

**Баглай О.П.**<https://orcid.org/0009-0003-7814-0206>

Український державний університет залізничного транспорту

## ВИКОРИСТАННЯ СУЧАСНИХ СТАТИСТИЧНИХ МЕТОДІВ У ЗВАРЮВАННІ МЕТАЛОКОНСТРУКЦІЙ

Статтю присвячено аналізу сучасних статистичних методів та інформаційних технологій у задачах контролю якості, прогнозування дефектів і оптимізації процесів зварювання металоконструкцій. Показано обмеженість традиційних емпіричних підходів у використанні великих масивів технологічних даних реального часу. У роботі розглянуто перехід від суб'єктивного контролю до даних-орієнтованого аналізу процесів зварювання в контексті Індустрії 4.0. Особливу увагу приділено методам описової та математичної статистики для виявлення закономірностей між вхідними параметрами процесу зварювання (струм, напруга) та вихідними характеристиками зварного шва (стабільність процесу, ймовірність утворення дефектів).

На основі датасету *Metal Arc Welding – Predictive Quality Data* [1] виконано попередню обробку та стандартизацію часових рядів, візуалізацію даних, обчислення описової статистики, кореляційний аналіз (коефіцієнт Пірсона) [2, с. 3], перевірку нормальності розподілів за критерієм Шаніро-Уїлка [3, с. 92] та порівняння класів якості за допомогою непараметричного тесту Манна-Уїтні.

Візуалізація показала, що дефектні шви мають нестабільну дугу з глибокими провалами струму та піками напруги, тоді як якісні – плавну динаміку. Статистичний аналіз підтвердив значущі відмінності ( $p < 0.001$ ): середній струм у дефектних швах вищий на 0.492 стандартних відхилення, дисперсія струму більша на 11,7 %, а кореляція між струмом і напругою змінюється з негативної ( $r = -0.143$ ) на помірну позитивну ( $r = 0.286$ ). Розподіли ненормальні ( $p < 0.001$ ). Отримані статистичні характеристики були використані як ознаки для побудови моделей класифікації. Застосування логістичної регресії як параметричного статистичного методу дозволило досягти точності 70.1 % ( $ROC-AUC = 0.79$ ), тоді як непараметричний метод  $k$  найближчих сусідів ( $k$ -NN) продемонстрував вищу ефективність з точністю 76.4 % та  $ROC-AUC = 0.85$ , що підтверджує інформативність статистичних ознак для прогнозування якості зварного шва. Отримані дані можуть слугувати основою для розробки інтелектуальних систем моніторингу та автоматизованого управління процесами зварювання в реальному часі, що відповідає завданням цифрової трансформації промисловості та підвищення конкурентоспроможності українського машинобудування.

**Ключові слова:** зварювання металоконструкцій, статистичні методи, аналіз часових рядів, контроль якості, Індустрія 4.0, якість зварювання.

**Постановка проблеми.** Незважаючи на прогрес у технологіях зварювання, забезпечення стабільної якості зварних з'єднань у металоконструкціях залишається актуальною проблемою машинобудування та будівництва. Дефектні шви (тріщини, пористість, геометричні спотворення) є однією з основних причин передчасного руйнування конструкцій під дією навантажень, корозії та втоми матеріалу. За оцінками, дефекти зварювання спричиняють 20-30 % аварій на критичних об'єктах, що призводить до значних економічних втрат і загрози безпеці. [4, с. 1].

Традиційні методи контролю якості зварних швів базуються переважно на візуальному огляді,

вибіркових неруйнівних випробуваннях (ультразвук, рентген, магнітно-порошкова дефектоскопія) та досвіді зварювальника. Такі підходи мають суттєві обмеження:

– вони є післяпроцесними (виявляють дефекти вже після завершення зварювання);

– не враховують стохастичний характер процесу (випадкові коливання струму, напруги, швидкості подачі дроту, зовнішніх умов);

– не дозволяють прогнозувати ймовірність дефектів і оперативно коригувати параметри процесу [5, с. 406].

