

УДК 512.44

DOI [https://doi.org/10.24144/2616-7700.2026.49\(2\).245-254](https://doi.org/10.24144/2616-7700.2026.49(2).245-254)**С. В. Мошко¹, О. Ю. Кучанський²**

¹ Київський політехнічний інститут ім. Ігоря Сікорського,
магістрант факультету біомедичної інженерії
serhii.stud@gmail.com
ORCID: <https://orcid.org/0009-0002-6828-1833>

² Astana IT University, Астана, Республіка Казахстан,
професор школи штучного інтелекту та науки про дані,
доктор технічних наук, професор
kuchanskyi.o@gmail.com
ORCID: <https://orcid.org/0000-0003-1277-8031>

ATTENTION-VAE: НЕСУПЕРВІЗОВАНЕ РАННЄ ВИЯВЛЕННЯ СЕПСИСУ ЧЕРЕЗ ОЦІНКУ АНОМАЛІЙ НА ОСНОВІ РЕКОНСТРУКЦІЇ ВІТАЛЬНИХ ПОКАЗНИКІВ ВІДДІЛЕННЯ ІНТЕНСИВНОЇ ТЕРАПІЇ

У статті розглянуто можливості застосування несупервізованого глибокого навчання для раннього виявлення сепсису у відділеннях інтенсивної терапії на основі аналізу часових рядів вітальних показників. Для досягнення мети дослідження було поставлено та розв'язано такі завдання: огляд існуючих підходів до виявлення аномалій у медичних часових рядах; формальна постановка задачі виявлення сепсису як задачі детекції аномалій; розробка архітектури Attention-VAE, що поєднує рекурентний GRU-енкодер, механізм багатоголової уваги та варіаційний декодер. Архітектуру навчено виключно на записах пацієнтів із нормальним фізіологічним перебігом — без використання розмічених даних щодо сепсису. Результатом роботи моделі є показник аномальності, що відображає ступінь відхилення поточного стану пацієнта від засвоєних нормальних закономірностей, а також карти уваги та профілі реконструкційних похибок за окремими показниками, що забезпечують подвійну інтерпретованість рішень. Застосування моделі апробовано на базі клінічних даних PhysioNet Challenge 2019 з використанням п'яти вітальних показників у шестигодинних вікнах спостереження; досягнуто $F1 = 0,8286$, $ROC-AUC = 0,8893$, частку хибнопозитивних результатів 0,15% — показники, що демонструють якість одного рівня із супервізованими методами. Обмеженням дослідження є один датасет та необхідність мультицентрової валідації. Результати дослідження є складовою магістерської дисертації та можуть бути використані у системах клінічної підтримки прийняття рішень.

Ключові слова: варіаційний автоенкодер; механізм уваги; виявлення сепсису; виявлення аномалій у часових рядах; навчання без учителя; PhysioNet Challenge 2019; інтерпретоване машинне навчання.

1. Вступ. Сепсис залишається однією з провідних причин смертності у відділеннях інтенсивної терапії (ВІТ). За оцінками ВООЗ, щорічно реєструється близько 49 мільйонів випадків, з яких понад 11 мільйонів завершуються летально [1]. Кожна година затримки антибіотикотерапії підвищує ризик летального наслідку на 7–8% [2]. Відтак рання діагностика є критичною умовою ефективного лікування.

Традиційні підходи ґрунтуються на клінічних шкалах (SOFA, qSOFA) та лабораторних біомаркерах, проте мають суттєві обмеження: шкали є реактивними і фіксують стан, коли патологічний процес уже набув клінічної форми; лабораторні дані надходять із затримкою; дискретні порогові критерії не враховують

багатовимірну динаміку фізіологічних показників [3]. Водночас у сучасних ВІТ безперервно реєструються значні обсяги даних моніторингу — ЧСС, SpO₂, MAP, температура, RR — які містять ранні сигнали розвитку сепсису.

Методи глибокого навчання відкривають нові можливості для аналізу багатовимірних медичних часових рядів. Підходи на основі виявлення аномалій є перспективними для раннього виявлення сепсису, оскільки не потребують точних міток початку патологічного процесу [4]. Момент onset сепсису в ретроспективних базах визначається неоднозначно та залежить від протоколів конкретного закладу.

