

А. С. Сверстюк^{1, 2}
Л. Є. Мосій¹

БАГАТОФАКТОРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ ДЛЯ ПРОГНОЗУВАННЯ КАРДІОЛОГІЧНОГО ДІАГНОЗУ НА ОСНОВІ ФУНКЦІЇ АМПЛІТУДНОЇ ВАРІАБЕЛЬНОСТІ

¹Тернопільський національний технічний університет імені І. Пулюя;

²Тернопільський національний медичний університет імені І. Я. Горбачевського

Запропоновано багатофакторну регресійну модель для автоматичного прогнозування кардіологічного діагнозу на основі статистичних характеристик функції амплітудної варіабельності електрокардіосигналів. Модель використовує сім статистично значущих предикторів (середнє арифметичне, медіану, моду, стандартне відхилення, вибірккову дисперсію, ексцес та асиметрію), виділених методом покрокової регресії з прямим відбором змінних з початкового набору 13 показників. Критерієм включення предикторів встановлено рівень значущості $p < 0,05$. У моделі для прогнозування кардіологічного діагнозу на основі функції амплітудної варіабельності враховані вагові коефіцієнти кожного предиктора, визначені методом найменших квадратів. Найбільший внесок у формування діагностичного висновку мають показники центральної тенденції ($\beta = 201,78$ для середнього та $\beta = 68,69$ для медіани).

Модель забезпечує ефективну диференціацію між трьома клінічними станами: умовною нормою, порушеннями ритму серця у вигляді екстрасистолії та морфологічними патологіями проведення збудження, представленими неповною блокадою лівої ніжки пучка Гіса, з коефіцієнтом детермінації Нейджелкерка $R^2 = 0,991$. Для норми характерні мінімальні значення амплітудної варіабельності (математичне сподівання $0,00003 \dots 0,00064$ мВ), тоді як у разі екстрасистолії спостерігається зростання на декілька порядків. Валідація моделі на репрезентативній вибірці з 204 електрокардіосигналів (102 норма, 51 екстрасистолія, 51 блокада) підтвердила її високу статистичну значущість (F-критерій Фішера 953,93, $p < 0,001$) та відповідність основним припущенням регресійного аналізу. Аналіз залишкових відхилень продемонстрував їхній нормальний розподіл та гомоскедастичність, що свідчить про адекватність моделі.

Запропонований підхід поєднує переваги високої інтерпретованості класичних статистичних методів з інноваційним використанням функції амплітудної варіабельності для комплексного аналізу морфологічних і ритмічних ознак кардіосигналів. Практичне значення полягає у створенні математичного інструментарію для автоматизованих систем діагностики серцево-судинних захворювань та систем підтримки ухвалення клінічних рішень.

Ключові слова: функція амплітудної варіабельності, багатофакторна регресійна модель, електрокардіосигнал, кардіологічна діагностика, покрокова регресія, екстрасистолія, блокада лівої ніжки пучка Гіса, коефіцієнт прогнозування CDPCAVF, статистичні предиктори, автоматична класифікація серцевих патологій.

Вступ

Вивчення варіабельності амплітуди за допомогою регресійних моделей привернуло значну увагу в різних наукових та інженерних дисциплінах, що спонукало до формування потужної наукової бази, яка досліджує різні аспекти цієї проблематики. Сучасний стан досліджень характеризується наявністю двох основних підходів до моделювання часових рядів: традиційних статистичних методів та новітніх методів машинного навчання, кожен з яких має свої переваги та обмеження.

Статистичні моделі пропонують високу інтерпретованість результатів, проте вони можуть мати труднощі зі складними нелінійними даними часових рядів [1]. З іншого боку, машинне навчання продемонструвало вражаючі можливості в різних застосуваннях прогнозування, але воно характеризується кількома недоліками, такими як обчислювальна витратність та необхідність спеціалізованої експертизи для впровадження та інтерпретації [2], [3]. Застосування методів машинного нав-

чання до проблем прогнозування набуло значного імпульсу, що зумовлено доступністю великих наборів даних та прогресом в обчислювальній потужності [4]. Ці моделі пропонують перевагу подолання певних обмежень, пов'язаних з класичними методами аналізу часових рядів, такими як AR, MA або ARIMA [5].

Регресійні моделі стали потужним інструментом для аналізу та прогнозування даних часових рядів у різних галузях, пропонуючи гнучку структуру для фіксації часових залежностей та динамічних патернів [6]. У контексті аналізу варіабельності амплітуди регресійні моделі надають засоби для характеристики флуктуацій та патернів, властивих амплітудам сигналів, що дозволяє отримати уявлення про основні процеси та потенційні аномалії [7].

Здатність цих моделей адаптуватися до різноманітних характеристик даних у поєднанні з їхньою обчислювальною ефективністю сприяла їх широкому впровадженню в галузях від фінансів та економіки до обробки сигналів і кліматології. Використовуючи притаманну структуру автокореляції даних часових рядів, регресійні моделі пропонують економний, але ефективний підхід до моделювання варіабельності амплітуди, полегшуючи такі завдання, як прогнозування, виявлення аномалій та аналіз точок зміни [8]. Для моделювання багатовимірних часових рядів без припущень про теоретичну структуру традиційно використовуються векторні регресії, які добре підходять для лінійної та незалежної від стану еволюції [9], [10].

Хоча регресійні моделі традиційно використовувалися для лінійного прогнозування часових рядів, їхні обмеження у фіксації нелінійних патернів мотивували розробку складніших методів [11]. Базова регресійна модель передбачає лінійний зв'язок між минулими та теперішніми значеннями, що може не відповідати дійсності для складних систем, які демонструють нелінійну поведінку. Щоб подолати це обмеження, дослідники вивчали різні розширення регресійних моделей, включаючи нелінійні регресійні моделі, порогові регресійні моделі та гібридні підходи, які поєднують регресійні моделі з методами машинного навчання [12]. Ці передові методи пропонують розширені можливості для фіксації складних залежностей та нелінійної динаміки в даних часових рядів, розширюючи застосовність регресійного моделювання до ширшого кола реальних проблем.

