УДК 004.8:004.056:336.71
DOI: 10.60022/3(6)-19S

Пекуровський Гліб Валерійович
кандидат технічних наук, доцент кафедри технічного захисту інформації,
Національний університет «Київський авіаційний інститут», Україна
Pekurovskyi Hlib
PhD in Technical Sciences, Associate Professor
at the Department of Technical Information Security
National University "Kyiv Aviation Institute", Ukraine
ORCID: 0009-0006-1993-8615

Пекуровський Єгор Валерійович
здобувач третього (освітньо-наукового) рівня вищої освіти
кафедри фінансів, банківської та страхової справи
Навчально-наукового інституту управління економіки та бізнесу
Міжрегіональна Академія Управління персоналом, Україна
Pekurovskyi Yehor
PhD student at the Department of Finance, Banking and Insurance
Educational and Scientific Institute of Management, Economics and Business
Interregional Academy of Personnel Management, Ukraine
ORCID: 0009-0003-6592-7704

БАЛАНСУВАННЯ ДАНИХ У НЕЙРОМЕРЕЖЕВИХ СИСТЕМАХ ВИЯВЛЕННЯ ФІНАНСОВОГО ШАХРАЙСТВА

Анотація. *У роботі досліджується проблема виявлення шахрайських банківських транзакцій в умовах критичного дисбалансу класів. Порівнюються чотири стратегії балансування навчальної вибірки: базова (без балансування), зважені класи, метод SMOTE та випадковий андерсемплінг. Класифікацію виконує нейронна мережа у формі багатошарового перцептрона (MLP). Запропоновано метрику зваженої ефективності виявлення (Weighted Detection Efficiency, WDE), що нормалізує якість моделі з урахуванням коефіцієнта дисбалансу класів і дозволяє коректно порівнювати методи балансування між собою. Результати експериментів показали, що найвище значення WDE демонструє базова модель без балансування, однак вона суттєво програє за повнотою за повнотою виявлення шахрайств. Оптимальним балансом між прецизійністю і повнотою характеризується метод SMOTE.*

Ключові слова: *нейронна мережа, шахрайство у банківській сфері, дисбаланс класів, SMOTE, зважені класи, андерсемплінг, WDE.*

DATA BALANCING IN NEURAL NETWORK-BASED FINANCIAL FRAUD DETECTION SYSTEMS

Abstract. *This paper investigates the critical problem of detecting fraudulent banking transactions under severe class imbalance conditions, which is a common challenge in modern financial data science. In real-world datasets, the number of legitimate transactions significantly outnumbers fraudulent ones, severely hindering the training process and predictive accuracy of machine learning algorithms. To address this issue, this study evaluates and compares four distinct training sample balancing strategies: the baseline approach (no balancing), class weighting, the Synthetic Minority Over-sampling Technique (SMOTE), and random under-sampling.*

The classification task is performed using an artificial neural network configured as a Multilayer Perceptron (MLP). Due to the extreme disproportion of classes, traditional evaluation metrics such as standard accuracy fail to provide an objective assessment of model performance. To overcome this limitation, the authors propose a novel metric — Weighted Detection Efficiency (WDE). This metric normalizes the overall quality of the model by incorporating the class imbalance coefficient, thereby enabling a statistically mathematically justified and correct comparison between different balancing techniques.

Experimental results demonstrated that while the baseline model without any balancing achieved the

highest nominal WDE value due to its high precision on the majority class, it suffered from a critically low recall rate, failing to detect the majority of actual fraud cases. On the other hand, resampling techniques provided a more robust representation of the minority class. The synthesis of the experimental data indicates that the SMOTE method delivers the optimal balance between precision and recall, significantly improving the model's ability to identify fraudulent activities without generating an unacceptable volume of false positives. The findings of this research can be practically applied to enhance the reliability of automated anti-fraud monitoring systems in banking and financial institutions.

Keywords: *neural network, banking fraud, class imbalance, SMOTE, weighted classes, under-sampling, WDE (Weighted Detection Efficiency).*

Постановка проблеми. Виявлення шахрайських фінансових транзакцій є одним із ключових завдань кібербезпеки у банківській сфері. Щорічні збитки від карткового шахрайства у світі становлять близько $34 млрд дол. США [1]. В Україні обсяг втрат від незаконних операцій з використанням платіжних карток у 2025 році збільшився на 24% – до загальної суми 1,4 млрд грн [2]. Це зумовлює необхідність постійного удосконалення автоматизованих систем моніторингу банківських транзакцій.

Відповідно до аналітичних даних, найбільша частка махінацій припадає на транзакції без фізичної присутності картки (card-not-present fraud) [3], тобто онлайн-платежі та дистанційні перекази. У відповідь регулятори посилюють вимоги до систем внутрішнього контролю банків: зокрема, стандарт PCI DSS зобов'язує фінансові установи впроваджувати автоматизований моніторинг підозрілих транзакцій у режимі реального часу. Це створює практичний запит на ефективні алгоритми класифікації, здатні працювати в умовах жорсткого дисбалансу між легітимними та шахрайськими операціями.

Методи машинного навчання демонструють високу ефективність у задачах класифікації транзакцій. Однак ключовою проблемою в цій предметній області є критичний дисбаланс класів: частка шахрайських транзакцій у реальних датасетах становить, як правило, менше ніж 0,5% від загальної кількості. Це суттєво ускладнює навчання класифікаторів, оскільки моделі схильні ігнорувати мінорний клас, досягаючи при цьому формально високої точності.

Для усунення дисбалансу запропоновано низку підходів: методи надвибірки (SMOTE), методи підвибірки (Random Undersampling), а також алгоритмічні рішення, насамперед зважування класів у функції втрат. Однак жодна з цих стратегій не є універсально оптимальною. Ключова проблема полягає в тому, що реальне операційне середовище банківської системи є динамічним. Адже рівень дисбалансу класів постійно змінюється під впливом сезонного характеру попиту, маркетингової активності ритейлерів, поведінкового профілю клієнтів та рівня агресивності зовнішніх атак тощо. Стратегія балансування, оптимальна для розподілу транзакцій у звичайний день, може виявитись неефективною або фінансово збитковою в умовах пікового навантаження чи скоординованої атаки. Крім того, порівняльний аналіз підходів ускладнюється відсутністю єдиної метрики оцінювання, яка б одночасно враховувала ступінь дисбалансу, пріоритетність виявлення шахрайства над хибними спрацюваннями та операційну вартість помилок класифікації.