2. Огляд літератури. Прямими архітектурними попередниками Attention-VAE є дві роботи. OmniAnomaly [6] поєднує GRU з VAE, доповненим стохастичним з'єднанням прихованих змінних та planar normalizing flow; на трьох стандартних benchmark-датасетах (SMD, SMAP, MSL) досягнуто $F1 = 0,86$ — цим підтверджено ефективність GRU-VAE як базової архітектури для послідовних даних, що лягло в основу розробленої моделі. MA-VAE [8] є архітектурно найближчою до запропонованої: BiLSTM-VAE з механізмом multi-head attention для виявлення аномалій у даних автомобільного тестування; автори демонструють, що multi-head attention суттєво підвищує інтерпретованість рішень порівняно з базовим BiLSTM-VAE. Принципова відмінність Attention-VAE від MA-VAE — використання GRU замість BiLSTM (менша кількість параметрів при порівнянній якості моделювання) та застосування до біомедичного домену.

Систематичний огляд 80 досліджень [9] фіксує, що типові значення ROC-AUC для ML-систем виявлення сепсису знаходяться у діапазоні 0,79–0,96, при цьому переважна більшість підходів є супервізованими класифікаторами, що потребують розмічених даних; проблема відтворюваності залишається невирішеною через різноманітність використовуваних датасетів і визначень сепсису, що підкреслює гостру потребу в несупервізованих підходах. DTAE [10] є єдиною знайденою роботою, що поєднує несупервізований автоенкодерний підхід із задачею виявлення сепсису: модель використовує Transformer-based denoising autoencoder із детермінованим декодером, що позбавляє її можливості кількісно оцінювати невизначеність у латентному просторі — принципова відмінність від Attention-VAE.

Серед супервізованих підходів орієнтиром якості слугує робота [11], у якій LSTM-модель з механізмом уваги демонструє $AUC = 0,892$ на вибірці понад 100 000 пацієнтів; систематичний аналіз ваг уваги в цій роботі підтверджує можливість автоматичної ідентифікації клінічно значущих часових вікон — принципово схожий механізм інтерпретації реалізовано в Attention-VAE, але у несупервізованому контексті.

Табл. 1 систематизує порівняння розробленого підходу з найближчими аналогами за чотирма ключовими вимірами: архітектура, тип навчання, наявність механізму уваги та домен застосування.

Запропонована архітектура спирається на встановлені у літературі принципи поєднання GRU-VAE для несупервізованого виявлення аномалій у багатовимірних часових рядах [6] та інтеграції multi-head attention у VAE [8]. Внеском цієї роботи є адаптація такої комбінації до домену біомедичних часових рядів відділення інтенсивної терапії з додатковим механізмом подвійної інтерпретації — часової (карти уваги) та ознакової (декомпозиція реконструкційної похибки

Таблиця 1.

Порівняння з існуючими підходами

Модель	Архітектура	Unsup.	Attn.	Домен	Venue
OmniAnomaly [6]	GRU-VAE + NF	+	–	IT	KDD 2019
MA-VAE [8]	BiLSTM-VAE + MHA	+	+	Automotive	NCTA 2023
DTAE [10]	Transformer-DAE	+	+	Sepsis	AAAI 2022
Zhang et al. [11]	LSTM + Att	–	+	Sepsis	Patterns 2021
Attention-VAE (наш)	GRU-VAE + MHA	+	+	Sepsis / PhysioNet 2019	–

за каналами вітальних показників), орієнтованим на клінічні вимоги до пояснюваності рішень у медичних системах підтримки прийняття рішень.

3. Методологія дослідження.

3.1. Постановка задачі. Задачу формалізовано як виявлення аномалій у багатовимірних часових рядах. Нехай $X \in \mathbb{R}^{72 \times 5}$ – часове вікно ($W = 72$ кроки, 6 годин при 5-хвилинному інтервалі; $d = 5$ вітальних показників). Модель навчається виключно на нормальних спостереженнях $\mathcal{D}_{train} = \{X(i)|y(i) = 0\}$ та обчислює показник аномальності $s(X) \in \mathbb{R}_+$, де високі значення відповідають аномальним станам.

У межах дослідження запропоновано архітектуру Attention-VAE, яка складається з чотирьох компонентів:

GRU-енкодер перетворює вхідне вікно $X \in \mathbb{R}^{72 \times 5}$ на послідовність прихованих станів $H \in \mathbb{R}^{72 \times 64}$. Двошаровий GRU з dropout $p = 0,2$. Вибір GRU замість LSTM обумовлено меншою кількістю параметрів при порівнянні якості моделювання часових залежностей [13].