Проте, незважаючи на появу складніших методів прогнозування, регресійні моделі зберігають свою актуальність завдяки своїй простоті, інтерпретованості та обчислювальній ефективності. Їхня лінійна структура дозволяє здійснювати пряму оцінку параметрів та інтерпретацію моделі, надаючи цінну інформацію про основну динаміку часового ряду. До того ж, регресійні моделі можна легко інтегрувати з іншими статистичними методами та методами машинного навчання, утворюючи гібридні підходи, які використовують сильні сторони різних методологій [13].

Застосування регресійних моделей до аналізу варіабельності амплітуди представляє унікальні виклики та можливості. Варіабельність амплітуди часто демонструє нестационарну поведінку, що означає, що її статистичні властивості змінюються з часом. Ця нестационарність може виникати з різних джерел, таких як зовнішні впливи, зміни режимів або внутрішня динаміка досліджуваної системи.

Для ефективного моделювання нестационарної варіабельності амплітуди важливо застосовувати адаптивні методи, які можуть коригувати параметри моделі з часом. Цього можна досягти за допомогою таких методів, як оцінка ковзного вікна, рекурсивна оцінка або моделювання простору станів. Також, важливо враховувати наявність викидів або екстремальних значень, які можуть непропорційно впливати на параметри моделі та спотворювати аналіз. Для пом'якшення впливу викидів можуть застосовуватися робастні методи оцінювання, що забезпечують надійність результатів.

У сфері аналізу часових рядів наявність стохастичних регресорів, що демонструють одночасну екзогенність щодо залишкового шуму, разом з автокореляцією та гетероскедастичністю, створює суттєві виклики [14]. Традиційні статистичні моделі, такі як регресійна інтегрована модель ковзного середнього (ARIMA) та сезонна ARIMA, знайшли широке застосування в різних галузях, демонструючи багатообіцяючу продуктивність у фіксації лінійних трендів та сезонності [15]. Проте багато реальних часових рядів демонструють як лінійні, так і нелінійні патерни, що вимагає розробки гібридних моделей, які можуть ефективно фіксувати обидва типи залежностей [11].

Відсутні дані (поширена проблема в аналізі часових рядів) можуть бути отримані за допомогою одновимірних методів імпутації завдяки їхній адаптивності, що забезпечує надійність аналізу [16]. Альтернативні підходи включають використання методів оптимізації, які мінімізують відповідний критерій, розглядаючи залишки як вільні змінні, що дозволяє точну інтерполяцію відсутніх значень [17].

Щоб подолати обмеження традиційних регресійних моделей, методи машинного навчання стали потужними інструментами для прогнозування часових рядів [18]. Зокрема, рекурентні нейронні

мережі, особливо мережі довгої короткочасної пам'яті (LSTM), продемонстрували кращу продуктивність порівняно з регресійними інтегрованими моделями ковзного середнього у фіксації довгострокових залежностей. Ефективність нейронних мереж у прогнозуванні часових рядів впливає з їхньої здатності вивчати складні нелінійні зв'язки між минулими та теперішніми значеннями, адаптуючись до еволюційних патернів у даних.

Застосування алгоритмів машинного навчання та глибокого навчання — це значний прогрес у вирішенні проблем прогнозування в часових рядах, що дає точніші результати ніж звичайне моделювання на основі регресії [19]. Для обробки нелінійних часових рядів нейронні мережі показали перспективність, хоча вони є складнішими та важчими для інтерпретації [20].

Моделі глибокого навчання, такі як рекурентні нейронні мережі, досягли вражаючого успіху в фіксації складних часових патернів. Методи глибокого навчання стали популярнішими, оскільки вони можуть вивчати складні представлення даних, інтегруючи спеціалізовані архітектурні припущення, які відображають тонкощі базових наборів даних [21]. До того ж моделі машинного навчання, зокрема методи на основі дерев та підходи глибокого навчання, набули популярності завдяки своїй здатності обробляти складні нелінійні залежності в даних [22].

Для систематизації розглянутих підходів до аналізу варіабельності амплітуди подамо їхню порівняльну характеристику (табл.).

Порівняльна характеристика методів аналізу варіабельності амплітуди

Метод	Переваги	Недоліки	Сфера застосування	Обчислювальна складність
Класичні AR моделі	висока інтерпретованість; теоретична обґрунтованість; швидка оцінка параметрів	лише лінійні залежності; припущення про стаціонарність; обмежена точність	стаціонарні процеси з лінійною динамікою	низька $O(n)$
ARIMA/SARIMA	робота з нестационарними рядами; врахування сезонності; широка підтримка в ПЗ	складність вибору параметрів; лінійність моделі; чутливість до викидів	економічні та фінансові часові ряди	середня $O(n^2)$
GARCH моделі	моделювання змінної волатильності; врахування кластеризації волатильності	складність оцінювання; нелінійність лише в дисперсії	фінансові ринки, аналіз ризиків	середня $O(n^2)$
Нейронні мережі (RNN/LSTM)	нелінійні залежності; довгострокова пам'ять; висока точність	«чорна скринька»; потреба в великих даних; ризик перенавчання	складні нелінійні системи	висока $O(n^3)$
Гібридні моделі	поєднання переваг; адаптивність; покращена точність	складність реалізації; важкість інтерпретації; високі вимоги до ресурсів	універсальне застосування	висока $O(n^3)$
Ансамблеві методи	стійкість прогнозів; зменшення дисперсії; робастність	обчислювальна витратність; складність налаштування	критичні застосування з високими вимогами до точності	дуже висока $O(n^4)$

Для вибору моделі важливо проводити ретельне тестування та валідацію для забезпечення її надійності та здатності до узагальнення. Ансамблеві методи, які поєднують кілька окремих моделей, також можуть підвищити точність прогнозування, зменшуючи дисперсію та зміщення, а пеналізовані регресії можуть запобігти перенавчанню [23]. Точність моделей часових рядів має першорядне значення, проте окремі моделі часто мають труднощі з продуктивністю на різних типах даних часових рядів [24].

