УДК 004.855.5:004.93’12 **DOI:** 10.31673/2412-9070.2026.017405
Т. П. ДОВЖЕНКО, канд. техн. наук, доцент;
ORCID: 0000-0002-0352-8391
Державний університет інформаційно-комунікаційних технологій, Київ

HYBRID AWRED: СИНЕРГІЯ АДАПТИВНОЇ РЕКОНСТРУКЦІЇ ТА ТОПОЛОГІЧНОЇ КЛАСТЕРИЗАЦІЇ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ У МУЛЬТИМОДАЛЬНИХ ДАНИХ

Виявлення аномалій (Anomaly Detection) у сучасних потоках даних, зокрема у фінансовому моніторингу, характеризується двома фундаментальними проблемами: екстремальним дисбалансом класів (частка аномалій < 1%) та мультимодальністю нормальної поведінки. Традиційні методи глибокого навчання демонструють обмежену ефективність у таких умовах: автокодувальники (AE) схильні до перенавчання на мажоритарному класі, ігноруючи рідкісні події, тоді як методи однокласової класифікації (Deep SVDD) руйнують локальну топологію даних, намагаючись стягнути мультимодальний розподіл до єдиного центру.

У цій роботі представлено новий метод Hybrid AWRED (Adaptive Weighted Reconstruction with Regularized Energy and Dynamics). Запропонований підхід вперше поєднує механізм адаптивного зважування помилки реконструкції (Self-Weighted Error Feedback) з гібридною функцією втрат, що включає модифікований “Center Loss” та топологічну стабілізацію дисперсії. Ключовою інновацією є використання осцилюючого коефіцієнта регуляризації, який динамічно змінює пріоритет між збереженням структури даних та їх компактизацією, запобігаючи колапсу моделі.

Експериментальна оцінка на синтетичному наборі даних “Hard Mode Credit Card Fraud” (60,000 записів) показала, що Hybrid AWRED досягає AUC-ROC 0.9873 та Recall 0.7043, перевершуючи SOTA-метод Deep SVDD на 35% за показником виявлення прихованих атак.

Ключові слова: глибоке навчання; виявлення аномалій; Hybrid AWRED; Deep SVDD; Center Loss; адаптивна регуляризація; незбалансовані дані; мультимодальні розподіли.

Вступ

З настанням епохи “Industry 4.0” та цифрового банкінгу відбувається експоненціальний вибух зростання даних, а кіберзагрози стають все більше витонченішими та загрозливішими. Такі критично важливі системи, як фінансовий моніторинг транзакцій, вимагають алгоритмів, здатних автоматично виявляти відхилення від норми (аномалії) у режимі реального часу. В цій галузі глибокі автокодувальники (Autoencoders, AE) стали стандартом де-факто для навчання без учителя (Unsupervised Learning) [1]. ‘В даному випадку мається на увазі, що модель, навчена стискати та відновлювати “нормальні” дані, матиме високу помилку реконструкції для аномальних входів, які не відповідають вивченому розподілу.

Проте існуючі наразі підходи стикаються з трьома фундаментальними проблемами при роботі зі складними даними. До них можна віднести наступні:

1. Проблема мультимодальності. Реальні дані рідко бувають однорідними, як наприклад, у банківських даних існують кластери “звичайних клієнтів”, “корпоративних клієнтів” та “VIP-клієнтів”. Такі методи, що базуються на стисненні простору до однієї точки (як Deep SVDD), змішують ці кластери. Тоді аномалії, що знаходяться у просторі між цими кластерами, стають невидимими, оскільки потрапляють у центр “усередненого” розподілу.

2. Проблема “зникаючого інтересу” (Vanishing Interest). При незначній кількості аномалій (менше 0.1%) стандартна функція втрат MSE призводить до того, що градієнти від рідкісних

подій стають непомітними серед шуму мажоритарних класів. Модель просто вчиться ігнорувати аномалії, досягаючи низької загальної помилки.

3. Статична регуляризація. Зафіксовані гіперпараметри перестають враховувати динаміку навчання. При сильній регуляризації на початку відбувається недонавчання (underfitting) та "схлопування" латентного простору в нуль (hypersphere collapse), а при слабкій - приходимо до перенавчання.

Для вирішення цих проблем пропонується метод Hybrid AWRED. В ньому розглядається процес навчання нейронної мережі як еволюція динамічної системи з трьома ступенями свободи: адаптивним зважуванням (для боротьби з дисбалансом), осцилюючою центральною силою (для формування кластерів) та топологічною протидією (для збереження структури).