Механізм багатоголової уваги (4 голови) [7] формує зважене контекстне представлення прихованих станів:

$$\alpha_t = \text{softmax}(q^T k_t / \sqrt{d_k}), \quad c = \sum_{t=1} W \alpha_t h_t,$$

де q – вектор запиту, k_t – ключ для кроку t , d_k – розмірність ключа. Ваги α_t вказують, на які часові моменти модель звертає увагу при кодуванні ($\sum_a \alpha_a = 1$, $\alpha_a \geq 0$).

Варіаційний латентний шар формує параметри апостеріорного розподілу $q\phi(z|X) = \mathcal{N}(\mu, \text{diag}(\sigma^2))$, де $z \in \mathbb{R}^{16}$:

$$\mu = W_\mu \cdot \mathbf{c} + b_\mu, \quad \log \sigma^2 = W_\sigma \cdot \mathbf{c} + b_\sigma.$$

Семплювання: $z = \mu + \sigma \odot \varepsilon$, де $\varepsilon \sim \mathcal{N}(0, I)$ [5]. GRU-декодер реконструює $\hat{X} \in \mathbb{R}^{72 \times 5}$ з латентного вектора z .

Модель оптимізується за β -VAE функцією втрат із KL-відпалюванням:

$$\mathcal{L}(\theta, \varphi) = 1/(W \cdot d) \cdot \sum_{tj} (x_t - \hat{x}_t)^2 + \beta(t) \cdot DKL,$$

де $\beta(t) = \min(t/T_{warmup}, 1) \cdot \beta_{max}$ з $T_{warmup} = 30$ епох та $\beta_{max} = 0,5$. Значення $\beta_{max} < 1$ зміщує баланс на користь якості реконструкції, що обґрунтовано для anomaly detection.

Показник аномальності обчислюється як:

$$s(X) = 1/(W \cdot d) \cdot \sum_{t=1}^W \sum_{j=1}^d d(x_t - \hat{x}_t)^2 + \beta \cdot DKL.$$

Рішення приймається порівнянням із порогом τ , оптимізованим за F1-score на валідаційній вибірці:

$$\hat{y} = 1, \text{ якщо } s(X) > \tau; \hat{y} = 0, \text{ інакше.}$$

3.2. Дані та препроцесинг. У дослідженні використано набір даних PhysioNet Challenge 2019 [14], що містить записи пацієнтів відділень інтенсивної терапії (ВІТ). Для аналізу були відібрані п'ять вітальних показників: частота серцевих скорочень (HR), рівень насичення крові киснем (SpO₂), середній артеріальний тиск (MAP), температура тіла та частота дихання (RR). Формування вхідних даних здійснювалося шляхом побудови часових вікон тривалістю 6 годин із дискретністю 5 хвилин, що відповідало довжині вікна $W = 72$ часових кроків. Для підвищення кількості навчальних зразків та врахування динаміки зміни фізіологічних показників застосовувався ковзний механізм із кроком зсуву 1 година.

Попередня обробка даних включала нормалізацію ознак за методом z-score, що виконувалася відповідно до виразу $\hat{x}_{t,j} = (x_{t,j} - \mu_j) / (\sigma_j + \varepsilon)$, за статистиками навчальної вибірки, де μ_j та σ_j визначалися на основі статистичних характеристик навчальної вибірки. Для обробки пропущених значень застосовувався метод forward-fill у межах кожного часового вікна з подальшим виконанням нормалізації. Вікна, частка пропущених значень у яких перевищувала 30%, виключалися з подальшого аналізу. Використання такого підходу обґрунтовується результатами дослідження [12], де показано, що коректна обробка пропущених значень у багатовимірних клінічних часових рядах є критичним фактором для забезпечення високої якості прогнозування.

Розбиття набору даних здійснювалося на рівні пацієнтів, що дозволяло уникнути витоку інформації між підмножинами. При цьому 70% даних використовувалося для навчання моделі, 15% — для валідації та 15% — для тестування. Після виконання процедури препроцесингу сформована вибірка містила 5590 часових вікон, серед яких 4930 відповідали нормальному стану пацієнтів, а 660 — випадкам сепсису, що характеризувалося дисбалансом класів на рівні приблизно 11,8%.