Аналіз останніх досліджень і публікацій. Загальну методологію, математичний апарат і усталені практики машинного навчання фундаментально описано у [4].

У [5] досліджено вплив стратегії підвибірки на калібрування ймовірностей у задачах виявлення шахрайських транзакцій і показано, що некалібровані моделі можуть хибно інтерпретувати оцінки ризику при сильному дисбалансі.

Проблема дисбалансу класів є добре дослідженою у машинному навчанні. У [6] проведено систематичний огляд методів роботи з незбалансованими даними, класифікувавши їх на рівні даних (re-sampling), алгоритмічні та комбіновані підходи.

Метод SMOTE (Synthetic Minority Over-sampling Technique), запропонований у [7], полягає у генерації синтетичних прикладів мінорного класу шляхом інтерполяції між наявними зразками у просторі ознак. Метод довів свою ефективність у численних задачах класифікації та є одним із найбільш широко використовуваних підходів до надвибірки.

Практична реалізація алгоритмів машинного навчання мовою програмування Python описана у [8], а специфіка реалізації в умовах дисбалансу класів досліджена у [9].

Використання нейронних мереж у задачах виявлення фінансового шахрайства є усталеним підходом [10; 11]. Однак питання вибору оптимальної стратегії балансування для MLP-архітектур у контексті банківських транзакцій залишається актуальним, зокрема щодо метрики оцінювання, яка б враховувала ступінь дисбалансу.

У роботі [12] проведено порівняльний аналіз класичних та глибинних моделей (Logistic Regression, Random Forest, LightGBM, GRU) на датасеті IEEE-CIS Fraud Detection з рівнем дисбалансу $IR = 24{,}25$.

Попри те що модель GRU досягла повноти 0,54 для класу шахрайства, пряме порівняння з результатами на датасетах з вищим є некоректним без нормалізації на складність задачі. Запропонована у даній роботі метрика усуває цей недолік, дозволяючи зіставляти ефективність моделей незалежно від рівня дисбалансу у вихідних даних.

Метою статті є порівняльний аналіз стратегій балансування навчальної вибірки під час тренування багатошарового перцептрону в контексті виявлення шахрайських банківських транзакцій, а також розробка метрики зваженої ефективності виявлення (*WDE*). Параметри останньої обґрунтовуються через вартісну модель операційних втрат фінансової установи, що дає змогу коректно порівнювати алгоритми виявлення шахрайства в умовах динамічної зміни рівня дисбалансу транзакційних даних. На відміну від наявних метрик, *WDE* нормалізує якість класифікатора відносно поточного коефіцієнта дисбалансу класів, що дозволяє: по-перше, коректно порівнювати стратегії балансування між датасетами з різним рівнем дисбалансу; по-друге, відстежувати реальну ефективність моделі в умовах динамічної зміни операційного навантаження; по-третє, розмежовувати деградацію якості класифікатора від природного зростання складності задачі. Параметри метрики обґрунтовуються через вартісну модель операційних втрат, що забезпечує її практичну інтерпретовність у контексті фінансового моніторингу.

Виклад основного матеріалу. У роботі використовується реальний набір даних Credit Card Fraud Detection від дослідницької групи Machine Learning Group (MLG) при ULB (Université Libre de Bruxelles) [13]. Датасет містить 284 807 записів транзакцій, з яких 492 є шахрайськими, що відповідає частці 0,17%. Коефіцієнт дисбалансу класів (*IR*) становить 577,88, що є типовим для реальних систем моніторингу фінансових операцій.

Ознаковий простір включає 28 анонімізованих числових ознак V1–V28, отриманих методом головних компонент (*PCA*) із вихідних транзакційних даних, а також ознаку Amount, що відображає суму транзакції. Цільова змінна Class є бінарною і набуває значення 0 (легітимна транзакція) або 1 (шахрайська транзакція). Ознаку Time було виключено з процесу як нерелевантну для задачі класифікації.

Перед навчанням виконувалась стандартизація числових ознак методом z-нормалізації за допомогою StandardScaler. Розбиття на навчальну (80%) та тестову (20%) вибірки здійснювалось зі стратифікацією по цільовій змінній, що забезпечує збереження пропорцій класів в обох підвибірках.

Ознака Time у даному датасеті не несе значущої предикативної інформації. Адже вона містить лише порядковий номер транзакції у секундах від початку збору, а не реальний час доби чи тиждень.

Класифікатором у дослідженні виступає багатошаровий перцептрон (Multilayer Perceptron, MLP) [6], тобто повнозв'язна нейронна мережа прямого поширення. Архітектура мережі включає вхідний шар розмірністю 29 нейронів (відповідно до кількості вхідних ознак), два приховані шари з 64 та 32 нейронами відповідно, та вихідний шар з одним нейроном для бінарної класифікації.

В ролі функції активації прихованих шарів використовується ReLU (Rectified Linear Unit), що забезпечує ефективне навчання глибоких мереж за рахунок відсутності проблеми затухання градієнтів. Навчання виконується оптимізатором Adam з параметрами за замовчуванням. Для запобігання перенавчанню застосовується механізм ранньої зупинки (early stopping): навчання припиняється, якщо якість на валідаційній вибірці (10% від навчального набору) не покращується протягом 10 послідовних епох. Максимальна кількість епох обмежена значенням 50. Реалізацію виконано засобами бібліотеки scikit-learn (клас MLPClassifier).

Програмну реалізацію дослідження виконано мовою Python 3 з використанням наступних бібліотек: scikit-learn для реалізації моделі MLP, стандартизації даних, розбиття вибірки та обчислення метрик; imbalanced-learn для реалізації методів SMOTE та випадкового андерсемплінгу; pandas та numpy для обробки та маніпуляції даними; matplotlib та seaborn для побудови графіків та візуалізації результатів.

Реалізація охоплює наступні етапи: завантаження та стандартизація даних; формування варіантів навчальної вибірки відповідно до кожної стратегії балансування; навчання окремого екземпляру MLP для кожної стратегії з ідентичними гіперпараметрами; обчислення метрик прецизійності, повноти, F1, ROC-AUC, середньої прецизійності та *WDE* на спільній тестовій вибірці; побудова графіків ROC-кривих, кривих повноти та прецизійності, матриць помилок, порівняльної діаграми метрик та аналітичного графіка залежності *WDE* від *IR*.

Для стратегії зважених класів використовується метод *compute_sample_weight* з параметром *balanced*, що обчислює ваги зразків, обернено пропорційні до частоти їх класу. Це є рівнозначним до параметра *class_weight ='balanced'*, який MLPClassifier не підтримує напряму.