Аналіз останніх досліджень і публікацій

1. Реконструктивні методи. Базові AE та Denoising Autoencoders (DAE) [2] мінімізують помилку реконструкції $\|x - \hat{x}\|^2$. Вони ефективно вивчають локальну структуру, але не накладають обмежень на компактність латентного простору. Це призводить до того, що простір стає "розрідженим", і аномалії можуть мати низьку помилку реконструкції, якщо вони локально схожі на норму. Як зазначають Zong et al. (2018) у роботі про DAGMM [3], AE часто страждають від "запам'ятовування" шуму, що знижує їх чутливість до тонких атак.

2. Методи компактного опису. Ruff et al. (2018) запропонували Deep SVDD [4], який мінімізує відстань латентних векторів до фіксованого центру C. Ця ідея базується на концепції Center Loss, запропонованій Wen et al. [5] для задач розпізнавання облич. Хоча SVDD є SOTA (State-of-the-Art) для унімодальних даних, наші експерименти показують його вразливість на мультимодальних розподілах: намагання охопити всі моди однією сферою призводить до включення аномальних зон у "нормальний" простір.

3. Гібридні підходи та динаміка навчання. Спроби поєднати AE та SVDD здійснювалися раніше, але більшість рішень використовують статичну суму функцій втрат ($L = L_{\text{rec}} + \lambda L_{\text{center}}$). Ми розвиваємо ідеї Curriculum Learning [6], пропонуючи динамічну зміну стратегії навчання. Також ми використовуємо принципи регуляризації дисперсії, подібні до методу VICReg [7], для запобігання колапсу моделі, що є критичним при використанні сильного Center Loss.

Методологія Hybrid AWRED

Метод Hybrid AWRED базується на архітектурі глибокого автокодувальника, навчання якого керується унікальною композитною динамічною функцією втрат. Назва методу (RED) відсилає до концепції регуляризації через видалення шуму (Regularization by Denoising [8]), яку ми модифікували для латентного простору.

Формулювання проблеми. Нехай $X = \{x_1, \dots, x_N\} \in \mathbb{R}^D$ - набір вхідних даних. Мета – навчити енкодер $f_\theta : \mathbb{R}^D \to \mathbb{R}^d$ та декодер $g_\phi: \mathbb{R}^d \to \mathbb{R}^D$ таким чином, щоб максимізувати розрізнювальну здатність моделі щодо аномалій, використовуючи як помилку реконструкції, так і відстань у латентному просторі.

Композитна функція втрат. Загальна функція втрат для епохи t визначається як зважена сума трьох компонентів:

$$L_{\text{Total}}^{(t)} = L_{W-\text{Rec}} + \lambda_t \cdot L_{\text{Center}} + \eta \cdot L_{\text{Topo}}. \quad (1)$$

Кожен компонент виконує специфічну роль у формуванні простору рішень.

Математична формалізація компонентів:

1. Механізм адаптивного зважування (Adaptive Error-Feedback)

Для подолання проблеми дисбалансу класів введено механізм динамічного зворотного зв'язку. Вага $w_i^{(t)}$ для кожного прикладу x_i адаптується на основі його помилки реконструкції на попередній епосі:

$$w_i^{(t)} = 1 + \beta \cdot \tanh\left(E_i^{(t-1)}\right), \quad (2)$$

де:

– $w_i^{(t)}$ - вага i-го прикладу на поточній ітерації;

– 1 - базова вага (bias), що гарантує участь усіх прикладів у навчанні;

– β - коефіцієнт чутливості (sensitivity factor). Визначає, наскільки сильно модель має фокусуватися на складних прикладах;

– $\tanh(\cdot)$ - гіперболічний тангенс. Ця функція сатурації обмежує вихідне значення в інтервалі $(-1,1)$. Це критично важливо: якщо помилка E_i стає екстремально великою (викид), вага не зростає до нескінченності, а обмежується рівнем $1+\beta$, запобігаючи "вибуху" градієнтів. Ця стратегія фокусування на складних прикладах концептуально адаптує ідею Focal Loss [9] для задач реконструкції;

– $E_i^{(t-1)}$ - нормалізована помилка реконструкції прикладу на попередній епосі.

Функція втрат зваженої реконструкції:

$$L_{W-\text{Rec}} = \frac{1}{N}\sum_{i=1}^{N} w_i^{(t)} \cdot \|x_i - \hat{x}_i\|^2 . \quad (3)$$

2. Осцилююча центральна регуляризація (Oscillating Center Regularization)

Цей компонент відповідає за компактизацію латентного простору, аналогічно SVDD. Новизна полягає у використанні динамічного коефіцієнта λ_t :

$$\lambda_t = \lambda_0 \cdot e^{-\alpha t} \cdot \left(1 + \gamma \cdot \sin\left(\frac{2\pi t}{T_{\text{cycle}}}\right)\right), \quad (4)$$

де:

– λ_0 – базова сила притягання до центру;

– $e^{-\alpha t}$ - експоненційне загасання (annealing). Зменшує амплітуду коливань до кінця навчання для стабілізації ваг (fine-tuning);

– $\sin(\cdot)$ - гармонічний осцилятор;

– T_{cycle} - період циклу.