3.3. Базові моделі для порівняння. Для оцінювання ефективності запропонованого підходу проведено порівняння результатів із двома базовими моделями, які широко використовуються для аналізу часових рядів та задач класифікації. Першою моделлю був двошаровий класифікатор на основі архітектури LSTM, що містив 144 001 параметр та навчався у режимі контрольованого навчання шляхом мінімізації функції втрат крос-ентропії. Для запобігання перенавчанню використовувався механізм ранньої зупинки, який завершив процес навчання на 29-й епосі.

Другою базовою моделлю виступав алгоритм XGBoost [15], конфігурація якого включала 200 дерев рішень із максимальною глибиною 6. На відміну від LSTM-підходу, навчання моделі виконувалося на основі агрегованих статистичних характеристик часових вікон, зокрема середнього значення, стандартного відхилення, мінімального та максимального значень, а також показника тренду. Для забезпечення коректності порівняння обидві базові моделі використовували однакову процедуру попередньої обробки даних та ідентичне розбиття вибірки на рівні пацієнтів.

4. Результати дослідження. Результати оцінювання на тестовій вибірці наведено у Табл. 2.

Таблиця 2.

Порівняння моделей детекції сепсису

Модель	Accuracy	F1	Precision	Recall	Specificity	FPR	ROC-AUC	PR-AUC
Attention-VAE	0,9597	0,8286	0,9869	0,7140	0,9985	0,0015	0,8893	0,8246
LSTM Baseline	0,9537	0,7968	0,9943	0,6648	0,9994	0,0006	0,9066	0,8589
XGBoost	0,9656	0,8571	0,9901	0,7557	0,9988	0,0012	0,8928	0,8415

Attention-VAE демонструє $F1 = 0,8286$ — лише на 0,0285 нижче за найкращий supervised бейзлайн (XGBoost), незважаючи на відсутність доступу до розмічених даних під час навчання. Усі три моделі мають $Precision > 0,98$.

$FPR = 0,0015$ означає 15 хибних тривог на 10 000 нормальних спостережень — прийнятний рівень для клінічного середовища, що мінімізує alarm fatigue [16]. Калібрований поріг $\tau = 0,6559$. За ROC-AUC та PR-AUC різниця між моделями не перевищує 0,0173.

ROC-крива Attention-VAE характеризується стрімким підйомом у лівій частині (область низьких FPR), що узгоджується з принципом автоенкодера: записи з найбільшими відхиленнями від норми генерують найвищу реконструкційну похибку.

PR-криві всіх моделей демонструють характерний перегин у діапазоні Recall 0,65–0,80, що свідчить про підгрупу об'єктивно складних для детекції випадків — ймовірно, ранні стадії сепсису з помірними фізіологічними змінами.

Розподіл anomaly score для нормального класу має моду $\sim 0,15$ – $0,25$, для сепсису — $\sim 0,55$ – $0,70$. Розподіл для сепсису є бімодальним: перший пік ($\sim 0,40$ – $0,50$) відповідає ранньостадійним випадкам, другий ($\sim 0,75$ – $0,90$) — вираженим. Бімодальність відображає гетерогенність клінічних фенотипів сепсису.

Аналіз ваг уваги виявляє характерні відмінності між класами:

- Нормальні пацієнти: рівномірний розподіл ваг, ентропія $H(\alpha) \approx \log 72 \approx 4,28$.
- Пацієнти із сепсисом: ваги концентруються на сегментах погіршення, ентропія $H(\alpha) \in [2,5; 3,5]$.

Піки уваги збігаються з: підвищенням ЧСС на 15–20 уд./хв; зниженням SpO_2 на 3–5 в.п. нижче 95%; падінням MAP нижче 65 мм рт. ст. — ці маркери модель виявила самостійно, без використання клінічних знань під час навчання. Тест Манна-Уїтні підтверджує значущість різниці ($p < 0,001$).

t-SNE та UMAP візуалізації 16-вимірному латентному простору: нормальні записи утворюють щільний кластер; записи сепсису розподілені дифузно на периферії, утворюючи кілька під-кластерів. Вимір z_3 кодує загальний ступінь відхилення від норми; виміри z_7 , z_{11} відображають варіативність патологічних патернів.