Відтворюваність результатів забезпечується фіксацією генератора псевдовипадкових чисел (*random_state* = 42) на всіх етапах: розбитті вибірки, застосуванні SMOTE та андерсемплінгу, а також ініціалізації ваг нейронної мережі.

У дослідженні порівнюються чотири стратегії балансування:

1. Без балансування: базова модель, навчена на оригінальній незбалансованій вибірці;
2. Зважені класи: навчання з урахуванням sample_weight, де кожному зразку мінорного класу присвоюється вага, обернено пропорційна до частоти його класу;
3. SMOTE: синтетична надвибірка мінорного класу до рівноважного розподілу;
4. Андерсемплінг: випадкове скорочення мажоритарного класу до розміру мінорного.

Наявні стандартні метрики (F1-міра, ROC-AUC) не враховують безпосередньо ступінь дисбалансу при порівнянні різних методів балансування. Зокрема, дві моделі з однаковими значеннями F1 на датасетах з різним *IR* не є рівнозначними з точки зору складності задачі.

Пропонується метрика зваженої ефективності виявлення (Weighted Detection Efficiency, *WDE*): (β × Повнота + (1 - β) × Прецизійність) / (1 + α × IR), де параметри метрики мають наступний зміст (табл. 1):

Таблиця 1

Параметри метрики *WDE*

Параметр	Значення	Опис
Повнота	—	Частка виявлених шахрайських транзакцій
Прецизійність	—	Частка справді шахрайських серед позначених
IR	577,88	Коефіцієнт дисбалансу класів (кількість легітимних / шахрайських)
β (бета)	0,7	Вага повноти, що відображає пріоритет виявлення шахрайства над зниженням хибних спрацювань
α (альфа)	0,001	Коефіцієнт штрафу за дисбаланс, що регулює вплив *IR* на фінальну оцінку

Джерело: складено авторами на основі [13]

Знаменник (1 + α × IR) виступає штрафним множником: чим вищий дисбаланс класів, тим нижча фінальна оцінка *WDE* при однаковій якості класифікатора. Це дозволяє коректно порівнювати методи балансування навіть між датасетами з різним рівнем дисбалансу, нормалізуючи оцінку відносно складності задачі.

Значення *WDE*, близьке до 1, свідчить про високу виявлювальну здатність класифікатора в умовах сильного дисбалансу. Зменшення *WDE* вказує або на деградацію якості, або на підвищення дисбалансу класів.

Важливим етапом впровадження метрики *WDE* є вибір її вагових коефіцієнтів. Вибір значення параметра *β* обґрунтовується через вартісну модель операційних втрат. Співвідношення середньої вартості пропущеного шахрайства та вартості обробки хибної тривоги становить *Cfraud / Creview* = 122,21 / 5,00 ≈ 24,4. Це означає, що пропуск одного шахрайства обходиться фінансовій установі приблизно у 24 рази дорожче, ніж хибне спрацювання. Виходячи з цього, повнота має отримати значно вищу вагу у формулі. Теоретично максимальне значення *β*, обчислене як *Cfraud / (Cfraud + Creview)*, складає 0,96. Однак прийняте у роботі значення *β* = 0,7 є свідомо консервативним: воно відображає пріоритет повноти над прецизійністю, водночас не знецінюючи прецизійність повністю. Це відповідає практиці реальних систем моніторингу, де надмірна кількість хибних тривог створює операційне навантаження, яке також має враховуватись.

Вибір значення параметра α визначає силу штрафу за дисбаланс. При *α* = 0,001 і *IR* = 577,88 штрафний множник знаменника становить 1,578, тобто всі оцінки *WDE* знижуються приблизно на 37% порівняно з незваженою метрикою. Збільшення *α* до 0,005 зробило б штраф надмірним (WDE ≈ 0,20 для всіх стратегій), а зменшення до 0,0005 — недостатнім (*WDE* ≈ 0,61). Значення *α* = 0,001 забезпечує змістовний діапазон *WDE* ∈ [0,40; 0,51] для досліджуваних стратегій, що дозволяє чітко диференціювати їх між собою.

Для підтвердження стійкості результатів виконано аналіз чутливості: ранжування стратегій за *WDE* зберігається при варіюванні β у діапазоні [0,5; 0,8]. При *β* = 0,9 перші позиції займають зважені класи та SMOTE, що відповідає логіці: за максимального пріоритету повноти перевагу отримують стратегії з вищою повнотою. Це підтверджує, що вибір конкретного значення *β* має узгоджуватись із операційними пріоритетами конкретної фінансової установи.

Хоча в рамках даного дослідження всі стратегії оцінюються на одному датасеті з фіксованим *IR*, метрика *WDE* спроєктована для ширшого застосування, зокрема, для коректного порівняння результатів між різними датасетами або відтворення результатів з літератури, де рівень дисбалансу може суттєво відрізнятись.

Для оцінювання практичної придатності стратегій балансування з точки зору фінансової установи введено вартісну модель операційних втрат. Модель враховує два типи витрат:

1. Втрати від пропущених шахрайств, тобто середня сума шахрайської транзакції прийнята рівною $122.21, що відповідає значенню, задокументованому у відкритому датасеті банківських транзакцій;

2. Витрати на обробку хибних тривог, тобто вартість ручної перевірки одного помилкового спрацювання прийнята рівною $5.00, що відповідає оцінкам операційних витрат підрозділів фінансового моніторингу, наведеним у літературі.

Загальні операційні втрати для тестової вибірки обчислюються як:

$$Losses = FN \times Cfraud + FP \times Creview,$$

де *FN*— кількість пропущених шахрайських транзакцій, *FP* — кількість хибних спрацювань, *Cfraud* = $122.21, *Creview* = $5.00.

Зведені результати експериментів для всіх чотирьох стратегій балансування наведені у табл. 2.

Таблиця 2

Результати класифікації MLP за різними стратегіями балансування

Стратегія балансування	Прецизійність	Повнота	F1-міра	ROC-AUC	*WDE*
Без балансування	0,8261	0,7755	0,8000	0,9799	0,5011
Зважені класи	0,2762	0,8878	0,4213	0,9795	0,4463
SMOTE	0,6721	0,8367	0,7455	0,9677	0,4990
Андерсемплінг	0,0172	0,9082	0,0338	0,9337	0,4062

Джерело: складено авторами на основі [13]

Базова модель (без балансування) демонструє найвищі значення прецизійності (0,826) та ROC-AUC (0,9799), однак повнота складає лише 0,776, що означає пропуск близько чверті шахрайських транзакцій. Для систем виявлення шахрайських транзакцій такий результат є прийнятним, але не оптимальним, бо критичною є мінімізація пропущених випадків.