Унаслідок цього значна частина дефектів виявляється лише на етапі експлуатації, що зни-

жує надійність конструкцій і збільшує витрати на ремонт. Водночас розвиток сенсорних систем, цифрових технологій збору даних та концепцій Індустрії 4.0 створює передумови для переходу до проактивного, даних-орієнтованого управління якістю зварювання [6, с. 21]. Основною проблемою залишається відсутність системного застосування сучасних статистичних методів для аналізу й прогнозування параметрів зварювального процесу в реальному часі, що обмежує можливості раннього виявлення нестабільності дуги та автоматизованого коригування режимів зварювання.

**Аналіз останніх досліджень і публікацій.** Аналіз наукових публікацій свідчить про поступову еволюцію підходів до забезпечення якості та надійності зварних з'єднань у металоконструкціях – від емпіричних методів підбору параметрів до комплексного використання статистичних, чисельних та інформаційних технологій аналізу даних. Сучасні дослідження зосереджені на оптимізації параметрів процесу зварювання, прогнозуванні експлуатаційних характеристик і автоматизованому контролю якості.

У роботі Abatta-Jacome та ін. (2025) [7, с. 13] для оцінки механічних властивостей зварних і незварних арматурних прутків зі сталі ASTM A706 застосовано класичні статистичні методи (критерії Шапіро-Уїлка, Андерсона-Дарлінга та Манна-Уїтні). Отримані результати підтвердили ефективність статистичних тестів для виявлення прихованих змін міцнісних характеристик і обґрунтували доцільність їх використання для стандартизації процесів зварювання, що узгоджується з підходами, використаними в даній роботі.

У публікації Muhammed та ін. (2020) [8, с. 22] автори показали можливість прогнозування корозійної стійкості та міцнісних характеристик на основі статистичних моделей, що ілюструє перехід від якісного аналізу до формалізованих, даних-орієнтованих підходів, подібних до застосованих у цій роботі.

У роботі Zhang та ін. (2020) [9, с. 13] проаналізовано комп'ютеризовані методи виявлення дефектів зварювання за рентгенографічними зображеннями та показано обмеження традиційних неруйнівних методів, зокрема їх післяпроцесний характер і залежність від оператора. Автори обґрунтовують доцільність переходу до статистичних і даних-орієнтованих підходів прогнозування дефектів у реальному часі.

Дослідження Sumesh та ін. (2021) [10, с. 8] підтвердило ефективність використання машинного навчання (SVM) для прогнозування якості зварю-

вання на основі сенсорних даних, що підкреслює перспективність статистичних і даних-орієнтованих методів контролю процесу.

Незважаючи на це, системний аналіз часових рядів електричних параметрів у реальному часі та їх інтеграція в автоматизовані системи моніторингу залишаються недостатньо вивченими.

**Постановка завдання.** Метою статті є аналіз сучасних статистичних методів обробки даних зварювальних процесів, виявлення закономірностей між параметрами процесу (струм, напруга) та стабільністю зварного шва, а також обґрунтування перспективності статистичного підходу для автоматизованого контролю якості в умовах Індустрії 4.0.

Для досягнення цієї мети необхідно виконати такі завдання:

1. Проаналізувати сучасні підходи до застосування статистичних методів і даних-орієнтованих моделей у задачах аналізу та контролю якості зварних з'єднань;

2. Виконати попередню обробку та стандартизацію часових рядів електричних параметрів процесу зварювання на основі відкритого датасету;

3. Здійснити візуалізацію даних, обчислення описових статистичних характеристик і кореляційний аналіз для виявлення відмінностей між якісними та дефектними швами;

4. Перевірити статистичні гіпотези щодо відмінностей між класами якості з використанням непараметричних критеріїв, а також оцінити можливість застосування моделей логістичної регресії та методу k найближчих сусідів для класифікації якості зварного шва;

5. Узагальнити результати та оцінити ефективність статистичних методів і моделей для автоматизованого контролю якості зварювання.

**Виклад основного матеріалу.** Аналіз сучасних підходів до застосування статистичних методів у задачах контролю якості зварних з'єднань, проведений у попередньому розділі, показав, що більшість досліджень зосереджена на постфактум-оцінці механічних властивостей матеріалів або на обмеженому використанні статистики для окремих параметрів процесу. Водночас системний аналіз часових рядів електричних параметрів (струм, напруга) у реальному часі залишається недостатньо вивченим.

