

## ПОБУДОВА АЛГОРИТМУ ВИПАДКОВОГО ЛІСУ ДЛЯ ДІАГНОСТИКИ ІШЕМІЧНОЇ ХВОРОБИ СЕРЦЯ ЗА ПОТОКАМИ ВІДЕОДАНИХ ЕХОКАРДИОГРАФІЙ

Є.А. Настенко<sup>1,2</sup>, В.Б. Максименко<sup>1,2</sup>, С.В. Поташев<sup>2</sup>, В.А. Павлов<sup>1</sup>, В.О. Бабенко<sup>1\*</sup>,  
С.В. Рисін<sup>1,2</sup>, О.В. Матвійчук<sup>1</sup>, В.В. Лазоришинець<sup>2</sup>

<sup>1</sup>КПІ ім. Ігоря Сікорського, Київ, Україна

<sup>2</sup>Національний інститут серцево-судинної хірургії імені М.М. Амосова, Київ, Україна

\*Corresponding author: vbabenko2191@gmail.com

Received 28 February 2021; Accepted 26 March 2021

**Проблематика.** Дослідження останніх років показують, що серцево-судинні захворювання, включаючи ішемічну хворобу серця, є основною причиною смертності й одним із основних факторів інвалідності в усьому світі. Виявлення випадків цього типу захворювань за останні 30 років збільшилось з 271 до 523 млн, а кількість смертей – з 12,1 до 18,6 млн. Серцево-судинні захворювання є головною причиною смертності населення України, і за цим показником країна лишається одним зі світових лідерів. Ішемічна хвороба серця є провідним чинником втрати здоров'я України, і для її своєчасного виявлення все більше використовують сучасні методи діагностики, включаючи алгоритми машинного навчання.

**Мета.** За даними спекл-трекінг ехокардіографії методом випадкового лісу побудувати класифікаційні алгоритми діагностики порушень кінематики скорочень лівого шлуночка серця у хворих на ішемічну хворобу серця в умовах стану спокою та при застосуванні ехострестесту з добутамінною пробою.

**Методика реалізації.** Методом спекл-трекінг ехокардіографії було обстежено 40 пацієнтів із наявністю ішемічної хвороби серця та 16 осіб, у яких патологію серця не виявлено. Ехокардіографія реєструвалась у В-режимі в трьох позиціях: по довгій вісі, у 4- та 2-камерній позиціях. Усього було використано 6245 кадрів відеопотоку: 1871 – без порушень серцевої діяльності та 4374 – за наявності патології під час обстеження. 56 пацієнтів (2509 кадрів відеоданих) обстежено без застосування добутамінної проби та 38 пацієнтів (3736 кадрів відеоданих) – із застосуванням ехострестесту з добутамінною пробою, якщо порушень у стані спокою виявлено не було. Під наглядом анестезіолога застосовувались дози добутаміна 10, 20 і 40 мкг. Як інформативні ознаки були використані дані текстурного аналізу зображень. Для побудови алгоритму виявлення ішемічної хвороби серця було застосовано алгоритм випадкового лісу.

**Результати.** На першому етапі дослідження було побудовано діагностичні алгоритми нормопатологія для стану спокою та доз добутаміну 10, 20 і 40 мкг. Перед застосуванням алгоритму вибірки випадковим чином було поділено на навчальні (70 %) й тестові (30 %). Класифікатори було оцінено за показниками точності, чутливості та специфічності. За даними тестових вибірок точність діагностичних висновків варіювалась від 97 до 99 %. На другому етапі дослідження, для підвищення універсальності моделей, було побудовано класифікатор для всіх зображень, без поділення на дози добутаміну. Точність за тестовими вибірками також була в межах від 96,6 до 97,8 %. Для побудови діагностичних алгоритмів методом випадкового лісу було використано дані текстурного аналізу зображень.

**Висновки.** Отримано високоточні моделі класифікації за допомогою використання алгоритму випадкового лісу. Розроблені моделі можуть бути застосовані для аналізу ехокардіограм, отриманих у В-режимі на обладнанні, яке не оснащене технологією спекл-трекінг.

**Ключові слова:** випадковий ліс; діагностика; спекл-трекінг ехокардіографія; ехострестест із добутамінном; ішемічна хвороба серця.

### Вступ

Знаходження нових інформативних ознак нормальної та патологічної деформації міокарда за даними спекл-трекінг ехокардіографії (СТЕ) є актуальним завданням на сучасному етапі розвитку медицини та розробки нових медико-

технічних діагностичних засобів. Необхідність стандартизації діагностичних підходів при застосуванні СТЕ привела до створення медиками та виробниками медичного обладнання сумісного ініціативного документа, що регламентував основні процедури й алгоритми обробки інформації СТЕ [1]. Цей документ наразі став

міжнародним робочим стандартом. Основні положення цього документа набули подальшого розвитку і в Україні [2].

Застосування стрес-тестів разом із STE-обстеженням дає змогу виявляти навіть латентні порушення роботи серця в пацієнтів із різними стадіями розвитку ішемічної хвороби серця (ІХС).

Інтенсивний розвиток сучасних інформаційних технологій стимулював пошук нових і вдосконалення відомих комп'ютерних алгоритмів для підтримки прийняття медичних рішень щодо наявності таких порушень та їх ранньої діагностики.