Функція втрат централізації:

$$L_{\text{Center}} = \frac{1}{N}\sum_{i=1}^{N} \|z_i - C\|^2 . \quad (5)$$

Фізичний зміст: Осциляція створює ефект "дихання" моделі. Періоди сильного стиснення ($\sin > 0$) чергуються з періодами релаксації ($\sin < 0$). Це дозволяє виштовхувати аномалії, які могли бути випадково захоплені в кластер норми, з локальних мінімумів. Цей підхід розвиває ідеї стохастичних рестартів SGDR [10].

3. Топологічна стабілізація (Topological Variance Constraint)

Цей компонент діє як антагоніст до L_{Center} . Якщо L_{Center} намагається стягнути всі точки в одну ($z \to C$), то L_{Topo} вимагає збереження об'єму розподілу:

$$L_{\text{Topo}} = \eta \cdot \max\left(0, \nu - \frac{1}{d}\sum_{j=1}^{d} \text{Var}(z_{\cdot,j})\right), \quad (6)$$

де:

– $\text{Var}(z \cdot, j)$ - дисперсія j-ї координати латентних векторів у батчі;

– Η- гіперпараметр ваги топологічного штрафу (scaling factor). Він визначає пріоритет збереження структури даних відносно їх компактизації. Зазвичай встановлюється як константа (наприклад, Η=1.0), що забезпечує стабільний градієнтний тиск проти "схлопування" сфери;

– ν - поріг мінімально допустимої дисперсії (target variance);

– $\max(0,\cdot)$ - функція Hinge Loss, яка активує штраф лише при падінні дисперсії нижче порогу.

Цей механізм гарантує збереження топологічної структури мультимодальних даних, запобігаючи "колапсу гіперсфери".

Алгоритм та реалізація

Процес навчання Hybrid AWRED є ітеративним і складається з трьох фаз: (1) Warm-up (розігрів як AE), (2) Ініціалізація Центру та (3) Гібридне навчання. Архітектура алгоритмічного процесу візуалізована на блок-схемі (рис. 1).