Для практичної перевірки потенціалу статистичних методів у виявленні закономірностей між параметрами процесу та стабільністю шва використано відкритий датасет Metal Arc Welding – Predictive Quality Data [1]. Даний датасет містить

часові ряди значень струму та напруги під час зварювання, а також позначки якості зварного шва («добре» або «погано»), що дозволяє поєднувати статистичний аналіз параметрів процесу з оцінкою стабільності результатів.

Дослідження проведено за допомогою мови програмування Python з використанням бібліотек pandas для обробки даних, matplotlib та seaborn [11, с.1] для візуалізації, scipy для статистичних тестів [12, с.267], а також а також scikit-learn для побудови моделей класифікації.

На першому етапі дослідження було проведено попередню обробку даних. Початковий датасет містив 200 тис. зварювальних циклів із часовими рядами струму та напруги. Розподіл класів якості мав такий вигляд:

- клас «-1» (без мітки) – 112777 записів (53,9 %);
- клас «0» (погано) – 49742 записи (23,8 %);
- клас «1» (добре) – 46666 записів (22,3 %).

Після фільтрації залишено лише розмічені приклади (класи 0 та 1) загальною кількістю 96408 записів. Пропущених значень не виявлено. Виявлені поодинокі аномалії (0,01 %), що виходили за фізично обґрунтовані межі, були скориговані лінійною інтерполяцією. Усі параметри струму та напруги стандартизовано для забезпечення коректності подальшого статистичного аналізу [13, с. 119].

Після попередньої обробки та стандартизації даних розпочато аналіз агрегованих характеристик часових рядів. Для швидкості розрахунків

та зручності візуалізації спочатку використано частину датасету (nrows = 15000 записів). У цій вибірці розподіл класів якості склався так: 11681 запис класу «погано» (0) та 3319 записів класу «добре» (1).

Візуальний аналіз часових рядів струму та напруги (рис. 1) показав чіткі відмінності між якісними та дефектними швами: для класу «добре» характерна стабільна динаміка параметрів, тоді як для класу «погано» спостерігається хаотична поведінка з провалами струму та нестабільністю напруги.

Гістограма середнього струму (рис. 2) підтверджує ці відмінності, демонструючи ширший і зміщений розподіл для дефектних швів порівняно з компактним розподілом для якісних.

Діаграма розсіювання середнього струму та напруги (рис. 3) показує чітке розділення класів: якісні шви утворюють компактний кластер зі стабільними значеннями, тоді як дефектні характеризуються значним розкидом і екстремальними відхиленнями.

Візуальні спостереження підтвердили суттєві відмінності в поведінці параметрів процесу залежно від якості шва. Для кількісної оцінки цих відмінностей проведено статистичний аналіз агрегованих характеристик часових рядів за класами «добре» (1) та «погано» (0). Описова статистика представлена в (табл. 1,2). В таблиці 1 представлена описова статистика агрегованих ознак за класами якості, усі значення стандар-