Різні часові ряди можуть демонструвати різні ступені лінійності, нелінійності та сезонності, що може вплинути на продуктивність моделей прогнозування. Для підвищення ефективності прогнозування ефективним може бути поєднання лінійних і нелінійних моделей [25]. Використовуючи сильні сторони обох типів моделей, можна фіксувати основні патерни різних часових рядів з підвищеною точністю [26]. Прогнозування з використанням ансамблю кількох методів часто розглядається як оптимальний компроміс [27]. Моделі ансамблевого прогнозування підвищують різноманітність у прогностичних системах, дозволяючи всебічно отримувати цінну інформацію з прогнозів окремих моделей та зміцнюючи стабільність результатів у непередбачуваних обставинах. Ансамблеві методи послідовно підвищують ефективність прогнозування, особливо в імовірнісних умовах [28].

Ретельне врахування попередньої обробки даних, вибору моделі та налаштування гіперпараметрів є критично важливим для досягнення оптимальних результатів прогнозування з використанням методів машинного навчання [29]. Проте ці моделі часто вимагають значних обчислювальних ресурсів та експертизи для ефективної реалізації та інтерпретації [30].

Значний внесок у розвиток методів аналізу кардіосигналів та діагностики серцево-судинних захворювань зробили українські дослідники. І. Чайковський та співавтори [31] запропонували використання штучного інтелекту для моніторингу та корекції функціонального стану на основі ЕКС з використанням біонічного підходу, що базується на особливостях людської зорової системи. Їхні алгоритми сегментації з змінною роздільною здатністю демонструють переваги порівняно з класичними методами обробки циклічних сигналів та бінарних контурів зображень. С. М. Коваль та колектив авторів [32] розробили метод класифікації нестационарних серцевих сигналів на основі їх спектральних та ймовірнісних властивостей, використовуючи ймовірнісну модель серцевого сигналу як випадкового та апіорі нестационарного процесу. Запропоновані параметри нестационарності продемонстрували високу чутливість та інформативну значущість у статистичній класифікації серцевих сигналів. С. Лупенко та Я. Литвиненко [33] опрацьовували біосигнали на основі моделі циклічних ритмічно пов'язаних випадкових процесів, запропонували використання діагностичних ознак у вигляді коефіцієнтів декомпозиції статистичних оцінок біосигналів. У роботі було розроблено програмний комплекс для обробки та моделювання синхронно зареєстрованих кардіосигналів з використанням моделей та методів теорії циклічного функціонального зв'язку. Й. Лецишин та співавтори [34] побудували багатокомпонентну модель точки зміни варіабельності серцевого ритму, що враховує її складну структуру та відображає динаміку змін статистичних характеристик у часі. Модель виділяє множину кусково-стационарних та кусково-періодичних випадкових процесів у загальному класі нестационарних випадкових процесів.

Фундаментальні дослідження Я. Драгана та співавторів [35], [36] щодо обґрунтування структури систем дистанційної діагностики адаптаційних резервів серця та розробки математичного й алгоритмічно-програмного забезпечення комп'ютерних засобів статистичного опрацювання коливань (ритмічних процесів). Ці роботи заклали теоретичні основи для створення автоматизованих систем кардіологічної діагностики на базі статистичного аналізу біосигналів.

Представлені дослідження українських науковців демонструють комплексний підхід до вирішення проблем аналізу кардіосигналів, поєднуючи класичні статистичні методи з сучасними технологіями штучного інтелекту та біонічними принципами обробки інформації.

Метою роботи є розробка та валідація багатофакторної регресійної моделі для автоматичного прогнозування кардіологічного діагнозу на основі статистичних характеристик функції амплітудної варіабельності (ФАВ) електрокардіосигналів (ЕКС), що забезпечить об'єктивну кількісну оцінку функціонального стану серцево-судинної системи (ССС) та диференціацію між нормою та патологічними змінами серцевого ритму.

Результати дослідження

У роботі [37] запропонована ФАВ для комплексного аналізу морфологічних і ритмічних діагностичних ознак кардіосигналів. Попередні дослідження продемонстрували перспективність застосування цієї функції для комплексного аналізу морфологічних і ритмічних діагностичних ознак ЕКС. Зокрема, встановлено суттєві відмінності статистичних характеристик ФАВ між станами умовної норми та патологічними змінами серцевого ритму.

Для пацієнтів з умовною нормою характерна мінімальна амплітудна варіабельність: математичне сподівання знаходиться в межах 0,00003...0,00064 мВ, дисперсія становить 0,00010...0,00022 мВ², варіаційний розмах не перевищує 0,048...0,082 мВ. У разі екстрасистолії спостерігається статистично значуще зростання всіх досліджуваних параметрів: математичне сподівання збільшується до 0,070...0,452 мВ (у 2333...15067 разів), дисперсія досягає 78,44...719,20 мВ² (збільшення на 5...6 порядків), варіаційний розмах становить 46,2...122,9 мВ (збільшення у 960...1500 разів) [31].

Виявлені значні відмінності статистичних показників ФАВ між нормою та патологією свідчать про високий діагностичний потенціал запропонованого методу. Проте для практичного застосування ФАВ у клінічній діагностиці необхідно розробити математичну модель автоматичної класифікації стану пацієнта на основі обчислених параметрів. Враховуючи багатофакторну природу амплітудної варіабельності, доцільним є застосування методів багатофакторного регресійного аналізу.