Стратегія зважених класів підвищує повноту до 0,8878, проте різко знижує прецизійність до 0,2762 і таким чином модель генерує велику кількість хибнопозитивних рішень. Метод SMOTE забезпечує найбільш збалансований розподіл між прецизійністю (0,672) і повнотою (0,837) серед методів балансування, що підтверджується найвищим значенням F1-міри (0,746) серед тих, що застосовували балансування. Андерсемплінг дає найвищу повноту (0,9082), але критично низьку прецизійність (0,0172) через суттєве скорочення навчальної вибірки.

Найвищий *WDE* має базова модель (0,5011), що пояснюється її значно вищою прецизійністю. Проте цей результат є певною мірою оманливим: висока *WDE* досягається ціною низької повноти. Серед методів балансування найвищий *WDE* демонструє SMOTE (0,4990).

Таким чином, *WDE* виявляє важливу закономірність: при штрафуванні на дисбаланс базова модель та SMOTE демонструють практично ідентичні значення *WDE* (0,5011 та 0,4990 відповідно), тоді як андерсемплінг суттєво відстає (0,4062).

Рис. 5 демонструє залежність *WDE* від *IR* для всіх стратегій при фіксованих значеннях прецизійності та повноти. Зі зростанням *IR* криві *WDE* монотонно спадають, що ілюструє зростання складності задачі. Стратегії з вищою повнотою (зважені класи, андерсемплінг) мають вищий початковий рівень *WDE* при низьких значеннях *IR*, однак при реалістичних рівнях дисбалансу (*IR* > 100) різниця між стратегіями нівелюється.

Оскільки в рамках даного експерименту *IR* є константою для всіх стратегій, ранжування за *WDE* збігається з ранжуванням за зваженою F-мірою, що підтверджує внутрішню узгодженість метрики. Відмінність *WDE* від стандартних метрик проявляється при міждатасетному порівнянні, де нормалізація на *IR* стає принциповою.

Для ілюстрації міждатасетних можливостей метрики *WDE* наведемо порівняння з результатами роботи [9], де задача виявлення шахрайства вирішувалась на датасеті IEEE-CIS Fraud Detection — наборі даних e-commerce транзакцій з принципово іншим рівнем дисбалансу (*IR* = 24,25, частка шахрайства 3,96%). Нейромережева модель GRU у вказаній роботі досягла прецизійності 0,28 та повноти 0,54 для класу шахрайства. Попри те, що ці значення значно поступаються результатам нашого MLP+SMOTE (прецизійність 0,672, повнота 0,837), безпосереднє порівняння без нормалізації на дисбаланс було б некоректним: задача на IEEE-CIS є суттєво простішою через значно нижчий *IR*. *WDE* для GRU на IEEE-CIS становить 0,4511, тоді як *WDE* нашого MLP+SMOTE цей показник становить 0,4990. Попри те що задача на IEEE-CIS є простішою (*IR* = 24,25 проти 577,88), нормалізація на *IR* дозволяє

коректно зіставити результати: незначна перевага нашої моделі за *WDE* відображає саме вищу якість класифікатора, а не лише різницю у складності задач.

Практична цінність метрики *WDE* виявляється при аналізі її поведінки в умовах зміни рівня дисбалансу класів, що відповідає реальним операційним сценаріям банківської діяльності. Розглянемо три характерні сценарії (табл. 3).

Оскільки прецизійність і повнота класифікатора визначаються архітектурою моделі та навчальною вибіркою, а не поточним розподілом вхідних транзакцій, для сценарного аналізу використовуються фіксовані значення метрик з табл. 2. Варіюється лише *IR*, що відображає зміну операційного навантаження. Значення *IR* для кожного сценарію обрані аналітично як репрезентативні орієнтири, а не отримані з реальних спостережень.

Таблиця 3

Значення *WDE* для різних операційних сценаріїв

Стратегія	Звичайний день (IR = 577,88)	Пікові навантаження (IR = 2 500)	Хакерська атака (IR = 28)
Без балансування	0,5011	0,2259	0,7691
Зважені класи	0,4464	0,2012	0,6851
SMOTE	0,4990	0,2249	0,7659
Андерсемплінг	0,4062	0,1831	0,6234

Джерело: складено авторами на основі [13]

Базовий сценарій звичайного операційного дня (*IR* = 577,88) відповідає реальному розподілу транзакцій у датасеті. Базова модель та SMOTE демонструють практично ідентичні *WDE* (0,5011 та 0,4990).

У періоди пікового навантаження (сезонні розпродажі, святкові дні, *IR* = 2 500) кількість легітимних транзакцій різко зростає, тоді як абсолютна кількість шахрайських залишається відносно стабільною. (*IR* збільшується приблизно у 4 рази. Класичні метрики (F1, прецизійність) залишаються незмінними, оскільки якість класифікатора не змінилась. Однак *WDE* для всіх стратегій падає приблизно вдвічі — до діапазону 0,18–0,23. Це є коректним сигналом: задача стала суттєво складнішою, і та сама модель об'єктивно забезпечує нижчий рівень захисту в абсолютному вимірі. Підрозділ ризик-менеджменту отримує обґрунтований індикатор для посилення ручного контролю у цей період.

При скоординованій хакерській атаці кількість шахрайських транзакцій різко зростає, що знижує (*IR* до рівня ~28. *WDE* всіх стратегій зростає до діапазону 0,62–0,77 — не через те, що модель стала кращою, а тому, що задача стала простішою: шахрайські патерни з'являються значно частіше і їх легше виявити. Це дає важливий практичний висновок: зростання *WDE* під час атаки не є підставою для зниження пильності, а навпаки вказує на аномальну зміну розподілу даних, яка сама по собі є сигналом інциденту безпеки.

Таким чином, динамічне відстеження *WDE* у реальному часі дозволяє підрозділам фінансового моніторингу відрізняти природну зміну операційного навантаження від реальної деградації якості моделі, що є суттєвою перевагою над статичними метриками класифікації.

Результати вартісного аналізу для тестової вибірки (56 962 транзакції, 98 шахрайських) наведені у табл. 4.