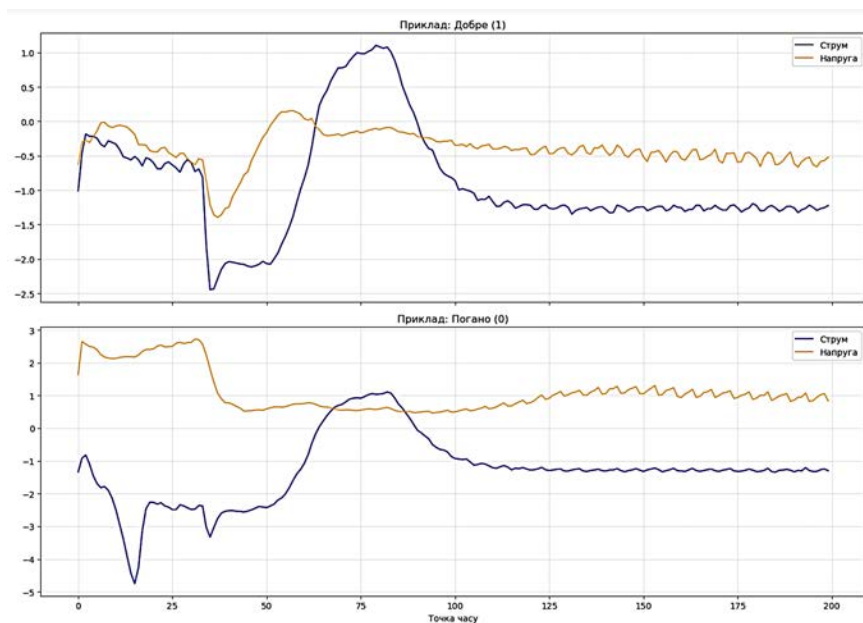


Рис. 1. Часові ряди

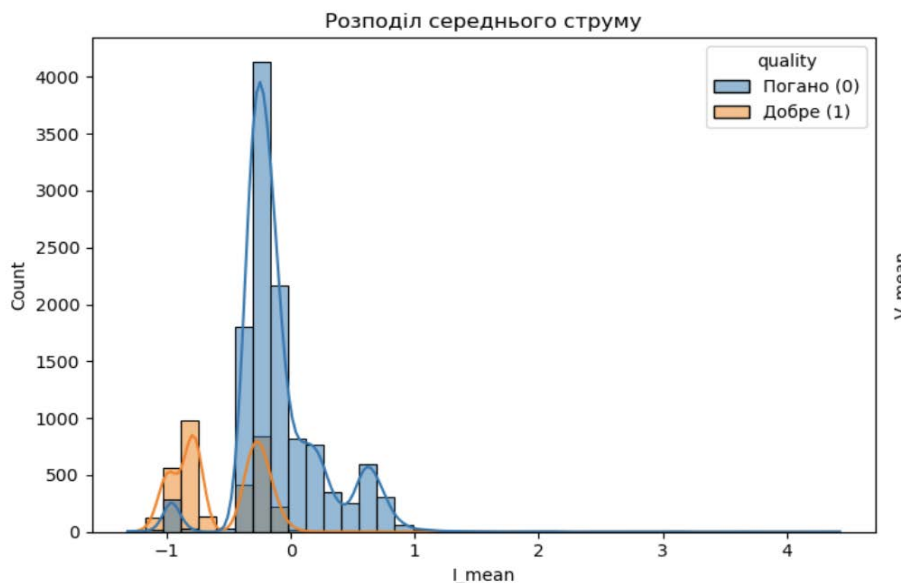


Рис. 2. Гістограма середнього струму

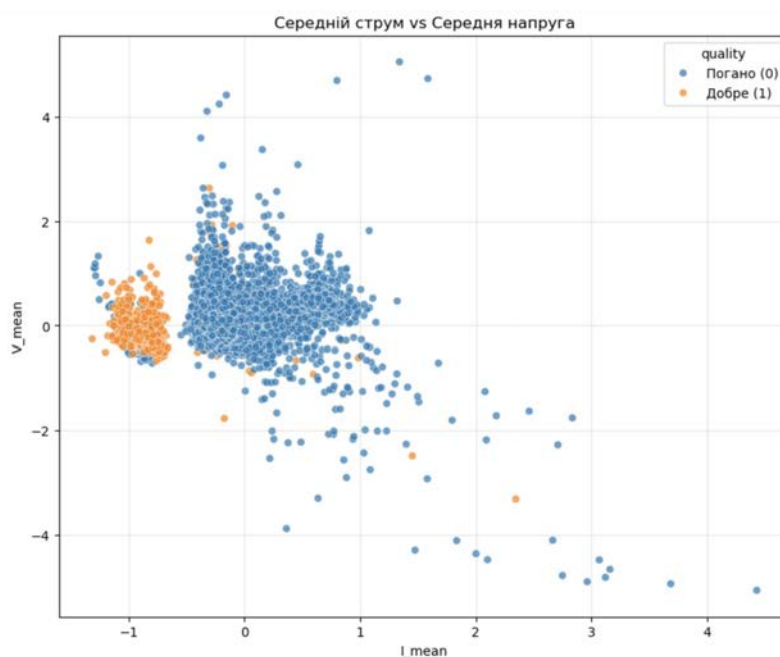


Рис. 3. Діаграма розсіювання

тизовани, тому середнє  $\approx 0$ , а стандартне відхилення  $\approx 1$ . Проаналізувавши отримані результати, можна побачити, що середній струм у дефектних швах вищий на 0.492 стандартних відхилення. Водночас дисперсія струму та напруги вища саме в якісних швах: на 11.7 % для струму та на 20.8 % для напруги. Це свідчить про те, що якісне зварювання характеризується контролю-

ваними коливаннями, тоді як дефектне – зміщенням середнього значення струму при меншій загальній варіативності.