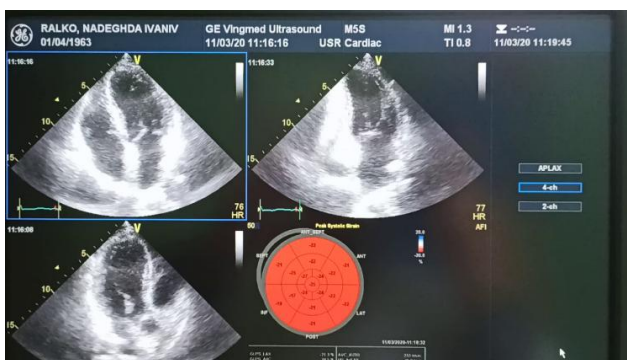
*Мета* нашої роботи – за даними спекл-трекінг ехокардіографії методом випадкового лісу побудувати класифікаційні алгоритми діагностики порушень кінематики скорочень лівого шлуночка серця у хворих на ІХС в умовах стану спокою та при застосуванні ехокардіографії із добутаміновою пробою.

## Матеріал і методи

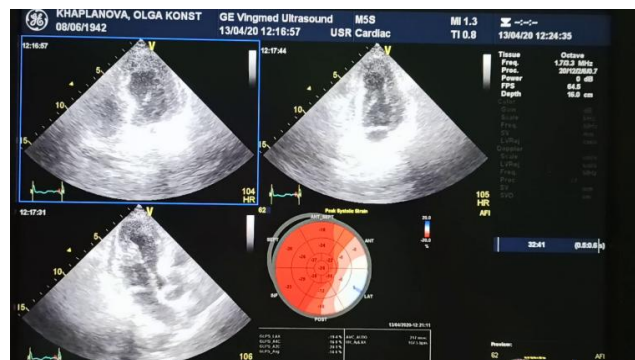
У роботі були використані дані STE-досліджень 16-ти пацієнтів, у яких під час обстежень порушень кінематики міокарда виявлено не було, та дані 40 пацієнтів, у яких під час обстеження були виявлені ознаки ІХС. При застосуванні цієї технології проводиться реєстрація ехокардіограми у В-режимі в інформативних проєкціях серця.

STE реєструвалась у В-режимі в трьох позиціях (рис. 1): по довгій вісі, у 4- та 2-камерній позиціях [1, 2].

Кожна позиція відповідала певній комбінації сегментів лівого шлуночка серця, розміщених у басейнах магістральних коронарних артерій.



а



б

**Рисунок 1:** Спекл-трекінг ехокардіографія: (а) серця в нормі (коло повністю червоне, всі сегменти лівого шлуночка серця (ЛШС) функціонують нормально), (б) при патології серця (коло змінює колір, стрейни деяких сегментів ЛШС патологічно змінені)

N1 = 56 пацієнтів (2509 спостережень) без застосування добутамінової проби та N2 = 38 пацієнтів (3736 спостережень) із застосуванням ехокардіографії (ЕСТ) із добутаміновою пробою, якщо порушень у стані спокою виявлено не було. За згодою пацієнтів під наглядом анестезіолога застосовувались дози добутаміна 10, 20 та 40 мкг. Проба негайно припинялась при появі найменших ознак дискомфорту та (або) порушень серцевої діяльності. У N3 = 3 пацієнтів в умовах застосування ЕСТ патології виявлено не було.

Усього було використано 6245 кадрів відеопотоку: 1871 – без порушень серцевої діяльності, та 4374 – за наявності патології при STE-обстеженні.

Дослідження були проведені відповідно до Гельсінської декларації Всесвітньої медичної асоціації “Етичні принципи медичних досліджень за участю людини у якості об’єкта дослідження”, Конвенції Ради Європи про права людини, Конвенції щодо захисту особи стосовно автоматичної обробки персональних даних.

## Результати

Кожне відео STE було розбито на фрейми, розрізані на три окремих зображення, кожне з яких відповідає різним проєкціям серця (рис. 2) при ехокардіографії.

Для кожної окремої проєкції були побудовані окремі класифікатори. Для того щоб розв’язати задачу бінарної класифікації (норма – патологія), з кожного фрейму були отримані інформативні ознаки (Feature Engineering). Для цього було застосовано текстурний аналіз, методика якого викладена в працях [3–7].

Кожен фрейм являє собою матрицю пікселів (точок зображення) розміром  $n \times m$ , де кожне значення пікселя має певний відтінок сірого. Оскільки записи СТЕ проводяться у відтинках сірого, то пікселі мають градації відтінків сірого в діапазоні від 0 (чорний колір) до 255 (білий колір).

Ідея текстурного аналізу полягає в розрахунку різного роду ознак на основі матриць розподілу відтінків сірого. Такими ознаками можуть бути: середнє значення градацій сірого в матриці, стандартне відхилення, медіана, асиметрія, ексцес тощо.

Перед отриманням інформативних текстурних ознак кожне зображення було нормалізовано. Це було зроблено для отримання ознак, незалежних від індивідуальних налаштувань яскравості та контрастності зображення кожним лікарем. Для нормалізації були застосовані методи горизонтальної та вертикальної диференціації, вперше представлені в праці [8].

Горизонтальна диференціація виконувалася за такими формулами:

$$g_{i,j}^{hor} = g_{i,j} - g_{i-1,j}, \quad g_{i,j}^{*hor} = g_{i,j}^{hor} + |g_{i,j}^{hor}|.$$

Візуально це має вигляд, як на рис. 3.