Таблиця 4

Операційні втрати за стратегіями балансування

Стратегія	Пропущено шахрайств	Хибні тривоги	Втрати від шахрайств, $	Витрати на перевірку, $	Загальні втрати, $
Без балансування	22	16	2 688,62	80,00	2 768,62
Зважені класи	11	228	1 344,31	1140,00	2 484,31
SMOTE	16	40	1 955,36	200,00	2 155,36
Андерсемплінг	9	5 076	1 099,89	25 380,00	26 479,89

Джерело: складено авторами на основі [13]

Вартісний аналіз суттєво змінює інтерпретацію результатів порівняно з класичними метриками. Метод SMOTE демонструє мінімальні загальні втрати ($2 155) завдяки оптимальному балансу між виявленням шахрайств і рівнем хибних тривог. Базова модель без балансування займає другу позицію ($2 769): вона генерує лише 16 хибних тривог, проте пропускає 22 шахрайські транзакції.

Найбільш наочним є результат андерсемплінгу: попри найвищу повноту і найменші втрати від пропущених шахрайств ($1 100), операційні витрати на обробку понад 5 000 хибних тривог складають $25 380, що робить цей метод фінансово непридатним.

Таким чином, оптимальна стратегія залежить від співвідношення *Cfraud / Creview* у конкретній установі. При типовому для роздрібного банкінгу співвідношенні ~24:1 перевагу має SMOTE як метод що забезпечує найкращий баланс між виявленням шахрайств і операційними витратами. За умови вищої середньої суми шахрайства, наприклад, у сегменті корпоративних транзакцій, рівновага зміщується на користь зважених класів, які забезпечують найнижчі втрати від пропущених шахрайств при прийнятному рівні хибних тривог.

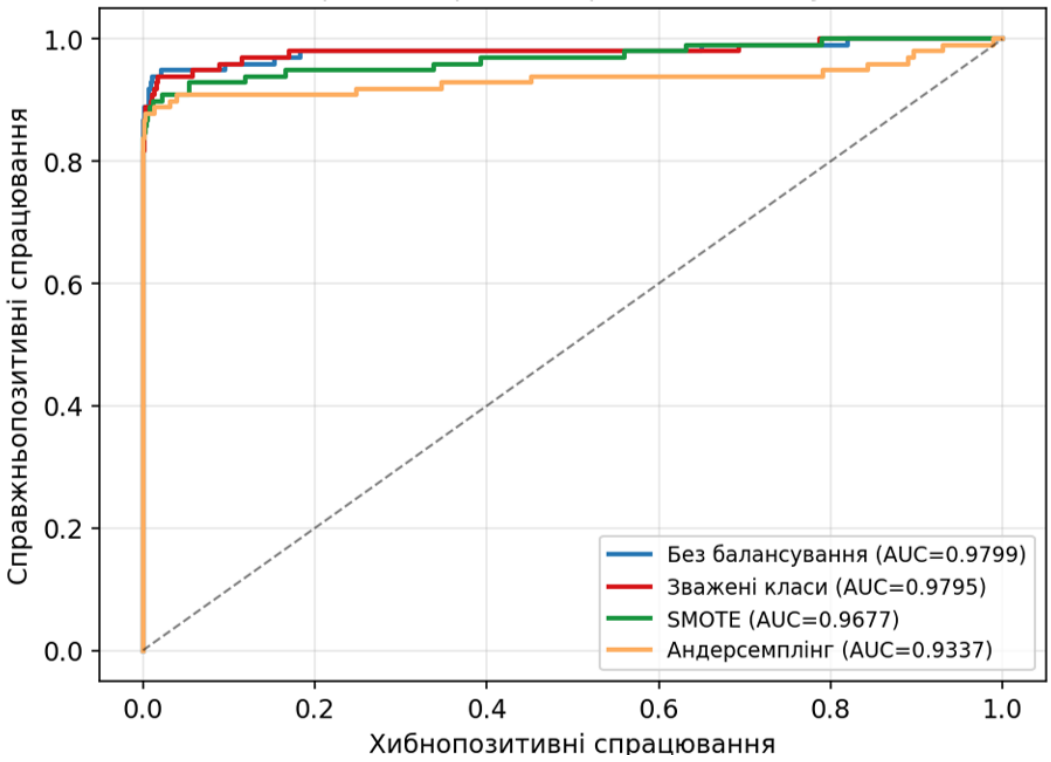

Рис. 1. ROC-криві для різних стратегій балансування
Джерело: складено авторами на основі [13]

ROC-криві (рис. 1) демонструють, що базова модель та зважені класи мають найвищі значення AUC (0,9799 та 0,9795), тобто практично ідентичні значення. SMOTE та андерсемплінг поступаються за AUC, що свідчить про часткову деградацію дискримінативної здатності моделі внаслідок зміни навчальної вибірки.

Криві прецизійності та повноти (рис. 2) є більш інформативними при сильному дисбалансі класів. Базова модель має найвищу середню точність (AP = 0,8376), проте крива обривається при низьких значеннях повноти, підтверджуючи неспроможність моделі виявляти більшість шахрайств. Метод SMOTE демонструє більш рівномірний хід кривої у широкому діапазоні повноти.