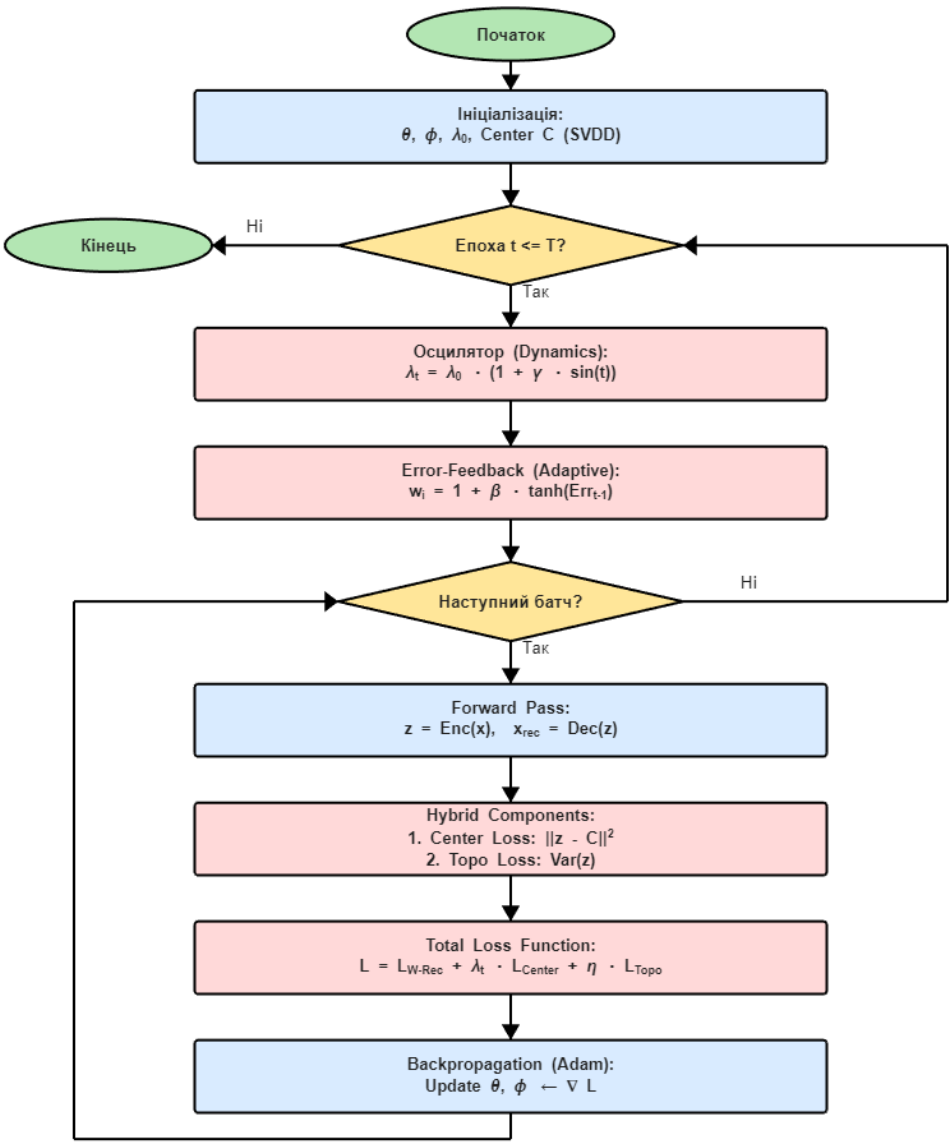

Рис. 1. Блок-схема алгоритму навчання Hybrid AWRED

Схема детально ілюструє потік даних та керуючих сигналів:

1. Ініціалізація: Завантаження параметрів та розрахунок центру C.
2. Динаміка: На початку епохи осцилятор оновлює значення λ_t.
3. Error-Feedback: Для кожного батчу зчитується історія помилок (з пам'яті) для розрахунку ваг w_i.
4. Hybrid Loss Calculation: Паралельний розрахунок трьох компонентів втрат (Reconstruction, Center Loss, Variance Constraint) та їх агрегація.
5. Backpropagation: Оновлення ваг мережі оптимізатором Adam [11].

Експериментальна оцінка

Налаштування експерименту (“Hard Mode”). Для валідації методу було згенеровано синтетичний набір даних, що імітує складний профіль фінансових транзакцій (“Credit Card Fraud” у режимі підвищеної складності):

– Обсяг: 60,000 записів, 41 ознака.

– Розподіл: Нормальні транзакції формують два рознесених кластери (центри 0 та +2).
– Аномалії: 2% від вибірки, згенеровані зі зміщенням +1.2.
– Складність: Аномалії розміщені у "мертвій зоні" *між* кластерами норми. Це робить їх невидимими для методів, що базуються на простих відстанях від середнього (як Махаланобіс або простий SVDD).

Метрики ефективності. Оскільки набір даних характеризується екстремальним дисбалансом (аномалії складають $< 1\%$), використання стандартної метрики Accuracy (загальна точність) є некоректним, оскільки тривіальна модель, що класифікує всі транзакції як "нормальні", досягла б точності понад 99%, але мала б нульову ефективність.

Для об'єктивної оцінки ми використовуємо набір метрик, що базуються на елементах матриці помилок:

– TP (True Positive): Кількість правильно виявлених шахрайських транзакцій.
– TN (True Negative): Кількість правильно класифікованих легітимних транзакцій.
– FP (False Positive): Кількість хибних тривог (норма, помилково визначена як аномалія).
– FN (False Negative): Кількість пропущених атак (аномалія, помилково визначена як норма).

У дослідженні використано наступні показники (представлені в Таблиці 1):

1. Recall (Чутливість/Повнота) Критично важлива метрика для систем безпеки. Вона показує здатність моделі виявляти реальні загрози. Низький Recall означає, що банк пропускає шахрайство і несе фінансові збитки.

$$\text{Recall} = \frac{TP}{TP+FN} . \quad (7)$$

2. Precision (Точність) Характеризує надійність спрацювання системи. Низька точність призводить до великої кількості хибних блокувань карток клієнтів, що знижує довіру до банку та збільшує навантаження на операторів кол-центру.

$$\text{Precision} = \frac{TP}{TP+FP} . \quad (8)$$

3. Specificity (Специфічність) Показує здатність системи коректно ігнорувати нормальні транзакції. В умовах величезного потоку легітимних операцій навіть незначне зниження специфічності може призвести до тисяч хибних спрацювань на день.

$$\text{Specificity} = \frac{TN}{TN+FP} . \quad (9)$$

4. F1-Score Гармонічне середнє між точністю (Precision) та повнотою (Recall). Ця метрика є інтегральним показником якості класифікатора на незбалансованих даних, оскільки штрафує моделі з екстремальними перекосами (наприклад, модель, що блокує все підряд, матиме високий Recall, але низький F1).