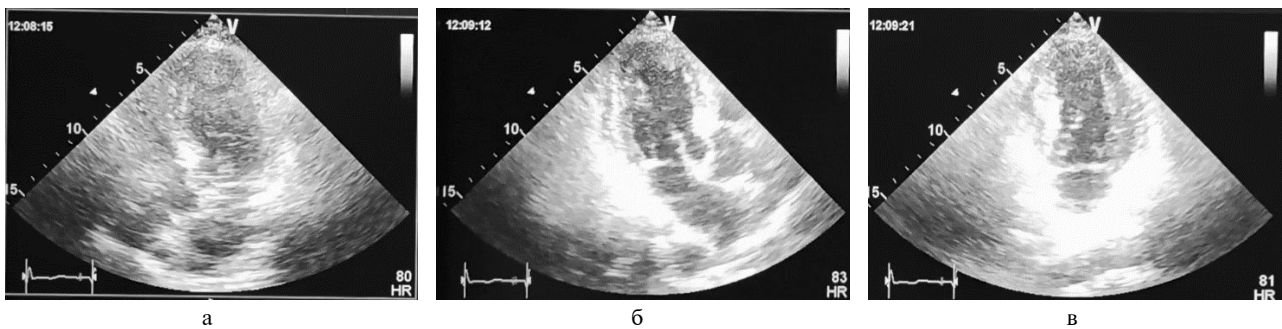


Рисунок 2: Проекції серця при ехокардіографії: (а) проекція № 1 – поздовжня вісь; (б) проекція № 2 – 2-камерна позиція; (в) проекція № 3 – 4-камерна позиція

Початкова матриця градацій сірого	Матриця після диференціювання	Диференційована матриця без негативних значень
53 55 56 61 67 67 65	2 1 5 6 0 -2	8 7 11 12 6 4
52 52 52 57 62 62 58	0 0 5 5 0 -4	6 6 11 11 6 2
60 60 59 60 61 58 54	0 -1 1 1 -3 -4	6 5 7 7 3 2
66 70 70 69 66 60 54	4 0 -1 -3 -6 -6	10 6 5 3 0 0
68 74 74 72 66 61 58	6 0 -2 -6 -5 -3	12 6 4 0 1 3
68 72 72 69 67 67 67	4 0 -3 -2 0 0	10 6 3 4 6 6
68 68 68 68 70 75 74	0 0 0 2 5 -1	6 6 6 8 11 5
67 68 68 69 71 74 75	1 0 1 2 3 1	7 6 7 8 9 7
65 68 71 73 73 73 73	3 3 2 0 0 0	9 9 8 6 6 6
66 72 76 76 74 73 74	6 4 0 -2 -1 1	12 10 6 4 5 7

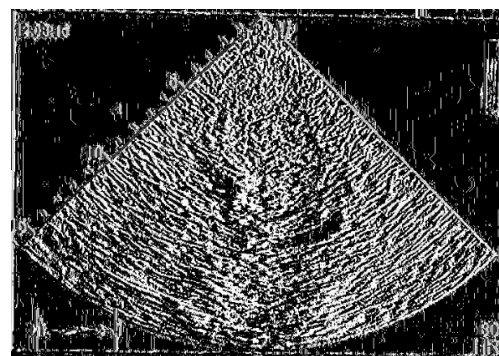
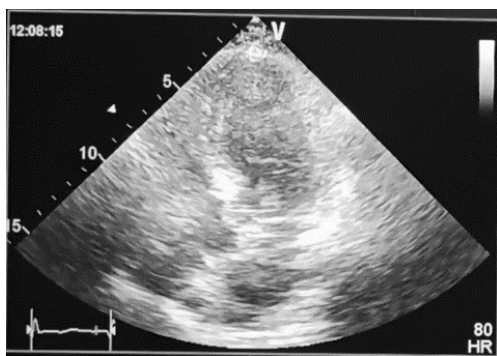


Рисунок 3: Горизонтальна диференціація: (а) схема горизонтальної диференціації, (б) оригінальне зображення, (в) зображення після диференціації



Вертикальна ж диференціація (рис. 4) виконується за формулами

$$g_{i,j}^{vert} = g_{i,j} - g_{i,j-1}, \quad g_{i,j}^{*vert} = g_{i,j}^{vert} + |g_{min}^{vert}|.$$

На основі отриманих після диференціації матриць були розраховані такі статистичні показники [9, 10]:

1. Середнє значення градацій сірого:

$$mean = \bar{g}_{i,j} = \frac{1}{n} \sum_1^n g_{i,j}.$$

2. Середньоквадратичне відхилення:

$$std = \sqrt{\frac{1}{n} \sum_1^n (g_{i,j} - mean)^2}.$$

3. Асиметрія:  $skew = \frac{\sum_1^n (g_{i,j} - mean)^3}{(n-1) \cdot std^3}.$

4. Експес:  $kurt = \frac{\sum_1^n (g_{i,j} - mean)^4}{(n-1) \cdot std^4}.$

5. Розмах:  $range = g_{max} - g_{min}.$

6. Медіана (median), перший квартиль (Q1), третій квартиль (Q3), 5-перцентиль (P5), 95-перцентиль (P95), міжквартильний розмах (IQR).

Кожна ознака п. 6 показана на рис. 5, виходячи з двовимірного розподілу частот, зображеного в лівій частині рисунка. Ці різниці були застосовані як додаткові інформативні ознаки для класифікації, оскільки є припущення, що при патології спостерігається більше патернів повторень градацій сірого.

Також ці статистичні показники були порашовані на основі матриць довжин однакових відтінків сірого (GLRLM) [11]. Ідеологія GLRLM показана на рис. 6. Матриця GLRLM підраховує кількість однорідних серій (1, 2 або 3 повтори) для кожного відтінку сірого. Статистичні ознаки, що були описані вище, розраховуються окремо для кожної серії.

Також із матриці GLRLM були отримані такі ознаки:

$$dif12 = AMo^1 - AMo^2,$$

$$dif13 = AMo^1 - AMo^3,$$

$$dif23 = AMo^2 - AMo^3,$$

де  $AMo^i$  – амплітуда моди нормалізованих частот матриці GLRLM за довжини  $i$ .