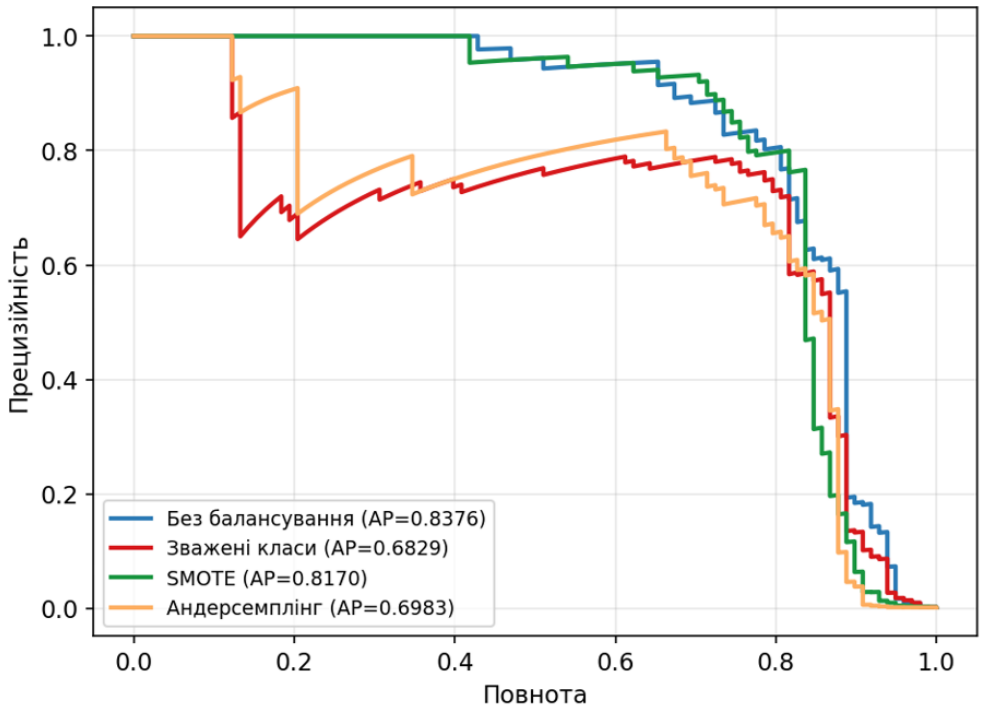

Рис. 2. Криві прецизійності та повноти для різних стратегій балансування
Джерело: складено авторами на основі [13]

Матриці помилок (рис. 3) наочно ілюструють компроміс між прецизійністю і повнотою для кожної стратегії. Базова модель пропускає 22 шахрайські транзакції із 98 у тестовій вибірці. Натомість андерсемплінг виявляє практично всі шахрайства (89 з 98), але генерує понад 5 000 хибних спрацювань, що є неприйнятним рівнем для операційної системи банку.

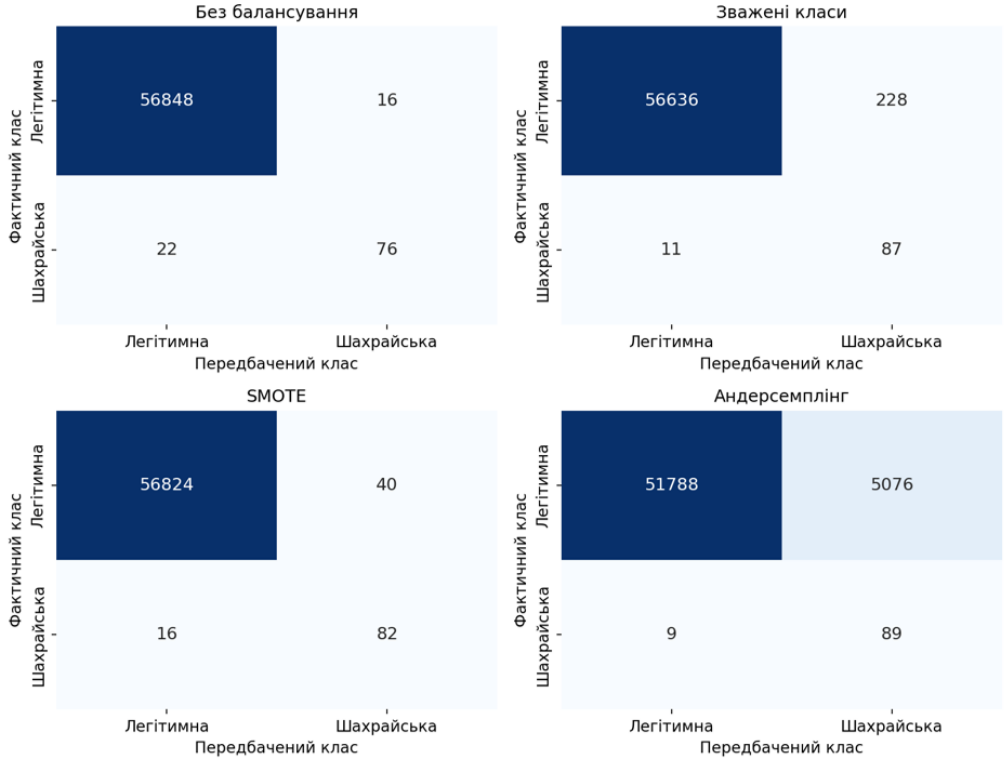

Рис. 3. Матриці помилок для всіх стратегій балансування
Джерело: складено авторами на основі [13]

Стовпчикова діаграма (рис. 4) узагальнює результати по всіх метриках. Візуально підтверджується відсутність домінуючої стратегії: базова модель лідирує за прецизійністю, ROC-AUC та *WDE*, тоді як методи балансування забезпечують вищу повноту за рахунок прецизійності. Компроміс між цими показниками є ключовою практичною проблемою при проєктуванні систем виявлення шахрайських транзакцій.

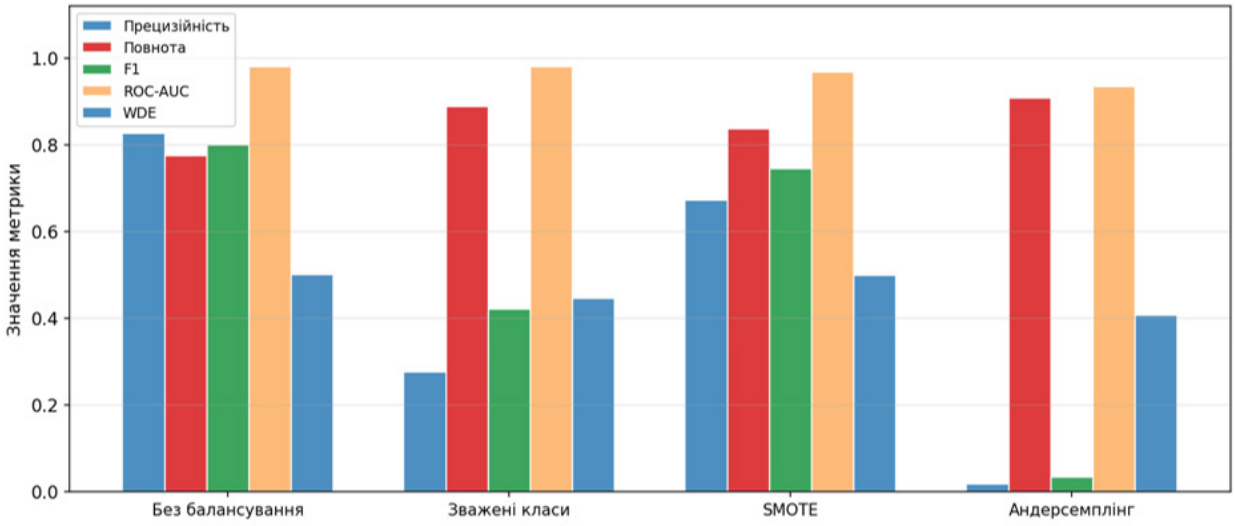

Рис. 4. Зведене порівняння метрик якості класифікатора за стратегіями балансування
Джерело: складено авторами на основі [13]

Залежність метрики *WDE* від коефіцієнта дисбалансу класів показана на рис. 5.