Отримані дані свідчать, що основною ознакою дефектності є не стільки дисперсія, скільки зміщення середнього струму, що може бути використано для автоматизованого виявлення відхилень.

Таблиця 1  
Описова статистика агрегованих ознак  
за класами якості

Параметр	Клас	Середнє	Стандартне відхилення
Середній струм	Добре (1)	-0,254	0,325
Середній струм	Погано (0)	0,238	0,539
Середня напруга	Добре (1)	-0,238	0,612
Середня напруга	Погано (0)	0,223	0,539
Дисперсія струму	Добре (1)	0,651	0,401
Дисперсія струму	Погано (0)	0,583	0,301
Дисперсія напруги	Добре (1)	0,645	0,581
Дисперсія напруги	Погано (0)	0,534	0,455

Для оцінки взаємозв'язку між параметрами проведено кореляційний аналіз. Кореляційний аналіз показав слабку позитивну залежність між середніми значеннями струму та напруги в усіх даних ( $r = 0.207$ ). Однак у якісних швах кореляція стає слабко негативною ( $r = -0.143$ ), що відповідає стабільній роботі дуги (зростання струму супроводжується зменшенням напруги). У дефектних швах кореляція помірна позитивна ( $r = 0.286$ ), що свідчить про порушення стабільності процесу – струм і напруга змінюються синхронно в одному напрямку, що характерно для неконтрольованої дуги.

Перевірка нормальності за критерієм Шапіро-Уїлка показала ненормальний характер розподілів ( $p < 0.001$ ), у зв'язку з чим для порівняння класів застосовано непараметричний тест Манна-Уїтні (табл. 2). Результати підтвердили статистично значущі відмінності між класами за всіма параметрами, що узгоджується з результатами візуального та кореляційного аналізу.

Отримані результати описового, кореляційного та непараметричного аналізу підтвердили наявність стійких відмінностей між якісними та дефектними зварними швами за ключовими параметрами процесу, що створює передумови для переходу від інтерпретаційного аналізу до формалізованого статистичного моделювання.

Таблиця 2  
Порівняння класів якості  
за тестом Манна-Уїтні

Параметр	Статистика тесту	p-value	Висновок
Середній струм	61557111172	$< 0,001$	Значущі відмінності
Середня напруга	515983654	$< 0,001$	Значущі відмінності
Дисперсія струму	123368111714	$1,20e-69$	Значущі відмінності
Дисперсія напруги	130908084229	$5,65e-259$	Значущі відмінності

На наступному етапі дослідження застосовано сучасні статистичні методи класифікації – логістичну регресію та метод k найближчих сусідів (k-NN), які, на відміну від класичних статистичних тестів, дозволяють здійснювати прогноз якості зварного шва на основі сукупності параметрів процесу. Логістична регресія використовується для моделювання ймовірності належності до класу, тоді як метод k-NN як непараметричний підхід не потребує припущень щодо форми розподілів, що є доцільним для зварювальних процесів із ненормальними статистичними характеристиками.

Для оцінки якості побудованих моделей використано стандартні метрики класифікації: accuracy, precision, recall, F1-score та площу під ROC-кривою (ROC-AUC). Оцінювання проводилося на тестовій вибірці, що не використовувалась під час навчання моделей.

Результати логістичної регресії свідчать про задовільну здатність моделі до розрізнення якісних і дефектних швів: точність класифікації становить 70,1 %, а значення ROC-AUC дорівнює 0,79, що вказує на добру дискримінаційну здатність моделі. Повнота для класу «добре» становить 74,2, що свідчить про ефективне виявлення якісних швів.

Метод k найближчих сусідів продемонстрував вищі показники ефективності порівняно з логістичною регресією: точність – 76,6 %, ROC-AUC – 0,85. Це свідчить про кращу адаптацію непараметричної статистичної моделі до складної структури даних та підтверджує перспективність її використання для автоматизованого контролю якості зварювання. Узагальнені результати оцінювання подано в табл. 3.