Для побудови та валідації багатофакторної регресійної моделі прогнозування кардіологічного

діагнозу (БРМПКД) на основі ФАВ сформовано репрезентативну вибірку, що включала 204 електрокардіосигнали (ЕКС). Структура вибірки така:

- 102 ЕКС (50 %) — записи пацієнтів з умовною нормою, що не мали клінічно значущих відхилень у роботі ССС, зокрема 51 з кардіостимулятором та 51 без кардіостимулятора;
- 51 ЕКС (25 %) — записи пацієнтів з порушеннями ритму серця, представленими різними формами екстрасистолії;
- 51 ЕКС (25 %) — записи пацієнтів з морфологічними патологіями проведення збудження, а саме неповною блокадою лівої ніжки пучка Гіса.

Така структура вибірки забезпечує збалансоване представлення основних типів кардіологічної патології та дозволяє розробити універсальну модель класифікації.

Первинний етап дослідження включав розрахунок комплексу з 13-ти статистичних показників для кожної реалізації ФАВ:

1. Mean (Середнє арифметичне) — характеризує центральну тенденцію розподілу амплітудної варіабельності.
2. Standard Error (Стандартна похибка) — оцінює точність визначення середнього значення.
3. Median (Медіана) — робастна міра центральної тенденції, стійка до викидів.
4. Mode (Мода) — найчастіше значення амплітудної варіабельності.
5. Standard Deviation (Стандартне відхилення) — характеризує розсіювання значень навколо середнього.
6. Sample Variance (Вибіркова дисперсія) — квадрат стандартного відхилення, міра варіабельності.
7. Kurtosis (Експес) — характеризує гостровершинність розподілу.
8. Skewness (Асиметрія) — визначає скошеність розподілу.
9. Range (Розмах) — різниця між максимальним та мінімальним значеннями.
10. Minimum (Мінімум) — найменше значення амплітудної варіабельності.
11. Maximum (Максимум) — найбільше значення амплітудної варіабельності.
12. Sum (Сума) — сумарне значення всіх елементів вибірки.
13. Confidence Level 95,0 % (Довірчий інтервал) — межі, в яких з імовірністю 95 % знаходиться істинне середнє.

Для побудови БРМПКД застосовано метод покрокової регресії з прямим відбором змінних. Статистична обробка даних проводилася з використанням програмного пакету Statistica 12.0 (StatSoft Inc., США). Критерієм включення предикторів у модель встановлено рівень значущості $p < 0,05$, критерієм виключення — $p > 0,10$. Застосування процедури покрокового регресійного аналізу дозволило з початкового набору 13 статистичних показників виділити 7 статистично значущих предикторів ($p < 0,001$), які увійшли до фінальної моделі: Mean (Середнє), Median (Медіана), Mode (Мода), StDev (Стандартне відхилення), SampleVar (Вибіркова дисперсія), Kurtosis (Експес) та Skewness (Асиметрія).

Аналіз стандартизованих коефіцієнтів регресії (рис. 1) дозволив оцінити відносний внесок кожного предиктора у формування прогностичного коефіцієнта.

Regression Summary for Dependent Variable: CDPCAVF (1 in FAV) R= ,99554012 R²= ,99110013 Adjusted R²= ,99078227 F(7,196)=3118,1 p<0,0000 Std.Error of estimate: ,10761						
N=204	b*	Std. Err. of b*	b	Std. Err. of b	t(196)	p-value
Intercept			2,3212	0,036143	64,2227	0,000000
Mean	1,66679	0,034653	201,7753	4,195013	48,0988	0,000000
Median	4,81349	0,130566	68,6924	1,863285	36,8663	0,000000
Mode	-1,50829	0,081852	-12,7402	0,691387	-18,4271	0,000000
StDev	-9,76335	0,261504	-26,9270	0,721220	-37,3353	0,000000
SampleVar	11,13001	0,330053	28,8454	0,855393	33,7219	0,000000
Kurtosis	0,09962	0,011519	0,0987	0,011416	8,6486	0,000000
Skewness	0,13099	0,011418	0,3290	0,028682	11,4722	0,000000

Рис. 1. Результат значущих факторів для коефіцієнта прогнозування кардіологічного діагнозу функції амплітудної варіабельності у разі проведення багатофакторного регресійного аналізу в програмі Statistica 12.0

Найбільший позитивний вплив мають показники Mean ($\beta = 201,7753$) та Median ($\beta = 68,6924$), що підкреслює важливість мір центральної тенденції для діагностики. Негативні коефіцієнти при Mode ($\beta = -12,7402$), StDev ($\beta = -26,9270$) вказують на зворотний зв'язок цих параметрів з діагностичним коефіцієнтом. Показники форми розподілу — Kurtosis ($\beta = 0,0987$) та Skewness ($\beta = 0,3290$) — мають менший, але статистично значущий вплив. Відображення $p = 0,000000$ є результатом об-

межень форматування в Statistica 12.0. Фактичні p -значення для всіх предикторів $\epsilon < 0,001$, з найменшими значеннями порядку 10^{-15} для основних предикторів.