Аналіз відношення $R_j = \text{MSE}_{j_{\text{sepsis}}} / \text{MSE}_{j_{\text{norm}}}$ виявляє ієрархію діагностичної цінності:

1. **ЧСС (HR)** — найбільше відносне зростання; тахікардія є найбільш раннім маркером системної запальної відповіді.
2. **МАР** — гіпотензія ($\text{MAP} < 65$) асоціюється з септичним шоком.
3. **SpO₂** — значне зростання для підгрупи з респіраторною дисфункцією.
4. **RR** — помірне зростання, відповідає критеріям SIRS/qSOFA.
5. **Температура** — найменше зростання через повільну динаміку та двоспрямованість (гіпер- і гіпотермія).

Ця ієрархія повністю узгоджується з клінічними знаннями, підтверджуючи медичну обґрунтованість навчених представлень.

5. Дослідження внеску окремих компонентів архітектури та аналіз чутливості до гіперпараметрів. Для ізольованої оцінки внеску архітектурних компонент та перевірки стабільності моделі до варіації ключових гіперпараметрів проведено серію з 9 додаткових експериментів на тій самій тестовій вибірці із фіксованим seed. У межах цієї серії повторно навчена базова конфігурація демонструє $F1 = 0,8217$ при порозі $\tau = 0,7764$, тоді як основна модель з Табл. 2 досягає $F1 = 0,8286$, при $\tau = 0,6559$. Різниця між тренуваннями моделі при різних ініціалізаціях ваг ($\Delta F1 \approx 0,007$ і $\Delta \tau \approx 0,12$) відображає типову варіативність глибокого навчання та враховується при інтерпретації результатів ablation/sensitivity. Поріг для кожної конфігурації серії калібрувався окремо за F1-критерієм на валідаційній вибірці. Результати наведено у Табл. 3.

Таблиця 3.

Абляційне дослідження та аналіз чутливості

Конфігурація	F1	Precision	Recall	ROC-AUC	FPR	$\Delta F1$
Baseline (повна модель)	0,8217	0,9893	0,7027	0,8788	0,0012	—
Без attention (mean pooling)	0,8182	0,9866	0,6989	0,8756	0,0015	-0,003
$\beta = 0,1$	0,8184	0,9689	0,7083	0,8755	0,0036	-0,003
$\beta = 0,3$	0,8124	0,9812	0,6932	0,8701	0,0021	-0,009
$\beta = 1,0$	0,8129	0,9865	0,6913	0,8777	0,0015	-0,009
$\text{dim}(z) = 8$	0,8302	0,9844	0,7178	0,8705	0,0018	+0,008
$\text{dim}(z) = 32$	0,8375	0,9871	0,7273	0,8847	0,0015	+0,016
heads = 1	0,7803	0,9855	0,6458	0,8788	0,0015	-0,041
heads = 8	0,8247	0,9868	0,7083	0,8809	0,0015	+0,003

Заміна механізму уваги на просте усереднення прихованих станів GRU за часовою віссю (mean pooling) не призводить до значущого зниження F1 ($\Delta F1 = -0,003$), що пояснюється здатністю GRU самостійно агрегувати часові залежності. Однак зменшення кількості голів уваги до однієї (heads = 1) знижує

F1 на 0,041 — це підтверджує, що ключовим архітектурним рішенням є саме багатоголова конфігурація, яка дозволяє паралельно моделювати різні патерни взаємодії між каналами вітальних показників. Примітно, що одноголова конфігурація дає гірший результат, ніж повна відсутність уваги: одна голова уваги без регуляризації через диверсифікацію може фокусуватися на шумових артефактах, тоді як mean pooling забезпечує робастне нейтральне агрегування. Окрім кількісного впливу, механізм уваги забезпечує клінічну інтерпретованість рішень (підрозділ 4.4), що є самостійною контрибуцією для систем підтримки прийняття рішень у медичному домені.