Таблиця 3  
Порівняння ефективності  
статистичних моделей

Модель	Accuracy, %	Precision	Recall	F1-score	ROC-AUC
Логістична регресія	70,1	67,4	74,2	70,6	0,79
k-NN	76,6	76,0	75,0	75,5	0,85

Матриці помилок для логістичної регресії та методу k-NN (рис. 4) демонструють, що модель k-NN забезпечує більш збалансовану класифікацію обох класів, зменшуючи кількість хибних класифікацій як для дефектних, так і для якісних швів. Це підтверджує її вищу чутливість до локальних особливостей даних.

ROC-криві для обох моделей (рис. 5) підтверджують перевагу методу k-NN, площа під кривою

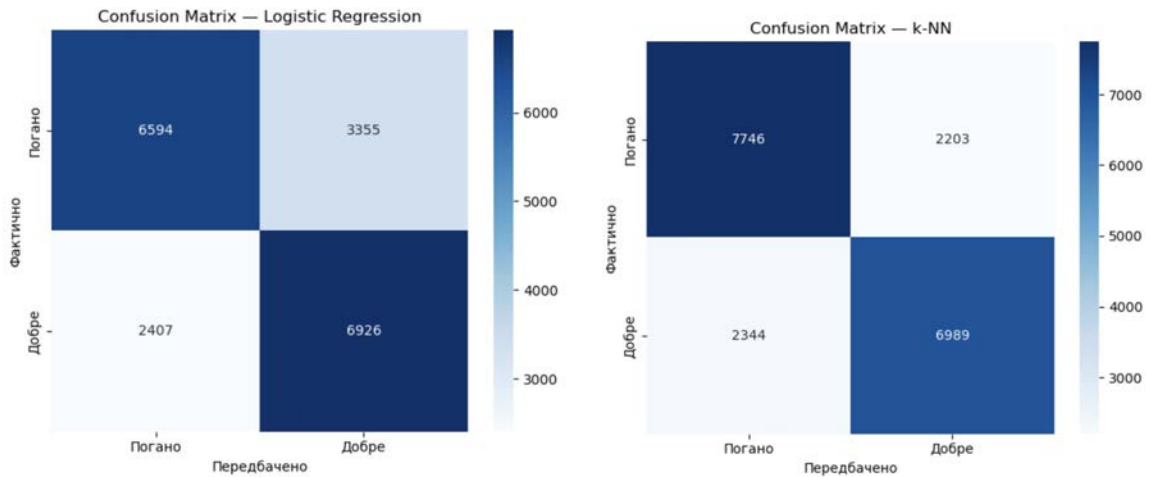


Рис. 4. Матриці помилок для логістичної регресії та k-NN

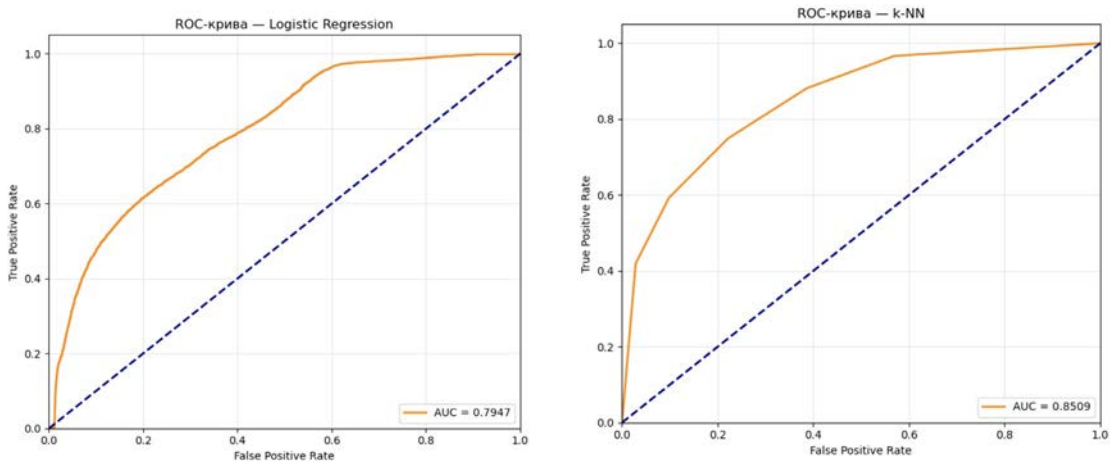


Рис. 5. ROC-криві статистичних моделей класифікації

якого становить 0,85, порівняно з 0,79 для логістичної регресії. Це свідчить про вищу здатність непараметричної статистичної моделі коректно розрізняти класи за різних порогів прийняття рішення, що є критично важливим для задач автоматизованого контролю якості зварювання.