На основі отриманих результатів регресійного аналізу сформовано математичну модель для розрахунку коефіцієнта прогнозування кардіологічного діагнозу функції амплітудної варіабельності (CDPCAVF — Cardiac Diagnosis Prediction Coefficient of the Amplitude Variability Function):

$$\text{CDPCAVF} = 201.7753 * \text{Mean} + 68.6924 * \text{Median} - 12.7402 * \text{Mode} - 26.9270 * \text{StDev} + 28.8454 * \text{SampleVar} + 0.0987 * \text{Kurtosis} + 0.3290 * \text{Skewness} + 2.3212,$$

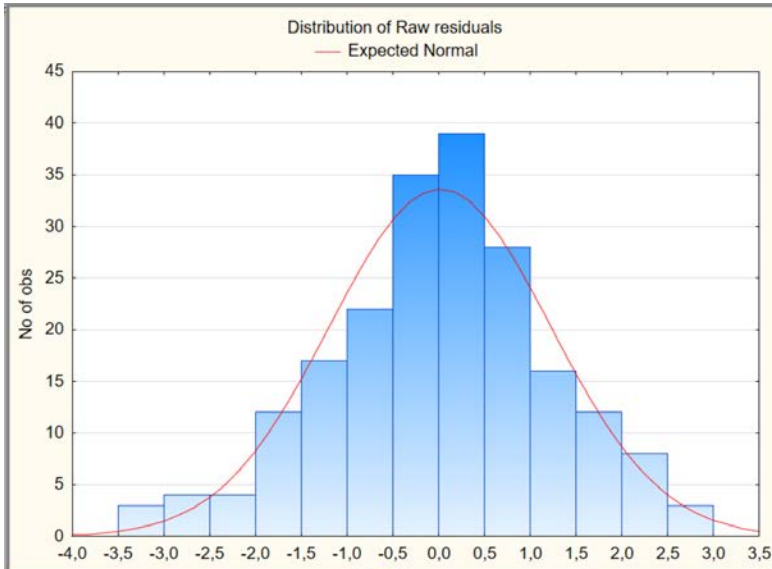


Рис. 2. Гістограма залишкових відхилень багаторефлексивної регресійної моделі прогнозування кардіологічного діагнозу на основі функції амплітудної варіабельності

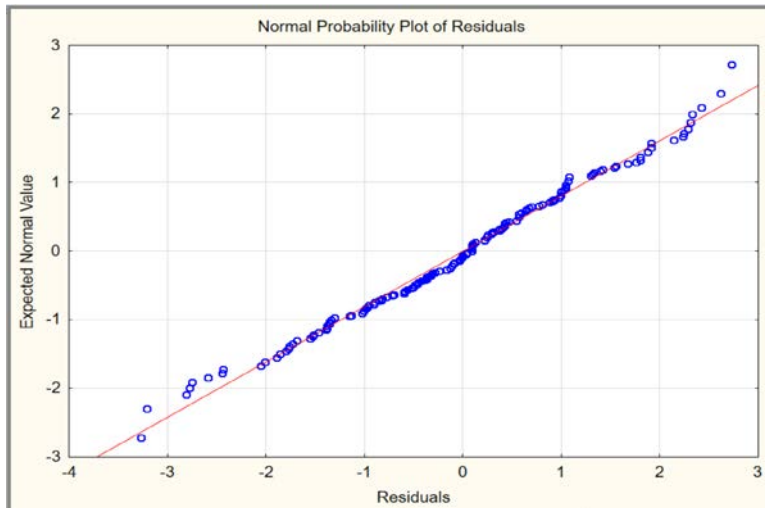


Рис. 3. Нормально-ймовірнісний графік залишкових відхилень багаторефлексивної регресійної моделі прогнозування кардіологічного діагнозу на основі функції амплітудної варіабельності

(воронкоподібного розширення, криволінійних трендів) свідчить про виконання умови гомоскедастичності.

Для оцінювання статистичної значущості моделі в цілому проведено дисперсійний аналіз (ANOVA) (рис. 5).

де всі параметри розраховуються для конкретної реалізації ФАВ досліджуваного пацієнта.

Критичним етапом оцінювання якості регресійної моделі є детальний аналіз залишкових відхилень. Гістограма розподілу залишків (рис. 2) демонструє симетричний розподіл з вираженим центральним піком та поступовим зменшенням частот до периферії. Візуальний аналіз свідчить про наближення емпіричного розподілу до теоретичної кривої нормального розподілу, що є важливою передумовою для застосування методу найменших квадратів.

Для додаткової перевірки нормальності розподілу залишків побудовано нормально-ймовірнісний графік (Q-Q plot) (рис. 3). Аналіз графіка виявив, що емпіричні точки розташовуються вздовж теоретичної прямої без систематичних відхилень, що підтверджує відповідність розподілу залишків нормальному закону. Незначні відхилення в області екстремальних значень не перевищують допустимих меж і не впливають на загальну адекватність моделі.

Важливим припущенням регресійного аналізу є постійність дисперсії залишків (гомоскедастичність). Для перевірки цього припущення побудовано діаграму розсіювання залишків відносно прогнозованих значень (рис. 4). Візуальний аналіз діаграми демонструє хаотичний характер розподілу точок без вираженої залежності розкиду залишків від величини прогнозованих значень. Відсутність систематичних патернів

Результати аналізу демонструють високу статистичну значущість регресійної моделі: F-критерій Фішера становить 953,9314 за рівня значущості $p < 0,001$. Це свідчить про те, що розроблена модель забезпечує статистично достовірне прогнозування порівняно з нульовою гіпотезою (використання лише середнього значення).

Коефіцієнт детермінації Нейджелкера (Nagelkerke R^2) є універсальною мірою якості регресійної моделі, що показує частку варіації залежної змінної, яка пояснюється включеними предикторами. Для розробленої моделі $R^2 = 0,99110013$, що означає врахування 99,1 % варіації діагностичного коефіцієнта вибраними статистичними показниками ФАВ.

Такий високий показник коефіцієнта детермінації свідчить про:

- адекватний вибір предикторів для моделювання;
- відсутність значущих неврахованих факторів;
- високу прогностичну здатність моделі;
- можливість практичного застосування для автоматичної діагностики.

Отримана математична модель дозволяє здійснювати об'єктивну кількісну оцінку функціонального стану ССС на основі аналізу амплітудної варіабельності ЕКС. Високі значення коефіцієнта CDPCAUF свідчать про наявність патологічних змін, тоді як низькі значення характерні для нормального функціонування серця.