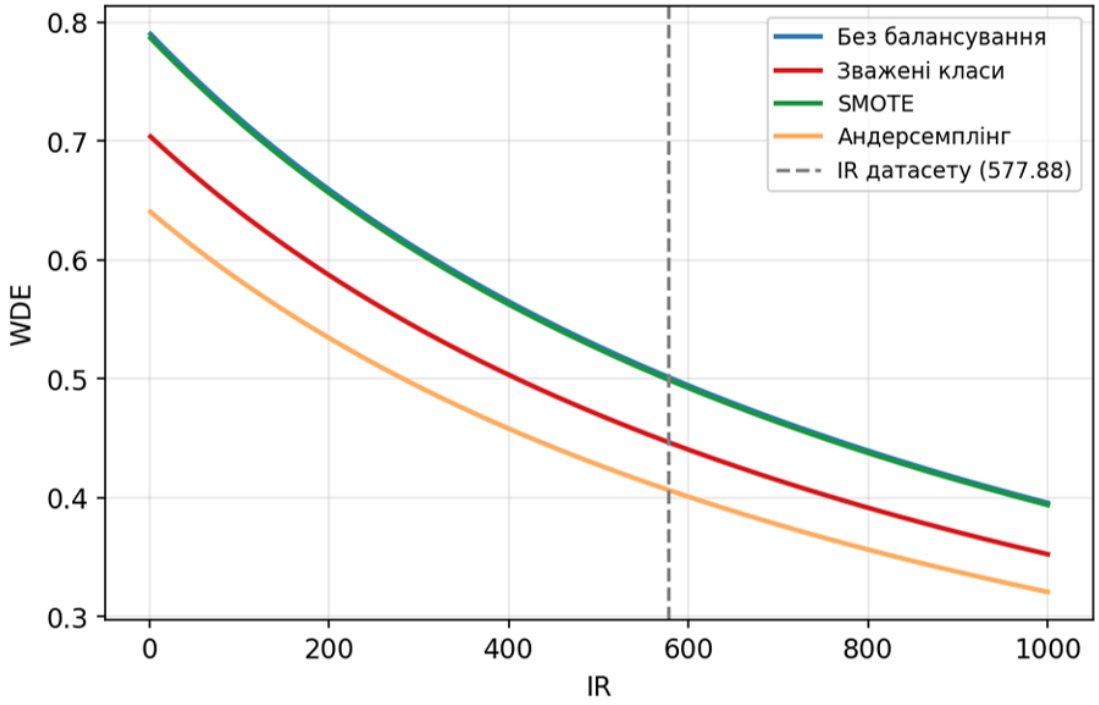

Рис. 5. Залежність метрики *WDE* від коефіцієнта дисбалансу класів (*IR*)
Джерело: складено авторами на основі [13]

Висновки. Отримані результати підтверджують відомий компроміс між прецизійністю і повнотою у задачах виявлення шахрайства. Жодна зі стратегій балансування не є абсолютно переважаючою, таким чином, вибір оптимального підходу залежить від операційних вимог конкретної фінансової установи.

Якщо пріоритетом є мінімізація збитків від пропущених шахрайств, то перевагу отримують методи з вищою повнотою (андерсемплінг, зважені класи). Якщо ж критичним є обмеження навантаження на підрозділ моніторингу, то перевагу має модель з вищою прецизійністю (базова або SMOTE).

Запропонована метрика *WDE* дає змогу ввести єдиний кількісний показник для порівняння стратегій з урахуванням дисбалансу. На відміну від F1-міри, *WDE* явно штрафує результати при високих значеннях *IR*, що відображає реальну складність задачі. Аналіз залежності *WDE* від *IR* (рис. 5) демонструє, що при збільшенні дисбалансу понад $IR \approx 200$ усі стратегії демонструють значне зниження ефективності, що вказує на необхідність використання складніших підходів, як-от ансамблевих методів або порогової оптимізації.

Додаткова практична цінність метрики *WDE* в умовах операційного моніторингу полягає в наступному. У періоди пікових навантажень (наприклад, сезонні розпродажі) природний коефіцієнт дисбалансу (*IR*) може різко зростати через наплив легітимних транзакцій. У таких умовах класичні метрики (F1, прецизійність) неминуче деградують, хибно сигналізуючи про «злам» моделі. Завдяки нормалізації на *IR*, метрика *WDE* дозволить підрозділам ризик-менеджменту відрізнити природну зміну розподілу даних від реальної втрати дискримінативної здатності нейромережі.

Слід зазначити, що архітектура MLP у даному дослідженні є цілеспрямовано спрощеною для забезпечення відтворюваності та порівнянності результатів, що залишає простір для подальшого вдосконалення за рахунок складніших архітектур.

У даній роботі проведено порівняльний аналіз чотирьох стратегій балансування класів при навчанні MLP-класифікатора для задачі виявлення шахрайських банківських транзакцій. Встановлено, що:

1. Базова модель без балансування досягає найвищих значень прецизійності (0,826) та ROC-AUC (0,9799), проте демонструє низьку повноту (0,776);
2. Метод SMOTE забезпечує найкращий баланс між прецизійністю та повнотою серед методів балансування, що підтверджується значенням $F1 = 0{,}746$;
3. Андерсемплінг досягає найвищої повноти (0,9082), але є непридатним для практичного застосування через надмірну кількість хибнопозитивних спрацювань;
4. Стратегія зважених класів займає проміжну позицію, демонструючи помірну повноту при суттєвому зниженні прецизійності.
5. Вартісний аналіз операційних втрат показав, що андерсемплінг, попри найвищу повноту, є фінансово непридатним через 5076 хибних тривог на тестовій вибірці; оптимальним з точки зору мінімізації сукупних витрат виявився метод SMOTE.

Запропонована метрика зваженої ефективності виявлення (*WDE*) є ключовим елементом наукової новизни роботи. *WDE* нормалізує якість класифікатора відносно коефіцієнта дисбалансу класів (*IR*), що дозволяє здійснювати коректне порівняння стратегій балансування незалежно від рівня дисбалансу в датасеті. Аналіз залежності *WDE* від *IR* показав, що при реалістичних рівнях дисбалансу ($IR > 400$) жодна зі стратегій не досягає $WDE > 0{,}5$, що вказує на необхідність подальших досліджень у напрямку ансамблевих та гібридних підходів до балансування.