Аналіз чутливості охоплює три ключові гіперпараметри архітектури: β у функції втрат, розмір латентного простору $\dim(z)$ та кількість голів уваги. Залежність F1 від β має U-подібну форму з оптимумом у baseline-значенні 0,5: відхилення в обидва боки ($\beta = 0,1$ та $\beta = 1,0$) знижують F1 на 0,003–0,009, що обґрунтовує обраний рівень регуляризації KL-дивергенції. За параметром $\dim(z)$ модель демонструє стабільність — F1 змінюється у межах $\pm 0,02$ від baseline; незначне покращення при $\dim(z) = 32$ (+0,016) знаходиться у межах варіативності між тренуваннями з різними значеннями seed, а базова конфігурація $\dim(z) = 16$ обрана з огляду на обчислювальну ефективність та зменшення ризику перенавчання. За кількістю голів уваги діапазон ширший: heads = 8 дає +0,003, тоді як heads = 1 дає -0,041 (див. підрозділ 5.1).

6. Обговорення. Різниця F1 між Attention-VAE та XGBoost становить лише 0,0285. З огляду на принципову різницю у вимогах до даних — Attention-VAE не потребує розмічених даних позитивного класу — цей результат є значущим. У клінічній практиці розмітка даних сепсису є суб'єктивною та ресурсоемною [4], що робить unsupervised-підхід більш придатним до практичного впровадження у нових закладах.

Attention-VAE надає подвійний механізм інтерпретації:

1. Часовий — ваги уваги відповідають: «коли почалося погіршення?»
2. Ознаковий — декомпозиція реконструкційної похибки: «що саме погіршується?»

Це відповідає вимогам FDA та ЕМА до пояснюваності алгоритмічних рішень у медичних системах [17].

FPR = 0,15% (15 хибних тривог на 10 000 спостережень) є критично важливим для практичного впровадження. Alarm fatigue — зниження уваги персоналу через надмірну кількість хибних сигналів — визнано одним із провідних факторів ризику у ВІТ [16].

Попри отримані результати, проведене дослідження має низку обмежень, які необхідно враховувати при інтерпретації результатів та оцінюванні можливостей практичного застосування запропонованого підходу.

1. Одноцентрові дані. Результати отримано на PhysioNet Challenge 2019. Для підтвердження узагальнюваності необхідна мультицентрова валідація.
2. Обмежений набір ознак. Лише 5 базових вітальних показників; інтеграція лабораторних даних (лактат, прокальцитонін) потенційно підвищить якість.
3. Ретроспективний характер. Оцінка реального клінічного впливу потребує проспективних досліджень.

7. Висновки. Запропоновано архітектуру Attention-VAE для unsupervised раннього виявлення сепсису на основі мультиваріативних часових рядів вітальних показників. Підхід поєднує: (1) навчання без розмічених даних сепсису; (2) конкурентоспроможну якість ($F1 = 0,83$ vs $0,86$ для XGBoost); (3) клінічно інтерпретовані рішення через механізм уваги та декомпозицію реконструкційних похибок.

Оцінювання на PhysioNet Challenge 2019 підтверджує, що модель самостійно навчається фокусуватися на маркерах сепсису — тахікардії, гіпотензії та десадурації. Латентний простір демонструє чітке розділення нормальних та аномальних патернів.

Напрями подальших досліджень: мультицентрова валідація, розширення набору ознак лабораторними даними, дослідження Transformer-encoder та проспективне оцінювання у клінічному середовищі.

Конфлікт інтересів

Автор заявляє, що не має конфлікту інтересів щодо даного дослідження, включаючи фінансовий, особистий, авторський або будь-який інший, який міг би вплинути на дослідження, а також на результати, представлені в даній статті.

Фінансування

Дослідження було проведено без фінансової підтримки.

Доступність даних

Набір даних PhysioNet Challenge 2019 є загальнодоступним за адресою: <https://physionet.org/content/challenge-2019/1.0.0>

Використання штучного інтелекту

Автори підтверджують, що при створенні даної роботи вони не використовували технології штучного інтелекту.

Внесок авторів

С. В. Мошко: концептуалізація, розробка методології, формальний аналіз, курація даних, програмна реалізація, візуалізація, написання — оригінальний проект. О. Ю. Кучанський: наукове керівництво, адміністрування проекту, написання — рецензування та редагування.

Авторські права ©



(2026). Мошко С. В., Кучанський О. Ю. Ця робота ліцензується відповідно до Creative Commons Attribution 4.0 International License.