Отримані результати підтверджують, що сучасні статистичні методи класифікації, зокрема логістична регресія та метод k найближчих сусідів, здатні ефективно використовувати агреговані статистичні характеристики часових рядів для прогнозування якості зварних швів. Особливо ефективним виявився непараметричний підхід k-NN, що узгоджується з ненормальним характером розподілів параметрів процесу.

**Висновки.** У результаті проведеного дослідження виконано поставлені завдання та отримано теоретично обґрунтовані й кількісно підтвержені результати щодо можливостей застосування статистичних методів для аналізу й прогнозування

якості зварних швів на основі часових рядів електричних параметрів процесу.

Проаналізовано сучасні підходи до застосування статистичних і даних-орієнтованих методів у задачах контролю якості зварювання. Встановлено, що більшість досліджень зосереджена на постпроцесному аналізі або використанні окремих статистичних тестів, тоді як системний аналіз часових рядів електричних параметрів у реальному часі залишається недостатньо розвиненим.

Виконано попередню обробку та стандартизацію даних відкритого датасету (200 тис. зварювальних циклів). Після фільтрації залишено 96408 розмічених записів (23,8 % – «погано», 22,3 % – «добре»). Аномальні значення становили 0,01 % і були скориговані. Усі ознаки стандартизовано ( $\mu \approx 0$ ,  $\sigma \approx 1$ ), що забезпечило коректність подальшого статистичного аналізу та моделювання.

Виявлено кількісні відмінності між класами якості. Середній струм у дефектних швах вищий на

0,492 стандартних відхилення. Дисперсія струму у якісних швах більша на 11,7 %, а напруги – на 20,8 %. Кореляція між середніми значеннями струму та напруги для якісних швів є слабко негативною ( $r = -0,143$ ), тоді як для дефектних – помірно позитивною ( $r = 0,286$ ). Непараметричний тест Манна-Уїтні підтвердив статистично значущі відмінності за всіма параметрами ( $p < 0,001$ ).

Оцінено ефективність статистичних моделей класифікації. Логістична регресія продемонструвала асигурацію 70,1 % та ROC-AUC 0,79, тоді як

метод k-NN – 76,6 % та 0,85 відповідно. Приріст точності становить 6,5 %, що свідчить про кращу адаптацію непараметричного підходу до складної структури даних.

Підтверджено практичну доцільність інтеграції статистичних методів у системи автоматизованого контролю якості. Отримані моделі забезпечують дискримінаційну здатність до 85 % (ROC-AUC) та можуть бути використані як основа для реалізації проактивного моніторингу процесу зварювання в рамках концепції Індустрії 4.0.