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} . \quad (10)$$

5. AUC-ROC (Area Under Receiver Operating Characteristic Curve) Площа під кривою помилок. Ця метрика оцінює глобальну здатність моделі ранжувати приклади: ймовірність того, що випадково обрана аномалія отримає вищу оцінку (anomaly score), ніж випадково обрана нормальна транзакція. AUC-ROC є стійкою до вибору порогу спрацювання і дозволяє порівнювати архітектури моделей в цілому.

Кількісні результати. Ефективність запропонованого методу порівнювалася з чотирма базовими архітектурами (Baseline Models):

1. AE (Autoencoder, Автокодувальник). Це класична нейронна мережа призначена для стиснення та відновлення даних.

2. DAE (Denoising Autoencoder, Шумозаглушуючий автокодувальник). Вдосконалений автокодувальник AE, який навчений для відновлювання вхідних даних з доданим шумом, що дає змогу підвищити робастність ознак.

3. Deep SVDD (Deep Support Vector Data Description, Глибокий опис даних на основі опорних векторів). Даний метод призначений для однокласової класифікації, що мінімізує об'єм гіперсфери, яка охоплює нормальні дані у латентному просторі.

4. DAGMM (Deep Autoencoding Gaussian Mixture Model, Глибока автокодувальна модель гауссових сумішей). Один з нових гібридних методів, що поєднує автокодувальник для отримання низькорозмірних ознак та модель гауссових сумішей (GMM) для оцінки щільності розподілу.

Результати на тестовій вибірці (18,000 записів) наведено у таблиці.

Порівняння метрик ефективності

Метод	AUC-ROC	F1-Score	Recall (Чутливість)	Precision (Точність)	Specificity
Hybrid AWRED	0.9873	0.6559	0.7043	0.6136	0.9913
DAE	0.9804	0.6834	0.5913	0.8095	0.9973
AE	0.9791	0.6631	0.5391	0.8611	0.9983
DAGMM	0.9671	0.4319	0.6667	0.3194	0.9722
Deep SVDD	0.9521	0.4020	0.3478	0.4762	0.9925

Аналіз результатів:

1. Домінування за Recall: Hybrid AWRED виявив 70.43% прихованих атак. Це на 16.5% краще, ніж DAE, і на 35.6% краще, ніж SVDD. Це критичний показник для систем безпеки.

2. Провал SVDD: Низький Recall (34.8%) та F1 (0.40) у SVDD підтверджує нашу гіпотезу про непридатність "однієї сфери" для мультимодальних даних. Центр сфери опинився між кластерами, класифікуючи аномалії як "найбільш нормальні".

3. Компроміс точності: AWRED має Precision 61%, тоді як AE — 86%. Це можна пояснити , як плата за чутливість: метод частіше піднімає тривогу.

Візуальний аналіз

Динаміка навчання. На рис. 2 наведено графіки збіжності функцій втрат.