Список використаної літератури

1. World Health Organization. (2020). Global report on the epidemiology and burden of sepsis. *WHO*. Retrieved from <https://www.who.int/publications/i/item/9789240010789>
2. Kumar, A., Roberts, D., Wood, K. E., & et al. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine*, 34(6), 1589–1596. <https://doi.org/10.1097/01.CCM.0000212262.65573.4E>
3. Singer, M., Deutschman, C. S., Seymour, C. W., & et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 801–810. <https://doi.org/10.1001/jama.2016.0287>
4. Harutyunyan, H., Khachatrian, H., Kale, D. C., & et al. (2019). Multitask learning and benchmarking with clinical time series data. *Scientific Data*, 6(96). <https://doi.org/10.1038/s41597-019-0103-9>
5. Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. In *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*. Retrieved from <https://arxiv.org/abs/1312.6114>
6. Su, Y., Zhao, Y., Niu, C., & et al. (2019). Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/3292500.3330672>
7. Vaswani, A., Shazeer, N., Parmar, N., & et al. (2017). Attention is all you need. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*. Retrieved from <https://arxiv.org/abs/1706.03762>
8. Correia, L., Goos, J.-C., Klein, P., & et al. (2023). MA-VAE: Multi-head attention-based variational autoencoder approach for anomaly detection in multivariate time-series. In *Proceedings of the 15th International Joint Conference on Computational Intelligence (NCTA)*. Retrieved from <https://arxiv.org/abs/2309.02253>
9. Zubair, M., Din, I., Sarwar, N., Elov, B., Makhmudov, S., & Trabelsi, Z. (2025). Revolutionizing sepsis diagnosis using machine learning and deep learning models: a systematic literature review. *BMC Infectious Diseases*, 25(1), 1396. <https://doi.org/10.1186/s12879-025-10766-4>
10. Dou, Y., Li, W., & Zomaya, A. Y. (2022). Transformer-based unsupervised learning for early detection of sepsis (Student Abstract). In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. <https://doi.org/10.1609/aaai.v36i11.21669>
11. Zhang, D., & et al. (2021). An interpretable deep-learning model for early prediction of sepsis in the emergency department. *Patterns*, 2(2), 100196. <https://doi.org/10.1016/j.patter.2021.100196>
12. Che, Z., Purushotham, S., Cho, K., & et al. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8, 6085. <https://doi.org/10.1038/s41598-018-24271-9>
13. Cho, K., van Merriënboer, B., Gulcehre, C., & et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. <https://doi.org/10.3115/v1/D14-1179>
14. Reyna, M. A., & et al. (2020). Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*, 48(2), 210–217. <https://doi.org/10.1097/CCM.0000000000004145>
15. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/2939672.2939785>
16. Cvach, M. (2012). Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation & Technology*, 46(4), 268–277. <https://doi.org/10.2345/0899-8205-46.4.268>
17. FDA. (2021). Artificial intelligence and machine learning in software as a medical device. *U.S. Food and Drug Administration*. Retrieved from <https://www.fda.gov/media/145022/download>

Moshko S. V., Kuchanskyi O. Y. Attention-VAE: Unsupervised early sepsis detection via reconstruction-based anomaly scoring of intensive care unit vital signs.

This paper examines the application of unsupervised deep learning for early sepsis detection in intensive care units based on vital sign time series analysis. To achieve the research objective, the following tasks were set and accomplished: a review of existing anomaly detection approaches for medical time series and deep learning methods for sepsis diagnosis; a formal problem statement framing sepsis detection as a multivariate time series anomaly detection task; the development of the Attention-VAE architecture combining a GRU-based recurrent encoder, multi-head scaled dot-product attention, and a variational decoder. The model is trained exclusively on records of patients with normal physiological status, without using any labeled sepsis data. The model output is an anomaly score reflecting the degree of deviation of the patient's current state from learned normal patterns, along with temporal attention maps and per-feature reconstruction error profiles that provide dual interpretability of decisions. The model was evaluated on the PhysioNet Challenge 2019 clinical database using five vital signs in six-hour observation windows, achieving $F1 = 0.8286$, $ROC-AUC = 0.8893$, and a false positive rate of 0.15% — results competitive with supervised methods. Limitations of the study include the single-centre nature of the data and the need for multi-centre validation. The research results constitute part of the master's thesis and may be applied in clinical decision support systems.

Keywords: variational autoencoder; attention mechanism; sepsis detection; time series anomaly detection; unsupervised learning; PhysioNet Challenge 2019; interpretable machine learning.

Отримано: 02.04.2026

Прийнято: 18.04.2026

Опубліковано: 30.04.2026