початкова матриця градацій сірого

53	55	56	61	67	67	65
52	52	52	57	62	62	58
60	60	59	60	61	58	54
66	70	70	69	66	60	54
68	74	74	72	66	61	58
68	72	72	69	67	67	67
68	68	68	68	70	75	74
67	68	68	69	71	74	75
65	68	71	73	73	73	73
66	72	76	76	74	73	74

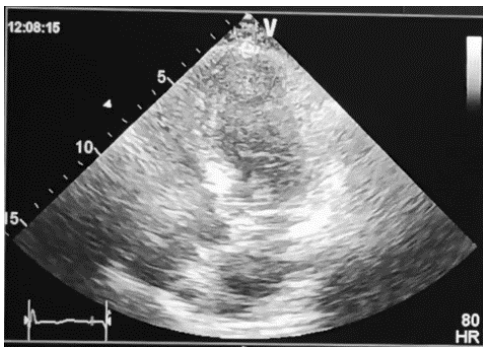
Матриця після диференціювання

-1	-3	-4	-4	-5	-5	-7
8	8	7	3	-1	-4	-4
6	10	11	9	5	2	0
2	4	4	3	0	1	4
0	-2	-2	-3	1	6	9
0	-4	-4	-1	3	8	7
-1	0	0	1	1	-1	1
-2	0	3	4	2	-1	-2
1	4	5	3	1	0	1

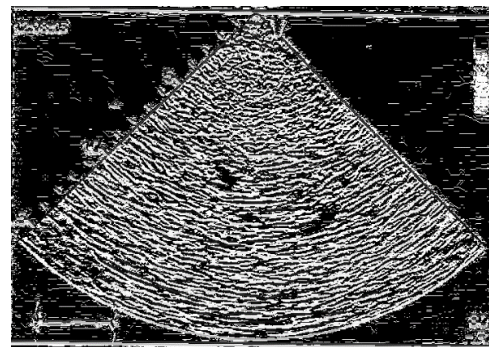
Диференційована матриця без негативних значень

6	4	3	3	2	2	0
15	15	14	10	6	3	3
13	17	18	16	12	9	7
9	11	11	10	7	8	11
7	5	5	4	8	13	16
7	3	3	6	10	15	14
6	7	7	8	8	6	8
5	7	10	11	9	6	5
8	11	12	10	8	7	8

а



б



в

Рисунок 4: Вертикальна диференціація: (а) схема вертикальної диференціації, (б) оригінальне зображення, (в) зображення після диференціації

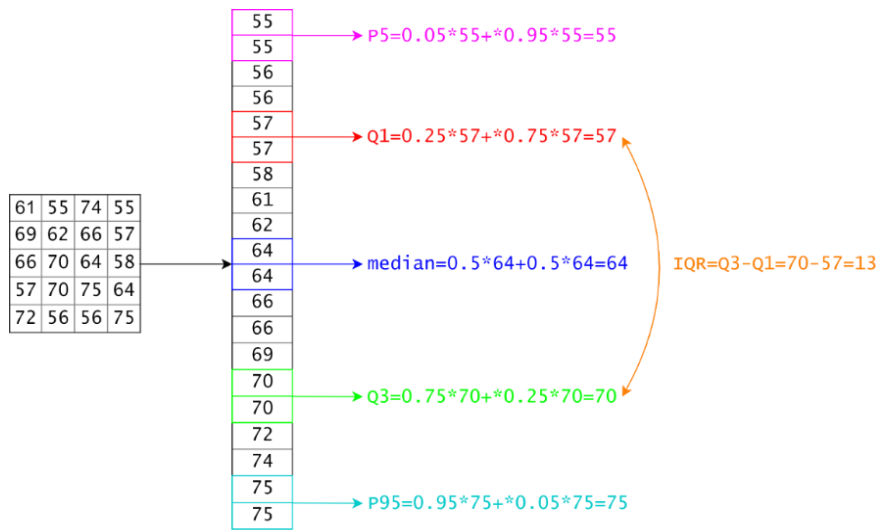


Рисунок 5: Квартили розподілу ознак, отриманих на основі текстурного аналізу

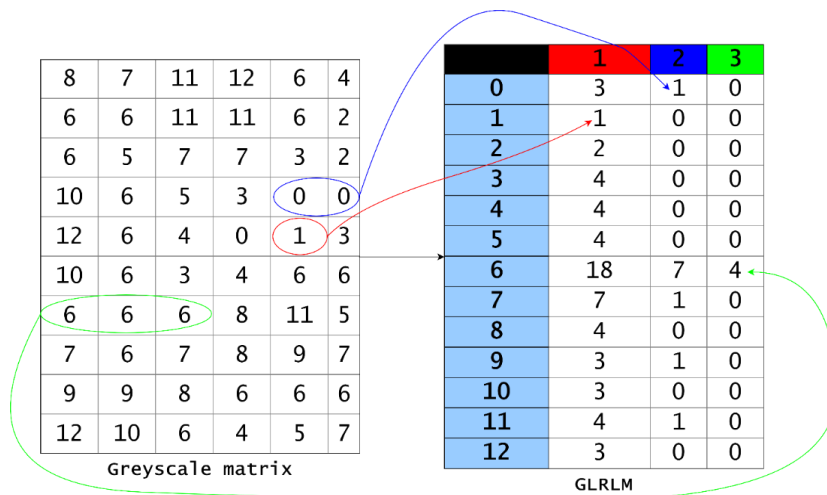


Рисунок 6: Схема матриці GLRLM

Розподіл різниць між амплітудами мод зображено на рис. 7. На рисунку червоними стрілками показано візуальну різницю між амплітудами мод різних довжин. Ці різниці були застосовані як додаткові інформативні ознаки для класифікації, оскільки при патології може спостерігатися більше або менше патернів повторень градацій сірого певних відтінків сірого.