Перспективними напрямами подальших досліджень є: застосування ансамблевих методів (XGBoost, Random Forest) та глибинних архітектур (LSTM, Autoencoder) з метою перевірки застосовності *WDE* для ширшого класу класифікаторів; розширення вартісної моделі шляхом врахування Customer Lifetime Value та сегментації транзакцій за сумою і типом; валідація метрики *WDE* на додаткових датасетах з різним рівнем дисбалансу для емпіричного підтвердження її міждатасетних властивостей; а також дослідження можливості динамічного калібрування параметра β на основі поточного співвідношення операційних витрат конкретної фінансової установи, що перетворить *WDE* на адаптивний інструмент ризик-менеджменту.

Література

1. Payment Card Fraud Worldwide. *Nilson Report*. 2025. Vol. 1298. URL: https://nilsonreport.com/newsletters/1298 (дата звернення: 23.04.2026).

2. Шахрайство з платіжними картками у 2025 році: кількість випадків знизилася, сума збитків — зросла. Національний банк України. URL: https://bank.gov.ua/ua/news/all/shahraystvo-z-platijnimi-kartkami-u-2025-rotsi-kilkist-vipadkiv-znizilasya-suma-zbitkiv--zrosla (дата звернення: 23.04.2026).

3. CNP Fraud Prevention for Issuers. *Nilson Report*. URL: https://nilsonreport.com/articles/cnp-fraud-prevention-for-issuers (дата звернення: 23.04.2026).

4. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016. 800 p.

5. Dal Pozzolo A., Caelen O., Johnson R. A., Bontempi G. Calibrating Probability with Undersampling for Unbalanced Classification. *2015 IEEE Symposium Series on Computational Intelligence (SSCI)*. Cape Town, South Africa, 2015. P. 159–166. DOI: https://doi.org/10.1109/SSCI.2015.33 (дата звернення: 23.04.2026).

6. He H., Garcia E. A. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*. 2009. Vol. 21, No. 9. P. 1263–1284. DOI: https://doi.org/10.1109/TKDE.2008.239 (дата звернення: 23.04.2026).

7. Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 2002. Vol. 16. P. 321–357. DOI: https://doi.org/10.1613/jair.953 (дата звернення: 23.04.2026).

8. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830.

9. Lemaître G., Nogueira F., Aridas C. K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*. 2017. Vol. 18, No. 17. P. 1–5.

10. Phua C., Lee V., Smith K., Gayler R. A Comprehensive Survey of Data Mining-based Fraud Detection Research // arXiv. 2010. DOI: https://doi.org/10.48550/arXiv.1009.6119 (дата звернення: 23.04.2026).

11. Ling C. X., Li C. Data Mining for Direct Marketing: Problems and Solutions. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*. New York, USA, 1998. P. 73–79.

12. Wang C., Nie C., Liu Y. Evaluating Supervised Learning Models for Fraud Detection: A Comparative Study of Classical and Deep Architectures on Imbalanced Transaction Data. *Proceedings of the International Conference on Management Innovation and Economic Development (MIED 2025)*. 2025. DOI: https://doi.org/10.2991/978-94-6463-835-6_65 (дата звернення: 23.04.2026).

13. Credit Card Fraud Detection Dataset / U. L. Leborgne, G. Bontempi; Machine Learning Group, Université Libre de Bruxelles. *Kaggle*, 2023. – URL: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (дата звернення: 23.04.2026).

References

1. Payment Card Fraud Worldwide. *Nilson Report*. 2025. Vol. 1298. Available at: https://nilsonreport.com/newsletters/1298 (accessed: 23.04.2026).

2. Shakhraistvo z platizhnymy kartkamy u 2025 rotsi: kilkist vypadkiv znyzylasia, suma zbytkiv — zrosla [Payment card fraud in 2025: fewer cases, higher losses]. Natsionalnyi bank Ukrainy. Available at: https://bank.gov.ua/ua/news/all/shahraystvo-z-platijnimi-kartkami-u-2025-rotsi-kilkist-vipadkiv-znizilasya-suma-zbitkiv--zrosla (accessed: 23.04.2026).

3. CNP Fraud Prevention for Issuers. *Nilson Report*. Available at: https://nilsonreport.com/articles/cnp-fraud-prevention-for-issuers (accessed: 23.04.2026).

4. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016. 800 p.

5. Dal Pozzolo A., Caelen O., Johnson R. A., Bontempi G. Calibrating Probability with Undersampling for Unbalanced Classification. *2015 IEEE Symposium Series on Computational Intelligence (SSCI)*. Cape Town, South Africa, 2015. P. 159–166. DOI: https://doi.org/10.1109/SSCI.2015.33 (accessed: 23.04.2026).

6. He H., Garcia E. A. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*. 2009. Vol. 21, No. 9. P. 1263–1284. DOI: https://doi.org/10.1109/TKDE.2008.239 (accessed: 23.04.2026).

7. Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 2002. Vol. 16. P. 321–357. DOI: https://doi.org/10.1613/jair.953 (accessed: 23.04.2026).

8. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830.

9. Lemaître G., Nogueira F., Aridas C. K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*. 2017. Vol. 18, No. 17. P. 1–5.

10. Phua C., Lee V., Smith K., Gayler R. A Comprehensive Survey of Data Mining-based Fraud Detection Research // arXiv. 2010. DOI: https://doi.org/10.48550/arXiv.1009.6119 (accessed: 23.04.2026).

11. Ling C. X., Li C. Data Mining for Direct Marketing: Problems and Solutions. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*. New York, USA, 1998. P. 73–79.

12. Wang C., Nie C., Liu Y. Evaluating Supervised Learning Models for Fraud Detection: A Comparative Study of Classical and Deep Architectures on Imbalanced Transaction Data. *Proceedings of the International Conference on Management Innovation and Economic Development (MIED 2025)*. 2025. DOI: https://doi.org/10.2991/978-94-6463-835-6_65 (accessed: 23.04.2026).

13. Credit Card Fraud Detection Dataset / U. L. Leborgne, G. Bontempi; Machine Learning Group,

Université Libre de Bruxelles. *Kaggle*, 2023. Available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (accessed: 23.04.2026).

Отримано: 24.04.2026
Прийнято до публікації: 31.05.2026
Опубліковано: 01.06.2026