Аналіз структури моделі виявив, що найбільший внесок у формування діагностичного висновку вносять показники центральної тенденції (Mean, Median), що відображають загальний рівень амплітудної варіабельності. Показники розсіювання (StDev, SampleVar) та форми розподілу (Kurtosis, Skewness) відіграють модульовальну роль, уточнюючи діагностичну оцінку.

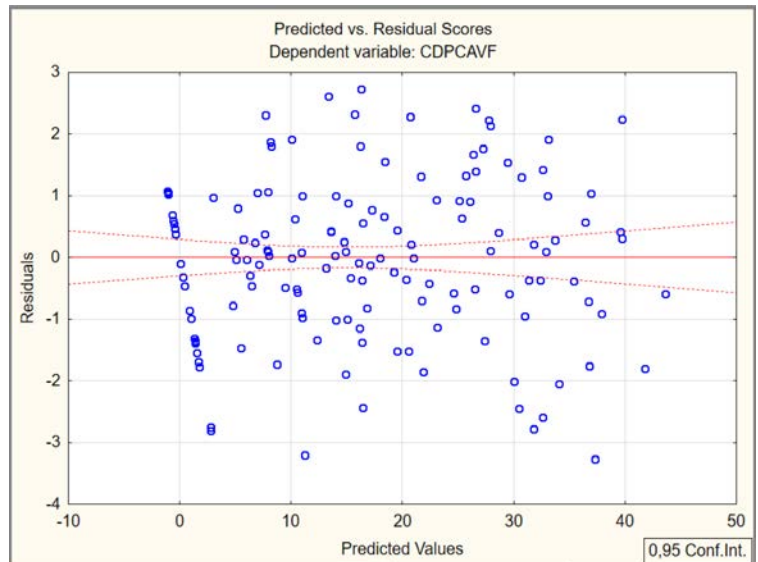


Рис. 4. Діаграма розсіювання залишкових відхилень регресійної моделі прогнозування кардіологічного діагнозу на основі функції амплітудної варіабельності

Analysis of Variance; DV: CDPCAUF (1 in FAV)					
Effect	Sums of Squares	df	Mean Squares	F	p-value
Regress.	23148,51	15	1543,234	953,9314	0,00
Residual	302,52	187	1,618		
Total	23451,03				

Рис. 5. Результати аналізу ANOVA

Висновки

У результаті проведеного дослідження розроблено та валідовано багатофакторну регресійну модель прогнозування кардіологічного діагнозу (CDPCAUF) на основі ФАВ ЕКС. Запропонована модель забезпечує об'єктивну кількісну оцінку функціонального стану ССС з високою точністю, про що свідчить коефіцієнт детермінації $R^2 = 0,991$, який демонструє врахування 99,1 % варіації діагностичного коефіцієнта вибраними статистичними показниками. Застосування методу покрокової регресії дозволило з початкового набору 13 статистичних характеристик ФАВ виділити 7 статистично значущих предикторів ($p < 0,001$), серед яких ключову роль відіграють показники центральної тенденції — Mean ($\beta = 201,78$) та Median ($\beta = 68,69$), що відображають загальний рівень амплітудної варіабельності. Показники розсіювання та форми розподілу виконують модульовальну функцію, уточнюючи діагностичну оцінку та підвищуючи специфічність моделі.

Комплексна статистична валідація підтвердила високу надійність та адекватність розробленої моделі. F-критерій Фішера становить 953,93 за рівня значущості $p < 0,001$, що свідчить про статистичну достовірність регресійної залежності. Аналіз залишкових відхилень продемонстрував їх відповідність нормальному закону розподілу та виконання умови гомоскедастичності, що підтверджує коректність вибраної лінійної структури моделі та відсутність систематичних похибок про-

гнозування. Запропонований підхід успішно поєднує переваги високої інтерпретованості класичних регресійних моделей з інноваційним використанням функції амплітудної варіабельності для комплексного аналізу морфологічних і ритмічних діагностичних ознак електрокардіосигналів.