Після отримання всіх необхідних ознак для побудови моделей класифікації було використано алгоритм випадкового лісу (Random Forest) [12–14]. Згідно з цим методом будується так званий ліс дерев, кожне з яких формує певне діагностичне рішення. Фінальне рішення щодо наявності або відсутності патології приймається методом голосування за більшістю рішень щодо приналежності досліджуваного об'єкта до певного класу (норма або патологія). Таким чином,

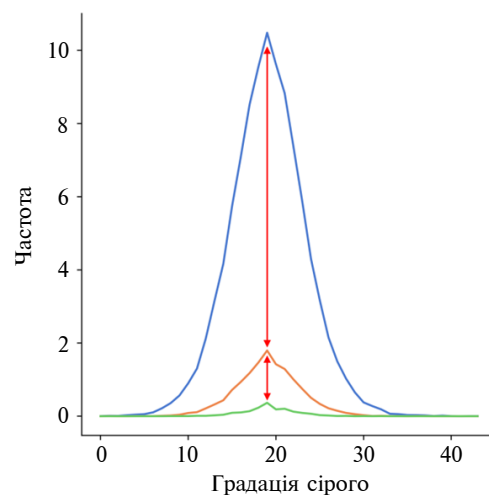


Рисунок 7: Різниці між амплітудами мод за результатами аналізу матриці GLRLM; довжина: — 1; — 2; — 3

реалізується підхід, який зветься “мудрість на товпу”. Рішення останнього зазвичай є точнішим за рішення за будь-яким із окремих дерев.

Перед застосуванням алгоритму вибірки випадковим чином були поділені на навчальні (70 %) і тестові (30 %). Класифікатори були оцінені за показниками точності, чутливості та специфічності (табл. 1). Спочатку класифікатори були побудовані для зображень, поділених на дози добутамінної проби.

Структура дерев випадкового лісу була досить складною і різнилась від однієї дози добутаміну до іншої. Приклад одного із дерев для класифікатора випадкового лісу подано на рис. 8.

На наступному етапі дослідження, для підвищення універсальності моделей на основі випадкового лісу, був побудований класифікатор норма–патологія для всіх зображень без розділення на дози добутаміну (табл. 2).

На рис. 9 продемонстрований приклад одного з дерев для класифікатора випадкового лісу, побудованого на другому етапі дослідження.

## Обговорення

Проведене дослідження дало 15 різних моделей випадкового лісу. Як видно з табл. 1, в якій при моделюванні враховувались дози добутаміну (що дало в сумі 12 моделей), були отримані високі характеристики точності, чутливості та специфічності. Аналогічним чином можна сказати про три моделі випадкового лісу (особливістю яких є неврахування дози добутаміну), якщо дивитись на табл. 2. Різниця в тому, що порівняно з деревами випадкового лісу, побудованими з поділенням вибірки на дози добутаміну, дерева, побудовані без урахування цих доз, мали більш комплексну структуру, проте моделі набули більшої універсальності.

Застосування випадкового лісу для класифікації даних СТЕ є не новим. У роботі [15] цей алгоритм був порівняний з іншими методами машинного навчання, такими як: дерево прийняття рішень, XGBoost, наївний байєсівський класифікатор, метод  $k$ -найближчих сусідів

**Таблиця 1:** Результати тестування (оцінки якості) отриманих моделей Random Forest (із діленням на дози добутаміну)

Позиція ехокардіограми	Навчальна вибірка (70 %)			Тестова вибірка (30 %)		
	Точність, %	Чутливість	Специфічність	Точність, %	Чутливість	Специфічність
00 rest						
4-cam	100	1	1	99,1	0,981	0,996
2-cam	100	1	1	99,3	0,985	1
long	100	1	1	99,6	0,987	1
10 mcg						
4-cam	100	1	1	98,6	0,975	0,996
2-cam	100	1	1	99	0,982	1
long	100	1	1	99,6	0,995	0,996
20 mcg						
4-cam	100	1	1	98,2	0,959	0,996
2-cam	100	1	1	98,7	0,965	1
long	100	1	1	98,2	0,921	1
40 mcg						
4-cam	100	1	1	97	0,969	0,971
2-cam	100	1	1	99,6	0,989	1
long	100	1	1	99,1	0,975	1

**Таблиця 2:** Результати тестування моделей випадкового лісу без урахування дози добутаміну

Позиція ехокардіограми	Навчальна вибірка (70%)			Тестова вибірка (30%)		
	Точність, %	Чутливість	Специфічність	Точність, %	Чутливість	Специфічність
4-cam	99,8	0,999	0,998	97	0,95	0,982
2-cam	100	1	1	97,6	0,963	0,987
long	100	1	1	96,8	0,915	0,994

