrey, E. M., Yin, M., Chen, J., Keaveny, A. P., Bridges, M., Bohte, A., Murad, M. H., Lomas, D. J., Talwalkar, J. A., & Ehman, R. L. (2015). Diagnostic performance of magnetic resonance elastography in staging liver fibrosis: A systematic review and meta-analysis of individual participant data. Clinical Gastroenterology and Hepatology, 13(3), 440-451.e6. https://doi.org/10.1016/j.cgh.2014.09.046
6. Friedrich-Rust, M., Ong, M. F., Martens, S., Sarrazin, C., Bojunga, J., Zeuzem, S., & Herrmann, E. (2008). Performance of transient elastography for the staging of liver fibrosis: A meta-analysis. Gastroenterology, 134(4), 960-974. https://doi.org/10.1053/j.gastro.2008.01.034
7. Barr, R. G., Wilson, S. R., Rubens, D., Garcia-Tsao, G., & Ferraioli, G. (2020). Update to the Society of Radiologists in Ultrasound liver elastography consensus statement. Radiology, 296(2), 263-274. https://doi.org/10.1148/radiol.2020192437
8. Xiao, G., Zhu, S., Xiao, X., Yan, L., Yang, J., & Wu, G. (2017). Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: A meta-analysis. Hepatology, 66(5), 1486-1501. https://doi.org/10.1002/hep.29302
9. Liu, Z., Huang, B., Wen, H., Lu, Z., Huang, Q., Jiang, M., Dong, C., Liu, Y., Chen, X., & Lin, H. (2023). Automatic diagnosis of significant liver fibrosis from ultrasound B-mode images using a handcrafted-feature-assisted deep convolutional neural network. IEEE Journal of Biomedical and Health Informatics, 27(10), 4938-4949. https://doi.org/10.1109/JBHI.2023.3295078
10. Lee, J. H., Joo, I., Kang, T. W., Paik, Y. H., Sinn, D. H., Ha, S. Y., Kim, K., Choi, C., Lee, G., Yi, J., & Bang, W. C. (2020). Deep learning with ultrasonography: Automated classification of liver fibrosis using a deep convolutional neural network. European Radiology, 30(2), 1264-1273. https://doi.org/10.1007/s00330-019-06407-1
11. Li, J., Qureshi, M., Gupta, A., Anderson, S. W., Soto, J., & Li, B. (2019). Quantification of degree of liver fibrosis using fibrosis area fraction based on statistical chi-square analysis of heterogeneity of liver tissue texture on routine ultrasound images. Academic Radiology, 26(8), 1001-1007. https://doi.org/10.1016/j.acra.2018.10.004
12. Wang, K., Lu, X., Zhou, H., Gao, Y., Zheng, J., Tong, M., Wu, C., Liu, C., Huang, L., Jiang, T., Meng, F., Lu, Y., Ai, H., Xie, X.-Y., Yin, L., Liang, P., Tian, J., & Zheng, R. (2019). Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: A prospective multicentre study. Gut, 68(4), 729-741. https://doi.org/10.1136/gutjnl-2018-316204
13. Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C. W., Kressel, H. Y., Rifai, N., Golub, R. M., Altman, D. G., Hooft, L., Korevaar, D. A., Cohen, J. F., & STARD Group. (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. BMJ, 351, h5527. https://doi.org/10.1136/bmj.h5527
14. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., Hoffman, M. M., Kengne, A. P., Lam, E., Lee, N., Loder, E. W., Maier-Hein, L., Mateen, B. A., McCradden, M. D., Oakden-Rayner, L., Ordish, J., Parnell, R., Rose, S., Singh, K., Wynants, L., & Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078378. https://doi.org/10.1136/bmj-2023-078378
15. Babenko, V. O., Nastenko, Ye. A., Solodushchenko, V. V., Pavlov, V. A., & Dykan, I. M. (2024). Iierarkhichnyi alhorytm multyklasyfikatsii stadii fibrozu pechinky z intehrovanym analizom oblastei interesu. Biomedical Engineering and Technology, 14(2). https://doi.org/10.20535/2617-8974.2024.14.301108
16. Babenko, V. O., Solodushchenko, V. V., Dyba, M. B., Hloba, M. V., & Tarasiuk, B. A. (2026). Efektyvnist multyparametrychnoi ultrazvukovoi diahnostyky z prohramamy shtuchnoho intelektu pry dyfuznykh zakhvoriuvanniakh pechinky u ditei. Clinical and Preventive Medicine, 1(47), 104-112. https://doi.org/10.31612/2616-4868.1.2026.13
17. Tran, L. C., Ley, D., Bourdon, G., Coopman, S., Lerisson, H., Tillaux, C., Béhal, H., Gottrand, F., & Aumar, M. (2022). Noninvasive pediatric liver fibrosis measurement: Two-dimensional shear wave elastography compared with transient elastography. Frontiers in Pediatrics, 10, 849815. https://doi.org/10.3389/fped.2022.849815