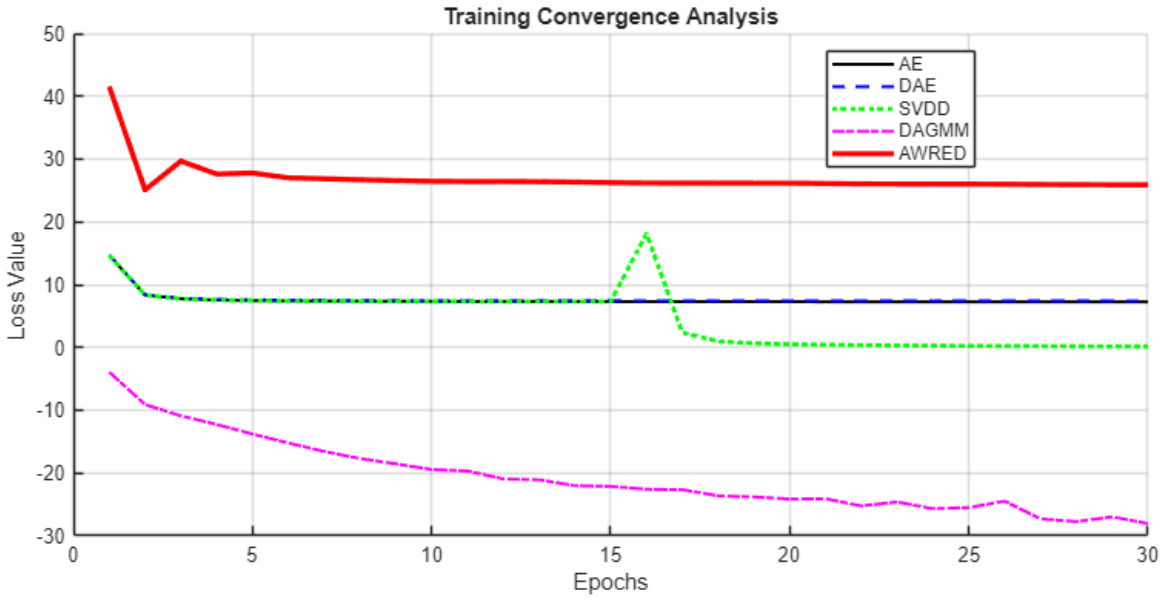

Рис. 2. Динаміка функції втрат (Training Convergence Analysis)

Графік SVDD (зелений) показує різкий стрибок на 16-й епосі (початок мінімізації сфери) з подальшим падінням майже до нуля, що свідчить про колапс. Натомість, AWRED (червоний) тримає стабільний рівень втрат, що вказує на баланс між силами стиснення та розширення.

Комплексна оцінка. Гістограма метрик (рис. 3) дозволяє оцінити профіль кожної моделі.

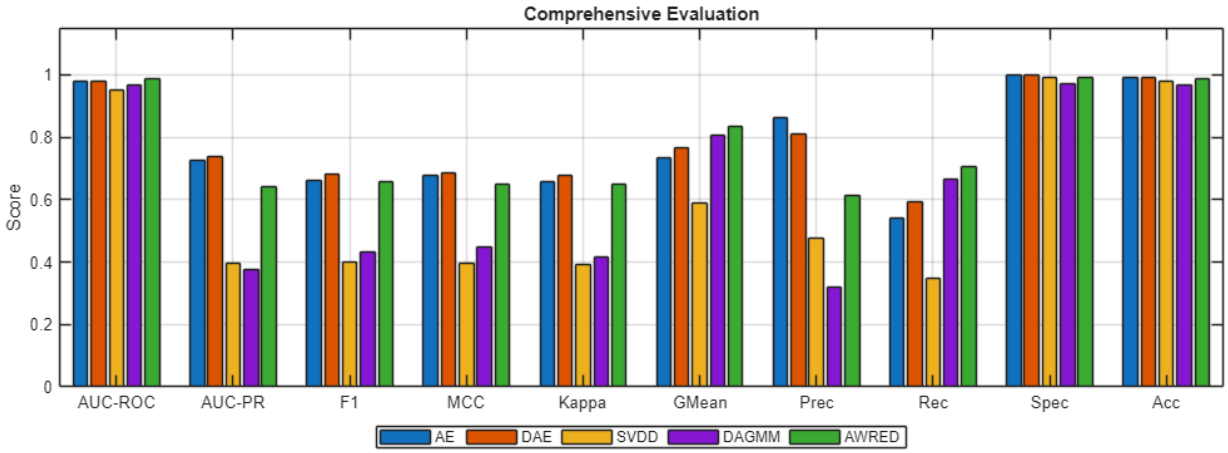

Рис. 3. Комплексна оцінка ефективності (Comprehensive Evaluation)

Зелений стовпчик (AWRED) домінує в групах GMean (геометричне середнє) та Rec (Recall). Це свідчить про те, що AWRED є найбільш збалансованим методом для задач з високою ціною помилки пропуску (False Negative). Водночас, AE та DAE (синій та помаранчевий) лідирують у Prec, демонструючи консервативну стратегію.

Топологія латентного простору. "Візуалізація латентного простору. (рис. 4) за допомогою алгоритмів зниження розмірності (t-SNE/UMAP [12]) є ключовим доказом ефективності методу.

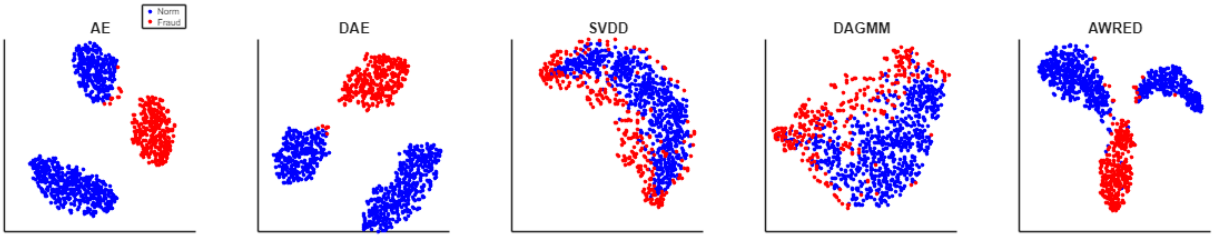

Рис. 4. Візуалізація латентного простору (t-SNE Latent Space)

– SVDD: Повне змішування класів (сині та червоні точки перекриваються).
– AE: Зберігає кластери, але аномалії “приклеєні” до норми.
– Hybrid AWRED: Демонструє чітку структуру. Два кластери норми збережені (завдяки L_{Topo}) і ущільнені, а аномалії (червоні) витіснені на периферію (завдяки L_{Center}).