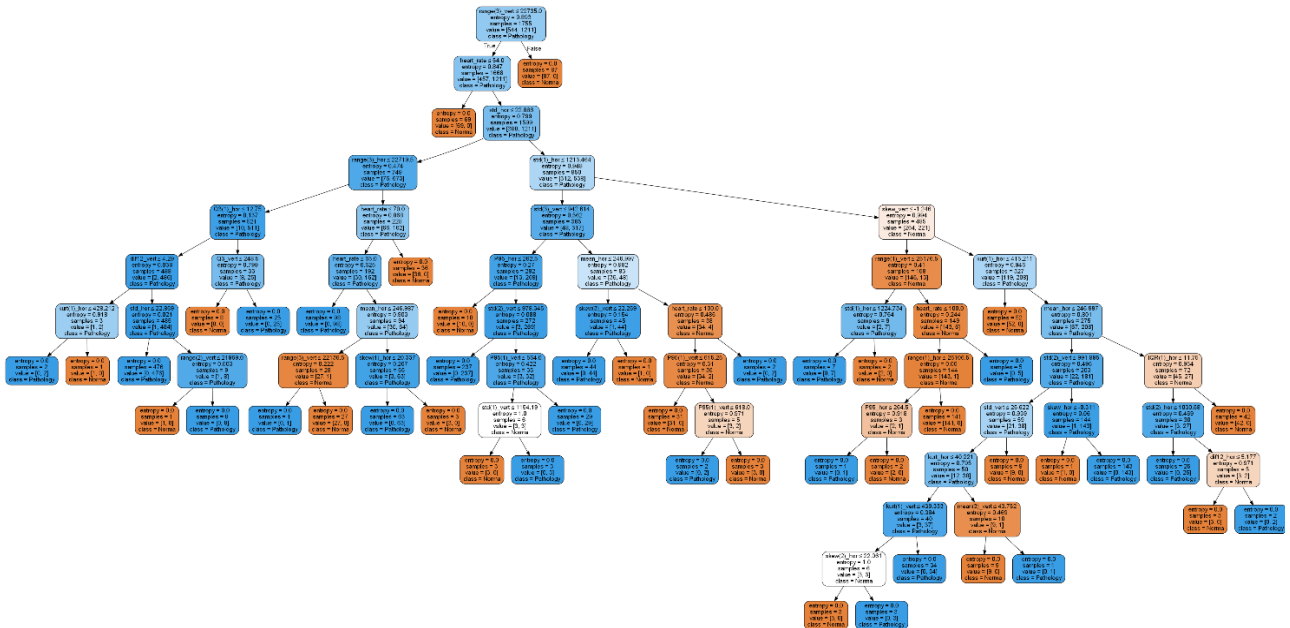


Рисунок 8: Дерево випадкового лісу класифікації для позиції “long” з нульовою дозою добутаміну

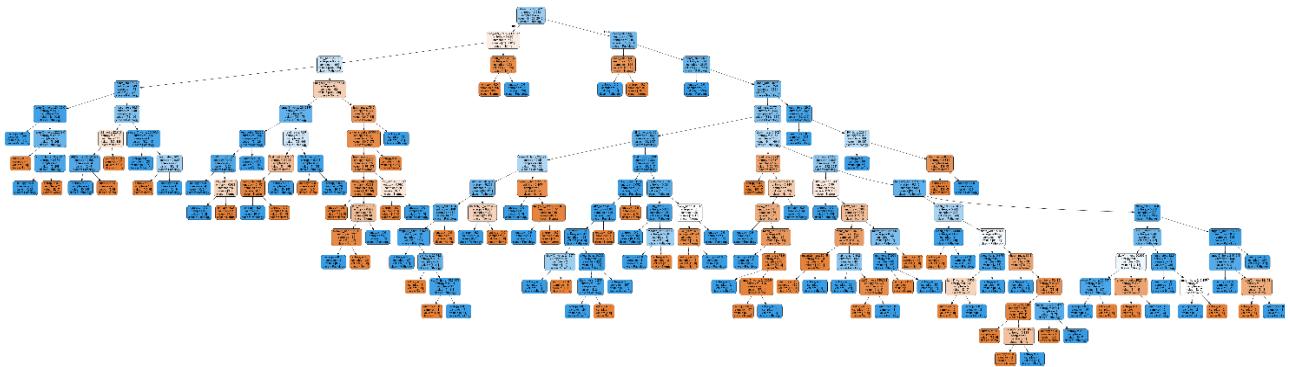


Рисунок 9: Дерево випадкового лісу класифікації для позиції “long” без урахування доз добутаміну

та логістична регресія. Модель випадкового лісу дала найкращий результат (90,5 % точності) на тестовій вибірці у задачі розпізнавання патології серця за даними СТЕ. Автори [16] використовували підходи машинного навчання для розпізнавання захворювання міокарда, де загальна модель видала 87 % чутливості та 82 % специфічності. Відмінність між дослідженнями у вказаних роботах і поточною роботою в тому, що в нашому дослідженні використовувались лише текстурні ознаки без урахування метаданих. Таким чином, можна сказати, що для отриманих моделей достатньо мати на вході лише зображення СТЕ, через аналіз якого з високою ймовірністю буде розпізнано ІХС.

Отримані нами результати свідчать, що метод побудови випадкового лісу за характеристиками текстури зображень є потужним засо-

бом створення діагностичних моделей. Це забезпечує їх високі прогностичні характеристики та можливість використання в автоматизованих системах підтримки прийняття медичних рішень, проте утруднює їх змістовну медичну інтерпретацію. Їх надто важка для сприйняття візуалізація (що можна частково спостерігати на рис. 8, 9) наводить на думку про те, що подальші роботи слід спрямувати на забезпечення аналітичними методами змістовної інтерпретації випадкового лісу для більш коректної діагностики.

Високий інтерес становить спроба застосування при побудові випадкового лісу конкретних характеристик СТЕ. У цьому випадку є сподівання отримати інформативні змістовні характеристики розвитку патологічних станів.

## Висновки

Застосування текстурних ознак відеопотоку даних ультразвукових обстежень методом СТЕ дає можливість побудувати високоточні діагностичні алгоритми для виявлення порушень коронарного кровообігу при ІХС.

Алгоритм побудови випадкового лісу є потужним і високоефективним засобом для роз-

робки алгоритмів діагностики ІХС за даними СТЕ. Точність діагностики за даними тестових вибірок становить 96,8–97,6 %.

Розроблені діагностичні алгоритми можуть бути застосовані при використанні більш простого ультразвукового діагностичного обладнання, яке працює в В-режимі, але не оснащено спекл-трекінг технологією.