Перспективи подальших досліджень охоплюють розширення спектру диференційованих патологій ССС, інтеграцію з гібридними підходами машинного навчання для підвищення робастності під час роботи з зашумленими сигналами, а також проведення розширеної клінічної апробації на репрезентативних вибірках пацієнтів різних вікових груп та з різноманітною кардіологічною патологією.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] B. Li, and Y. Lian, "A Forecasting Approach for Wholesale Market Agricultural Product Prices Based on Combined Residual Correction," *Applied Sciences*, no. 15(10), 5575, 2025. <https://doi.org/10.3390/app15105575>.
- [2] S. Ali, M. Ali, D. Bhatti, M. S., and B. J. Choi, "Explainable Clustered Federated Learning for Solar Energy Forecasting," *Energies*, no. 18(9), 2380, 2025. <https://doi.org/10.3390/en18092380>.
- [3] H. Zhang, Y. Liu, C. Zhang, and N. Li, "Machine Learning Methods for Weather Forecasting: a Survey," *Atmosphere*, no. 16(1), pp. 82, 2025. <https://doi.org/10.3390/atmos16010082>.
- [4] H. Hewamalage, K. Ackermann, and C. Bergmeir, "Forecast evaluation for data scientists: common pitfalls and best practices. Data Mining and Knowledge Discovery," no. 37(2), pp. 788-832, 2022. <https://doi.org/10.1007/s10618-022-00894-5>.
- [5] G. Skenderi, C. Joppi, M., and M. Cristani, "On the Use of Learning-Based Forecasting Methods for Ameliorating Fashion Business Processes: A Position Paper," *In Lecture Notes in Computer Science*, pp. 647-659, 2023. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-37742-6_50.
- [6] J. Kaur, K. S. Parmar, and S. Singh, "Autoregressive models in environmental forecasting time series: a theoretical and application review" *Environmental Science and Pollution Research*, no. 30(8), pp. 19617-19641, 2023. <https://doi.org/10.1007/s11356-023-25148-9>.
- [7] M. F. Rizvi, "ARIMA Model Time Series Forecasting," *International Journal for Research in Applied Science and Engineering Technology*, no. 12(5), pp. 3782-3785, 2024. <https://doi.org/10.22214/ijraset.2024.62416>.
- [8] "Vector Autoregressive Models for Multivariate Time Series. (n.d.)," in *Modeling Financial Time Series with S-PLUS®*, pp. 385-429. Springer New York. https://doi.org/10.1007/978-0-387-32348-0_11.
- [9] M. Di Mauro, G. Galatro, F. Postiglione, W. Song, and A. Liotta, "Multivariate Time Series Characterization and Forecasting of VoIP Traffic in Real Mobile Networks," *IEEE Transactions on Network and Service Management*, no. 21(1), pp. 851-865, 2024. <https://doi.org/10.1109/tnsm.2023.3295748>.
- [10] S. Verstyuk, "Modeling Multivariate Time Series in Economics: From Auto-Regressions to Recurrent Neural Networks," *SSRN Electronic Journal*, 2019. <https://doi.org/10.2139/ssrn.3357211>.
- [11] M. Khairalla, X. Ning, and N. AL-Jallad, "Modelling and optimisation of effective hybridisation model for time-series data forecasting," *The Journal of Engineering*, no. 2018 (2), pp. 117-122, 2018. <https://doi.org/10.1049/joe.2017.0337>.
- [12] Z. Fang, S. Yang, C. Lv, S. An, and W. Wu, "Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study," *BMJ Open*, no. 12 (7), e056685, 2022. <https://doi.org/10.1136/bmjopen-2021-056685>.
- [13] O. S. Alamu, and M. K. Siam, "Stock Price Prediction and Traditional Models: An Approach to Achieve Short-, Medium- and Long-Term Goals," *Journal of Intelligent Learning Systems and Applications*, no. 16(04), pp. 363-83, 2024. <https://doi.org/10.4236/jilsa.2024.164018>.
- [14] E. Buchweitz, S. Ahal, O. Papish, and G. Adini, *Two-Stage Regularization of Pseudo-Likelihood Estimators with Application to Time Series (Version 2)*. arXiv, 2020. <https://doi.org/10.48550/ARXIV.2007.11306>.
- [15] E. R. Isaac, and B. Singh, *QBSD: Quartile-Based Seasonality Decomposition for Cost-Effective RAN KPI Forecasting (Version 3)*. arXiv, 2023. <https://doi.org/10.48550/ARXIV.2306.05989>.
- [16] Z. Hossain, A. Rahman, M. Hossain, and J. H. Karami, "Over-Differencing and Forecasting with Non-Stationary Time Series Data," *Dhaka University Journal of Science*, no. 67(1), pp. 21-26, 2019. <https://doi.org/10.3329/dujs.v67i1.54568>.
- [17] Ilkka Karanta, "Optimization methods in time series interpolation," *Communications in Statistics - Simulation and Computation*, no. 22 (4), 1pp. 181-1203, 1993. <https://doi.org/10.1080/03610919308813148>.
- [18] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, "Deep Learning for Time Series Forecasting: A Survey," *Big Data*, no. 9(1), pp. 3-21, 2021. <https://doi.org/10.1089/big.2020.0159>.
- [19] Siami-Namini, S., Tavakoli, N., and A. S. Namin, *A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM (Version 1)*, arXiv, 2019. <https://doi.org/10.48550/ARXIV.1911.09512>.
- [20] S.-K. Sim, P. Maass, and P. G. Lind, "Wind Speed Modeling by Nested ARIMA Processes," *Energies*, no. 12(1), pp. 69, 2018. <https://doi.org/10.3390/en12010069>.
- [21] B. Lim, and S. Zohren, "Time-series forecasting with deep learning: a survey. Philosophical Transactions of the Royal Society," *Mathematical, Physical and Engineering Sciences*, no. 379(2194), 20200209, 2021. <https://doi.org/10.1098/rsta.2020.0209>.
- [22] T. Hall, and K. Rasheed, "A Survey of Machine Learning Methods for Time Series Prediction," *Applied Sciences*, 15(11), 5957, 2025. <https://doi.org/10.3390/app15115957>.
- [23] R. P. Masini, M. C. Medeiros, and E. F. Mendes, *Machine Learning Advances for Time Series Forecasting (Version 3)*, arXiv. 2020. <https://doi.org/10.48550/ARXIV.2012.12802>.
- [24] A. Sanghani, N. Bhatt, and N. C. Chauhan, *A Review of Soft Computing Techniques for Time Series Forecasting. Indian Journal of Science and Technology*, 9(S1), 2016. <https://doi.org/10.17485/ijst/2016/v9is1/99604>.
- [25] Ü. Ç. Büyüksahin, and Ş. Ertekin, "Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition," *Neurocomputing*, no. 361, pp. 151-163, 2019. <https://doi.org/10.1016/j.neucom.2019.05.099>.
- [26] L. Zhang, W. Bian, W. Qu, L. Tuo, and Y. Wang, "Time series forecast of sales volume based on XGBoost," *Journal of Physics: Conference Series*, no. 1873(1), 012067, 2021. <https://doi.org/10.1088/1742-6596/1873/1/012067>.