#### Список літератури:

1. Hahn Y. Metal Arc Welding – Predictive Quality Arc Welding Dataset (Version v2) [Electronic resource]. Zenodo, 2023. DOI: 10.5281/zenodo.10017718.
2. Šverko Z., Vrankić M., Vlahinić S., Rogelj P. Complex Pearson correlation coefficient for EEG connectivity analysis. *Sensors*. 2022. Vol. 22, no. 4. Article 1477. DOI: 10.3390/s22041477.
3. Hanusz Z., Tarasinska J., Zielinski W. Shapiro-Wilk test with known mean. *SHILAP Revista de Lepidopterología*. 2022. DOI: 10.57805/revstat.v14i1.180.
4. Kowalczyk M., Andruszko J., Stefanek P., Mazur R. Failure and degradation mechanisms of steel pipelines: analysis and development of effective preventive strategies. *Materials*. 2025. Vol. 18, no. 1. Article 134. DOI: 10.3390/ma18010134.
5. Zhang Y. та ін. Real-time monitoring of welding process using deep learning. *Journal of Manufacturing Processes*. 2023. DOI: 10.1016/j.jmapro.2023.01.012.
6. Баглай О.П., Тимофеева М.А. Застосування інформаційних технологій для контролю якості зварних з'єднань металоконструкцій // Вісник Херсонського національного технічного університету. – Херсон : ХНТУ, 2023. – Вип. 2(77). – С. 18-25. – URL: [https://journals.kntu.kherson.ua/index.php/visnyk\\_kntu/article/view/1019](https://journals.kntu.kherson.ua/index.php/visnyk_kntu/article/view/1019) (дата звернення: 08.02.2026).
7. Abatta-Jacome L., Rosero-Pazmiño D., Rosero-Vivas J., Chávez-Guerrero B. F., Barrionuevo G. O. Statistical evaluation of the mechanical properties of welded and unwelded ASTM A706 reinforcing steel bars of different commercial brands. *Metals*. 2025. Vol. 15, no. 12. Article 1307. DOI: 10.3390/met15121307.
8. Muhammed M., Mustapha M., Ginta T. L., Ali A. M. Statistical review of microstructure-property correlation of stainless steel: implication for pre- and post-weld treatment. *Processes*. 2020. Vol. 8, no. 7. Article 811. DOI: 10.3390/pr8070811.
9. Hou W., Zhang D., Wei Y., Guo J., Zhang X. Review on computer aided weld defect detection from radiography images. *Applied Sciences*. 2020. Vol. 10, no. 5. Article 1878. DOI: 10.3390/app10051878.
10. Sumesh A. та ін. Real-time weld quality prediction using a support vector machine based on optical sensor data. *Journal of Manufacturing Processes*. 2021.
11. Waskom M. L. Seaborn: statistical data visualization. *Journal of Open Source Software*. 2021. Vol. 6, no. 60. Article 3021. DOI: 10.21105/joss.03021.
12. Virtanen P. та ін. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020. Vol. 17. P. 261–272. DOI: 10.1038/s41592-019-0686-2.
13. Zikopoulos P., Eaton C., deRoos D. та ін. Understanding Big Data. New York : McGraw-Hill, 2020.

#### Baglai O.P. APPLICATION OF MODERN STATISTICAL METHODS IN WELDING OF METAL STRUCTURES

*The article is devoted to the analysis of modern statistical methods and information technologies in quality control, defect prediction, and optimization of welding processes for metal structures. The limitations of traditional empirical approaches in processing large volumes of real-time technological data are demonstrated. The study considers the transition from subjective control to data-driven analysis of welding processes within the framework of Industry 4.0. Particular attention is paid to descriptive and mathematical statistical methods for identifying relationships between input welding parameters (current, voltage) and output weld characteristics (process stability, probability of defect formation).*

*Based on the Metal Arc Welding – Predictive Quality Data dataset [1], preprocessing and standardization of time series were performed, including data visualization, descriptive statistical analysis, correlation analysis (Pearson correlation coefficient) [2, p. 3], normality testing using the Shapiro-Wilk test [3, p. 92], and comparison of quality classes using the nonparametric Mann-Whitney test.*

*Visualization revealed that defective welds are characterized by arc instability with deep current drops and voltage peaks, whereas high-quality welds exhibit smoother dynamics. Statistical analysis confirmed significant differences ( $p < 0.001$ ): the mean current in defective welds is higher by 0.492 standard deviations, current variance is 11.7% higher, and the correlation between current and voltage changes from weak negative ( $r = -0.143$ ) to moderately positive ( $r = 0.286$ ). The distributions were found to be non-normal ( $p < 0.001$ ). The obtained statistical characteristics were used as features for classification modeling.*

*The application of logistic regression as a parametric statistical method achieved an accuracy of 70.1% (ROC-AUC = 0.79), while the nonparametric  $k$ -nearest neighbors ( $k$ -NN) method demonstrated higher performance with an accuracy of 76.4% and ROC-AUC = 0.85, confirming the informativeness of the statistical features for weld quality prediction. The results may serve as a basis for developing intelligent monitoring systems and automated real-time welding process control, contributing to the digital transformation of industry and enhancing the competitiveness of Ukrainian mechanical engineering.*

**Keywords:** *welding of metal structures, statistical methods, time series analysis, quality control, Industry 4.0, weld quality.*

Дата першого надходження статті до видання: 27.02.2026  
Дата прийняття статті до друку після рецензування: 24.03.2026  
Дата публікації (оприлюднення) статті 11.05.2026