## References

- [1] Voigt JU, Pedrizzetti G, Lysyansky P, Marwick TH, Houle H, Baumann R, et al. Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging. *Eur Heart J Cardiovasc Imaging*. 2015;16(1):1-11. DOI: 10.1093/ehjci/jeu184
- [2] Lazoryshynets VV, Kovalenko VM, Rudenko AV, Ivaniv YA, Beshlyaga VM, Potashev SV, et al. Definition for a common standard for 2D speckle-tracking echocardiography (the Association of cardiovascular surgeons of Ukraine and the Ukrainian society of cardiology working group draft consensus). *Cardiol Card Surg Contin Prof Dev*. 2019;2:105-29. DOI: 10.30702/ccs.201905.02.2dst105129
- [3] Contreras-Ojeda SL, Sierra-Pardo C, Dominguez-Jimenez JA, Lopez-Bueno J, Contreras-Ortiz SH. Texture analysis of ultrasound images for pneumonia detection in pediatric patients. In: *Proceedings of 22nd Symposium on Image, Signal Processing and Artificial Vision; 2019; Institute of Electrical and Electronics Engineers Inc.* DOI: 10.1109/STSIVA.2019.8730238
- [4] Molinari F, Caresio C, Acharya UR, Mookiah MRK, Minetto MA. Advances in quantitative muscle ultrasonography using texture analysis of ultrasound images. *Ultrasound Med Biol*. 2015 Sep 1;41(9):2520-32. DOI: 10.1016/j.ultrasmedbio.2015.04.021
- [5] Kayal EB, Kandasamy D, Khare K, Bakhshi S, Sharma R, Mehndiratta A. Texture analysis for chemotherapy response evaluation in osteosarcoma using MR imaging. *NMR Biomed*. 2021 Feb 1;34(2):17. DOI: 10.1002/nbm.4426
- [6] Boulkenafet Z, Komulainen J, Hadid A. Face spoofing detection using colour texture analysis. *IEEE Trans Inf Forensics Secur*. 2016;11(8):1818-30. DOI: 10.1109/tifs.2016.2555286
- [7] Kavitha P, Prabakaran S. Designing a feature vector for statistical texture analysis of brain tumor. *Int J Eng Adv Technol*. 2019;8(5):1228-30.
- [8] Nastenko Ie, Dykan I, Tarasiuk B, Pavlov V, Nosovets O, Babenko V, Kruglyi V, Soloduschenko V. GMDH classification of liver states in diffuse diseases based on statistical features of the ultrasound images texture. *Induktivne Modellivannia Skladnyh System*. 2020;11:54-66.
- [9] Zayed N, Elnemr HA. Statistical analysis of haralick texture features to discriminate lung abnormalities. *Int J Biomed Imaging*. 2015;2015:267807. DOI: 10.1155/2015/267807
- [10] Vrbik I, Van Nest SJ, Meksjarun P, Loepky J, Brolo A, Lum JJ, et al. Haralick texture feature analysis for quantifying radiation response heterogeneity in murine models observed using Raman spectroscopic mapping. *PLoS One*. 2019;14(2):e0212225. DOI: 10.1371/journal.pone.0212225
- [11] Rini Novitasari DC, Lubab A, Sawiji A, Asyhar AH. Application of feature extraction for breast cancer using one order statistic, glcm, glrlm, and gldm. *Adv Sci Technol Eng Syst*. 2019;4(4):115-20. DOI: 10.25046/aj040413
- [12] Ali J, Khan R, Ahmad N, Maqsood I. Random forests and decision trees. *Int J Comp Sci Iss*. 2012;9(5):272-8.
- [13] Sharma H, Kumar S. A Survey on decision tree algorithms of classification in data mining. *Int J Sci Res*. 2016 Apr 5;5(4):2094-7. DOI: 10.21275/v5i4.nov162954
- [14] Fawagreh K, Gaber MM, Elyan E. Random forests: From early developments to recent advancements. *Syst Sci Control Eng*. 2014;2(1):602-9. DOI: 10.1080/21642583.2014.956265
- [15] Zhong J, Liu P, Li S, Huang X, Zhang Q, Huang J, et al. A comparison of three-dimensional speckle tracking echocardiography parameters in predicting left ventricular remodeling. *J Healthc Eng*. 2020 Aug 1;2020:1-9. DOI: 10.1155/2020/8847144
- [16] Narula S, Shameer K, Salem OAM, Dudley JT, Sengupta PP. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. *J Am Coll Cardiol*. 2016 Nov;68(21):2287-95. DOI: 10.1016/j.jacc.2016.08.062



Е.А. Настенко<sup>1,2</sup>, В.Б. Максименко<sup>1,2</sup>, С.В. Поташев<sup>2</sup>, В.А. Павлов<sup>1</sup>, В.О. Бабенко<sup>1</sup>, С.В. Рысин<sup>1,2</sup>, А.В. Матвийчук<sup>1</sup>, В.В. Лазоришинец<sup>2</sup>

<sup>1</sup>КПИ им. Игоря Сикорского, Киев, Украина

<sup>2</sup>Национальный институт сердечно-сосудистой хирургии им. Н.М. Амосова, Киев, Украина

## ПОСТРОЕНИЕ АЛГОРИТМА СЛУЧАЙНОГО ЛЕСА ДЛЯ ДИАГНОСТИКИ ИШЕМИЧЕСКОЙ БОЛЕЗНИ СЕРДЦА ПО ПОТОКАМ ВИДЕОДАНЫХ ЭХОКАРДИОГРАФИИ

**Проблематика.** Исследования последних лет показывают, что сердечно-сосудистые заболевания, включая ишемическую болезнь сердца, являются основной причиной смертности и одним из основных факторов инвалидности во всем мире. Выявление случаев данного типа заболеваний за последние 30 лет увеличилось с 271 до 523 млн, а количество смертей – с 12,1 до 18,6 млн. Сердечно-сосудистые заболевания являются главной причиной смертности населения Украины, и по этому показателю страна остается одним из мировых лидеров. Ишемическая болезнь сердца является ведущим фактором потери здоровья Украины, и для ее своевременного выявления все больше используют современные методы диагностики, включая алгоритмы машинного обучения.