18. Dyba, M. B., & Berezenko, V. S. (2023). Mozhlyvosti elastohrafii zsuvnoi khvyli v diahnostytsi fibrozu pechinky ta monitorynhu autoimunnykh zakhvoriuvan pechinky u ditei. Modern Pediatrics. Ukraine, 8(136), 34-41. https://doi.org/10.15574/SP.2023.136.34
19. Dietrich, C. F., Ferraioli, G., Sirli, R., Popescu, A., Sporea, I., Pienar, C., Kunze, C., Taut, H., Schrading, S., Bota, S., Schreiber-Dietrich, D., Fang, C., & Yi, D. (2019). General advice in ultrasound based elastography of pediatric patients. Medical Ultrasonography, 21(3), 315-326. https://doi.org/10.11152/mu-2063
20. Ferraioli, G., Barr, R. G., Berzigotti, A., Sporea, I., Wong, V. W.-S., Reiberger, T., Karlas, T., Thiele, M., Cardoso, A. C., Ayonrinde, O. T., Castera, L., Dietrich, C. F., Iijima, H., Lee, D. H., Kemp, W., Oliveira, C. P., & Sarin, S. K. (2024). WFUMB Guideline/Guidance on Liver Multiparametric Ultrasound: Part 1. Update to 2018 Guidelines on Liver Ultrasound Elastography. Ultrasound in Medicine & Biology, 50(8), 1071-1087. https://doi.org/10.1016/j.ultrasmedbio.2024.03.013
21. Nastenko, I., Maksymenko, V., Potashev, S., Pavlov, V., Babenko, V., Rysin, S., Matviichuk, O., & Lazoryshinets, V. (2021). Random Forest Algorithm Construction for the Diagnosis of Coronary Heart Disease Based on Echocardiography Video Data Streams. Innovative Biosystems and Bioengineering, 5(1), 61-69. https://doi.org/10.20535/ibb.2021.5.1.225794
22. Petrunina, O., Shevaga, D., Babenko, V., Pavlov, V., Rysin, S., & Nastenko, I. (2021). Comparative Analysis of Classification Algorithms in the Analysis of Medical Images From Speckle Tracking Echocardiography Video Data. Innovative Biosystems and Bioengineering, 5(3), 153-166. https://doi.org/10.20535/ibb.2021.5.3.234990
23. Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J. W. L., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R., Boellaard, R., Bogowicz, M., Boldrini, L., Buvat, I., Cook, G. J. R., Davatzikos, C., Depeursinge, A., Desseroit, M.-C., Dinapoli, N., Dinh, C. V., Echegaray, S., Lambin, P., Leijenaar, R. T. H., & van Griethuysen, J. (2020). The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology, 295(2), 328-338. https://doi.org/10.1148/radiol.2020191145
24. van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillion-Robin, J.-C., Pieper, S., & Aerts, H. J. W. L. (2017). Computational radiomics system to decode the radiographic phenotype. Cancer Research, 77(21), e104-e107. https://doi.org/10.1158/0008-5472.CAN-17-0339

Abstract

INTERPRETABLE AI-DRIVEN ULTRASOUND ANALYSIS FOR THE SELECTION OF PEDIATRIC PATIENTS FOR LIVER ELASTOGRAPHY: A PILOT STUDY

Volodymyr V. Soloduschenko[1,2], Maryna B. Dyba[1], Vitalii O. Babenko[1], Maryna V. Globa[2], Borys A. Tarasiuk[1]

[1]SI «Ukrainian Center of Maternity and Childhood of NAMS of Ukraine», Kyiv, Ukraine
[2]Shupyk National Healthcare University of Ukraine, Kyiv, Ukraine

Introduction. Standard two-dimensional B-mode liver ultrasound is the most accessible modality for primary diagnosis; however, its conventional visual interpretation in pediatric chronic diffuse liver diseases is limited by subjectivity and low sensitivity to early-stage fibrosis. Shear wave elastography requires specialized equipment and is frequently unavailable in primary care settings. The development of interpretable artificial intelligence (AI) tools for objective computer-aided analysis of echograms enables automated patient triage and rational identification of priority cohorts for referral to liver elastography.

Aim. To evaluate the diagnostic performance and clinical interpretability of a developed AI-based computer tool designed for the automated triage of children with chronic diffuse liver diseases prior to elastographic examination.

Materials and methods: In a retrospective single-center study, the clinical echogram archive of 105 children aged 5-18 years was analyzed. The total dataset comprised 661 regions of interest (ROIs) extracted from 337 two-dimensional B-mode ultrasound images acquired via expert-class ultrasound systems. Two-dimensional shear wave elastography (2D-SWE) served as the reference standard (stratification into stages F0 vs. F1–F4 based on verified kPa thresholds). A logistic regression model was trained using local grayscale intensity profiles (32 bins) to construct spatial local-support maps and compute an integrated patient score. Validation was performed using a 5-fold cross-validation scheme configured to strictly prevent data leakage (patient-level splitting). Additionally, a selective classification analysis was conducted by introducing an indeterminate/uncertainty zone (rejection option) for equivocal cases.

Results. In the binary task of differentiating normal tissue (F0) from fibrotic changes (F1–F4) at the patient level, the model achieved an area under the ROC curve (AUC) of 0.91 [95% confidence interval: 0.84-0.96], a balanced accuracy of 0.83 [0.76-0.90], a sensitivity of 0.80 [0.68-0.91], and a specificity of 0.86 [0.77-0.95] at a standard operating threshold of 0.50. Implementing a screening threshold of 0.32 increased the sensitivity for detecting fibrotic changes to 0.91. Exploratory selective classification analysis revealed that routing equivocal cases to the indeterminate category captured 15 out of 17 (88%) baseline model errors. After filtering out these indeterminate cases, the balanced accuracy on the retained subset of patients increased to 0.90, and specificity reached 1.00 with a sensitivity of 0.80.

Conclusions. The developed intelligent method provides high triage accuracy for children requiring subsequent elastographic fibrosis stratification. Spatial local-support maps render the model's logic clinically interpretable for clinicians by providing a visual explanation of the calculated score. While the selective classification method effectively routes equivocal cases to an expert review zone, the low coverage (only 31.4% of patients remained outside the indeterminate zone) limits its utility as a standalone screening tool.

Keywords: **pediatric hepatology, chronic diffuse liver diseases, liver fibrosis, ultrasound diagnostics, echography, artificial intelligence, interpretability, selective classification, medical triage**

Received: 07.05.2026
Accepted: 30.06.2026
Published: 11.09.2026