- [27] E. Vaiciukynas, P. Danenas, V. Kontrimas, and R. Butleris, "Two-Step Meta-Learning for Time-Series Forecasting Ensemble," *IEEE Access*, no. 9, pp. 62687-62696, 2021. <https://doi.org/10.1109/access.2021.3074891>.
- [28] M. Zanotti, *The Cost of Ensembling: Is it Always Worth Combining?* Elsevier BV, 2025. <https://doi.org/10.2139/ssrn.5287421>
- [29] P. Lara-Benitez, M. Carranza-García, and J. C. Riquelme, "An Experimental Review on Deep Learning Architectures for Time Series Forecasting," *International Journal of Neural Systems*, no. 31(03), 2130001, 2021. <https://doi.org/10.1142/s0129065721300011>.
- [30] A. Navon, and Y. Keller, *Financial Time Series Prediction Using Deep Learning (Version 1)*, arXiv, 2017. <https://doi.org/10.48550/ARXIV.1711.04174>.
- [31] I. Chaikovskiy, et al., "Artificial intelligence in monitoring and correction of functional state based on electrocardiosigna," in *Advances in Artificial Intelligence*, pp. 237-293, 2024. Elsevier. <https://doi.org/10.1016/b978-0-443-19073-5.00015-x>
- [32] S. M. Koval, et al., *Classification of nonstationary cardiac signals based on their spectral and probabilistic properties. In Information Technology in Medical Diagnostics I*, pp. 267-274, 2019. CRC Press. <https://doi.org/10.1201/9780429057618-31>.
- [33] S. Lupenko, Ia. Lytvynenko, A. Sverstiuk, B. Shelestovskyi, A. Horkunenko, "Software for statistical processing and modeling of synchronously registered cardio signals," *The Fourth International Workshop on Computer Modeling and Intelligent Systems (CMIS-2021)*. 27 April 2021, Zaporizhzhia, Ukraine. p. 194-205.
- [34] Y. Leshchyshyn, L. Scherbak, O. Nazarevych, V. Gotovych, P. Tymkiv, and G. Shymchuk, "Multicomponent Model of the Heart Rate Variability Change-point," in *2019 IEEE XVth International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, pp. 110-113, 2019. IEEE. <https://doi.org/10.1109/memstech.2019.8817379>.
- [35] Я. П. Драган, Ю. Б. Паляниця, О. В. Гевко, і І. Ю. Дедів, «Обґрунтування структури системи дистанційної діагностики адаптаційних резервів серця», *Науковий вісник НЛТУ України*, зб. наук.-техн. праць, (25.10), с. 255-259, 2015.
- [36] Я. Драган, «Математичне й алгоритмічно-програмне забезпечення комп'ютерних засобів статистичного опрацювання коливань (ритмічних процесів)», *Вісник Нац. ун-ту «Львівська політехніка»*, № 621, с. 124-130, 2008.
- [37] A. Sverstiuk, L. Mosiy, "Information technology for electrocardiographic signal analysis based on mathematical models of temporal and amplitude variability," *Computer systems and information technologies*, no. 2, pp. 36-44, 2025. <https://doi.org/10.31891/csit-2025-2-4>.

Рекомендована кафедрою біомедичної інженерії та оптико-електронних систем ВНТУ

Стаття надійшла до редакції 6.08.2025

Сверстюк Андрій Степанович — д-р техн. наук, професор, професор кафедри медичної інформатики, Тернопільського національного медичного університету імені І. Я. Горбачевського; професор кафедри комп'ютерних наук Тернопільського національного технічного університету імені Івана Пулюя, e-mail: sverstyuk@tdmu.edu.ua;

Мосій Любомир Євгенійович — аспірант кафедри комп'ютерних наук, e-mail: lmosiy@gmail.com.

Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль

A. S. Sverstiuk^{1, 2}

L. Ye. Mosiy¹

Multifactor Regression Analysis for Cardiac Diagnosis Prediction Based on Amplitude Variability Function

¹Ternopil Ivan Puluj National Technical University;

²Ternopil Ivan Horbachevsky National Medical University

Multifactor regression model for automatic cardiac diagnosis prediction based on statistical characteristics of the amplitude variability function of electrocardiographic signals is proposed. The model employs seven statistically significant predictors (mean, median, mode, standard deviation, sample variance, kurtosis, and skewness) selected by means of stepwise regression with forward variable selection from the initial set of 13 parameters. The predictor inclusion criterion established a significance level of $p < 0.05$. The model for prediction of cardiac diagnosis, based on the function of the amplitude variability takes into consideration weight coefficients for each predictor, determined by the least squares method. Central tendency measurements make the greatest contribution into diagnostic outcome formation ($\beta = 201.78$ for mean and $\beta = 68.69$ for median).

The model provides effective differentiation among three clinical states: conditional norm, cardiac rhythm disorders in the form of extrasystole, and morphological conduction abnormalities represented by incomplete left bundle branch block,

achieving a Nagelkerke coefficient of determination $R^2 = 0.991$. Normal conditions are characterized by minimal amplitude variability values (mathematical expectation 0.00003...0.00064 mV), while extrasystole demonstrates an increase of 3—4 orders of magnitude. Model validation on a representative sample of 204 electrocardiographic signals (102 normal, 51 extrasystole, 51 block) confirmed its high statistical significance (Fisher's F -criterion 953.93, $p < 0.001$) and compliance with fundamental regression analysis assumptions. Residual analysis demonstrated normal distribution and homoscedasticity, confirming model adequacy.

The proposed approach combines the advantages of high interpretability of classical statistical methods with innovative application of the amplitude variability function for comprehensive analysis of morphological and rhythmic characteristics of cardiac signals. The practical significance lies in creating mathematical tools for automated cardiovascular disease diagnostic systems and clinical decision support systems.

Keywords: amplitude variability function, multifactor regression model, electrocardiographic signal, cardiac diagnosis, stepwise regression, extrasystole, left bundle branch block, CDPCAVF prediction coefficient, statistical predictors, automatic classification of cardiac pathologies.

Sverstiuk Andriy S. — Dr. Sc. (Eng.), Professor, Professor of the Chair of Medical Informatics of Ternopil Ivan Horbachevsky National Medical University, Professor of the Chair of Computer Science of Ternopil Ivan Puluj National Technical University, e-mail: sverstyuk@tdmu.edu.ua;

Mosiy Lyubomyr Ye. — Post-Graduate Student of the Chair of Computer Science, e-mail: lmosiy@gmail.com