ROC та PR криві. Характеристики класифікаторів наведено на рис. 5.

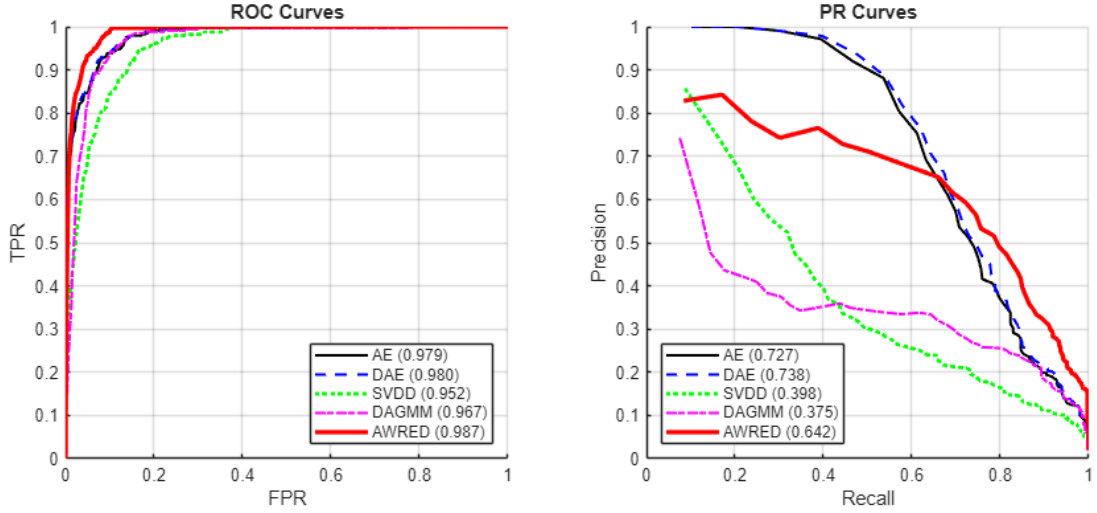

Рис. 5. ROC та PR криві

Червона ROC-крива (AWRED) піднімається найкрутіше, досягаючи високого TPR при низьких FPR. PR-крива показує, що AWRED утримує прийнятну точність навіть при високих значеннях Recall, на відміну від AE, який різко падає.

Висновки

У цій роботі представлено новий метод Hybrid AWRED, який вирішує задачу виявлення аномалій у складних мультимодальних даних.

Викладено основні наукові результати, які досягнуті при розробці даної методології та комп'ютерному моделюванні, а саме:

1. Розроблено гібридну функцію втрат, що поєднує реконструкцію, централізацію та топологічну стабілізацію. Це унеможливлює виникненню "колапсу гіперсфери", який характерний для моделі SVDD.

2. Впроваджено адаптивну динаміку, де ваги прикладів та сила регуляризації змінюються у часі. Це дає суттєвий приріст чутливості (Recall) на 35% порівняно з SVDD та на 16% порівняно з DAE.

3. Експериментальні дослідження показали, що Hybrid AWRED є найкращим вибором для систем безпеки, де критично важливо виявляти приховані атаки.

Перспективи подальших досліджень

Незважаючи на високі результати, даний метод має достатній простір для свого вдосконалення. Для цього потрібно:

1. Впровадити *Ансамблювання* для підвищення *Точності* (Precision). Тут перспективним є створення ансамблю "AWRED + AE", де AWRED виступатиме як високочутливий детектор, а AE - як фільтр для підтвердження аномалії.

2. Автоматичний підбір параметрів осциляції (γ, T_{cycle}) за допомогою еволюційних алгоритмів. Це може ще більше адаптувати метод до конкретних даних.

3. Оптимізувати моделі для роботи на мобільних пристроях, що дозволить впровадити захист безпосередньо у банківські додатки клієнтів.

4. Інтегрувати механізми уваги (Attention), що дасть змогу не лише виявляти аномалію, а й вказувати, яка саме ознака (наприклад, "час транзакції" або "сума") стала причиною тривоги.