**Цель.** По данным спэкл-трекинг эхокардиографии методом случайного леса построить классификационные алгоритмы диагностики нарушений кинематики сокращений левого желудочка сердца у больных ишемической болезнью сердца в условиях состояния покоя и при применении эхостресс-теста с добутаминовой пробой.

**Методика реализации.** Методом спэкл-трекинг эхокардиографии были обследованы 40 пациентов с наличием ишемической болезни сердца и 16, у которых патология сердца не обнаружена. Эхокардиография регистрировалась в В-режиме в трех позициях: по длинной оси, в 4- и 2-камерной позициях. Всего было использовано 6245 кадров видеопотока: 1871 – без нарушений сердечной деятельности и 4374 – при наличии патологии при обследовании. 56 пациентов (2509 кадров видеоданных) обследованы без применения добутаминовой пробы и 38 пациентов (3736 кадров видеоданных) – с применением эхостресс-теста с добутаминовой пробой, если нарушений в состоянии покоя обнаружено не было. Под наблюдением анестезиолога применялись дозы добутамина 10, 20 и 40 мкг. В качестве информативных признаков были использованы данные текстурного анализа изображений. Для построения алгоритма выявления ишемической болезни сердца был применен алгоритм случайного леса.

**Результаты.** На первом этапе исследования были построены диагностические алгоритмы норма–патология для состояния покоя и доз добутамина 10, 20 и 40 мкг. Перед применением алгоритма выборки случайным образом были разделены на учебные (70 %) и тестовые (30 %). Классификаторы были оценены по показателям точности, чувствительности и специфичности. По данным тестовых выборок точность диагностических выводов варьировалась от 97 до 99 %. На втором этапе исследования, для повышения универсальности моделей, был построен классификатор для всех изображений, без разделения на дозы добутамина. Точность по тестовым выборкам также находилась в пределах от 96,6 до 97,8 %. Для построения диагностических алгоритмов методом случайного леса были использованы данные текстурного анализа изображений.

**Выводы.** Получены высокоточные модели классификации посредством использования алгоритма случайного леса. Разработанные модели могут быть применены для анализа эхокардиограмм, полученных в В-режиме на оборудовании, которое не оснащено технологией спэкл-трекинг.

**Ключевые слова:** случайный лес; диагностика; спэкл-трекинг эхокардиография; эхостресс-тест с добутамином; ишемическая болезнь сердца.

Ie.A. Nastenko<sup>1,2</sup>, V.B. Maksymenko<sup>1,2</sup>, S.V. Potashev<sup>2</sup>, V.A. Pavlov<sup>1</sup>, V.O. Babenko<sup>1</sup>, S.V. Rysin<sup>1,2</sup>, O.V. Matviichuk<sup>1</sup>, V.V. Lazoryshinets<sup>2</sup>

<sup>1</sup>Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

<sup>2</sup>Amosov National Institute of Cardiovascular Surgery, Kyiv, Ukraine

## RANDOM FOREST ALGORITHM CONSTRUCTION FOR THE DIAGNOSIS OF CORONARY HEART DISEASE BASED ON ECHOCARDIOGRAPHY VIDEO DATA STREAMS

**Background.** Recent studies show that cardiovascular diseases, including coronary heart disease, are the leading causes of death and one of the main factors of disability worldwide. The detection of cases of this type of disease over the past 30 years has increased from 271 million to 523 million and the number of deaths – from 12.1 million to 18.6 million. Cardiovascular diseases are the main cause of death among the population of Ukraine and, according to this indicator, the country remains one of the world leaders. Coronary heart disease is the leading factor in the loss of health in Ukraine and modern diagnostic methods, including machine learning algorithms, are increasingly being used for timely detection.

**Objective.** According to the data of speckle-tracking echocardiography using the random forest method, construct classification algorithms for diagnosing violations of the kinematics of left ventricular contractions in patients with coronary heart disease at rest, and when using an echostress test with a dobutamine test.

**Methods.** Speckle-tracking echocardiography was used to examine 40 patients with coronary heart disease and 16 in whom no cardiac pathology was found. Echocardiography was recorded in B mode in three positions: along the long axis, in 4-chamber, and 2-chamber positions. In total, 6245 frames of the video stream were used: 1871 – without cardiac abnormalities, and 4374 – in the presence of pathology during the examination. 56 patients (2509 frames of video data) were examined without the use of a dobutamine test and 38 patients (3736 frames of video data) – using an echostress test with a dobutamine test if no disturbances were found at rest. Dobutamine doses of 10, 20, and 40 mcg were administered under the supervision of an anesthesiologist. The data of texture analysis of images were used as informative features. To build an algorithm for detecting coronary heart disease the random forest algorithm was applied.

**Results.** At the first stage of the study, the diagnostic algorithms norma–pathology for the state of rest and dobutamine doses of 10, 20, and 40 mcg were constructed. Before applying the algorithm the samples were randomly divided into training (70%) and test (30%). The classifiers were evaluated for accuracy, sensitivity, and specificity. According to the test samples, the accuracy of diagnostic conclusions varied from 97 to 99%. At the second stage of the study, to increase the versatility of the models, the classifier was built for all images, without dividing them into dobutamine doses. The accuracy for the test samples also ranged from 96.6 to 97.8%. To construct diagnostic algorithms by the random forest method the data of texture analysis of images were used.

**Conclusions.** High-precision classification models were obtained using the random forest algorithm. The developed models can be applied to the analysis of echocardiograms obtained in B mode on equipment that is not equipped with the speckle tracking technology.

**Keywords:** random forest; diagnostics; speckle-tracking echocardiography; echostress test with dobutamine; coronary heart disease.