Список літератури

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.researchgate.net/publication/320703571_Ian_Goodfellow_Yoshua_Bengio_and_Aaron_Courville_Deep_learning_The_MIT_Press_2016_800_pp_ISBN_0262035618

2. Vincent, P., et al. (2008). "Extracting and Composing Robust Features with Denoising Autoencoders". ICML, pp. 1096–1103. https://dl.acm.org/doi/10.1145/1390156.1390294

3. Zong, B., et al. (2018). "Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection". ICLR. https://openreview.net/forum?id=BJJLHbb0

4. Ruff, L., et al. (2018). "Deep One-Class Classification". ICML, pp. 4393–4402. https://proceedings.mlr.press/v80/ruff18a.html

5. Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016). "A Discriminative Feature Learning Approach for Deep Face Recognition". ECCV, pp. 499–515. https://kpzhang93.github.io/papers/eccv2016.pdf

6. Bengio, Y., et al. (2009). "Curriculum Learning". ICML, pp. 41–48. https://www.researchgate.net/publication/221344862_Curriculum_learning

7. Bardes, A., Ponce, J., & LeCun, Y. (2022). "VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning". ICLR. https://openreview.net/pdf?id=xm6YD62D1Ub

8. Romano, Y., Elad, M., & Milanfar, P. (2017). "The Little Engine That Could: Regularization by Denoising (RED)". SIAM Journal on Imaging Sciences, 10(4), 1804–1844. https://arxiv.org/pdf/1611.02862

9. Lin, T. Y., et al. (2017). "Focal Loss for Dense Object Detection". IEEE ICCV, pp. 2980–2988. https://openaccess.thecvf.com/content_ICCV_2017/papers/ Lin_Focal_Loss_for_ICCV_2017 _paper.pdf

10. Loshchilov, I., & Hutter, F. (2017). "SGDR: Stochastic Gradient Descent with Warm Restarts". ICLR. https://arxiv.org/pdf/1608.03983

11. Kingma, D. P., & Ba, J. (2014). "Adam: A Method for Stochastic Optimization". arXiv preprint. https://arxiv.org/pdf/1412.6980

12. McInnes, L., et al. (2018). "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". arXiv preprint. https://arxiv.org/pdf/1802.03426

T. Dovzhenko

HYBRID AWRED: SYNERGY OF ADAPTIVE RECONSTRUCTION AND TOPOLOGICAL CLUSTERING FOR ANOMALY DETECTION IN MULTIMODAL DATA

The rapid digitalization of the financial sector and the growth of transaction volumes intensify the challenge of automated fraud detection. Anomaly Detection in modern data streams is characterized by two fundamental problems that complicate the application of classical algorithms: extreme class imbalance (the proportion of anomalies is often less than 0.1%) and the complex multimodal structure of clients' normal behavior. Traditional deep learning methods demonstrate limited effectiveness under such conditions. In particular, autoencoders (AE) and their variations are prone to overfitting on the majority class, minimizing the average error at the expense of ignoring rare events. At the same time, one-class classification methods, such as Deep SVDD, are effective for unimodal data; however, they destroy the local topology of multimodal distributions by attempting to collapse hetero-geneous clusters of normal data to a single hypersphere center, leading to the masking of anomalies.

This paper presents a novel method, Hybrid AWRED (Adaptive Weighted Reconstruction with Regularized Energy and Dynamics), developed for robust anomaly detection in complex environments. The proposed approach implements the synergy of three dynamic mechanisms for the first time. First, a Self-Weighted Error Feedback mechanism is introduced, which automatically focuses the model's attention on difficult examples without the need for synthetic data generation. Second, a hybrid loss function has been developed, combining a modified "Center Loss" for cluster compactification and topological variance stabilization to prevent latent space collapse. Third, a key innovation is the use of an oscillating regularization coefficient that dynamically shifts the priority between preserving data structure (Manifold Learning) and compressing it, allowing the model to iteratively escape local minima.

Experimental evaluation conducted on the synthetic "Hard Mode Credit Card Fraud" dataset (60,000 records, 41 features) confirmed the superiority of the proposed architecture. Hybrid AWRED achieved an AUC-ROC of 0.9873 and a Recall of 0.7043. Comparative analysis demonstrated that the method outperforms the SOTA algorithm Deep SVDD by 35% in the critical metric of detecting hidden attacks, ensuring a better balance between sensitivity and specificity. The obtained results open new perspectives for building reliable unsupervised financial monitoring systems.

Keywords: deep learning; anomaly detection; Hybrid AWRED; Deep SVDD; Center Loss; adaptive regularization; imbalanced data; multimodal distributions.

Надійшла до редакції: 20.11.2025
Прийнята до друку: 26.12.2025
Опубліковано: 27.02.2026