НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФАКУЛЬТЕТ БІОМЕДИЧНОЇ ІНЖЕНЕРІЇ
(повна назва інституту/факультету)

кафедра БІОМЕДИЧНОЇ КІБЕРНЕТИКИ
(повна назва кафедри)

«На правах рукопису»
УДК _____

«До захисту допущено»
В.о.завідувача кафедри БМК

_____ Світлана АЛХІМОВА
(підпис) (ініціали, ПРІЗВИЩЕ

“___” __*травня*__2026р.

Магістерська дисертація
на здобуття ступеня магістра

за освітньо-науковою програмою *«Комп’ютерні науки»*
зі спеціальності *122 «Комп’ютерні науки»*

на тему: **Інтелектуальна система раннього виявлення сепсису на основі аналізу багатовимірних часових рядів життєвих показників пацієнтів**

Виконав: студент **II** курсу, групи _ЗК-41мн__

МОШКО СЕРГІЙ ВОЛОДИМИРОВИЧ _____
(прізвище, ім’я, по батькові) (підпис)

Науковий керівник: *проф. каф. біомедичної кібернетики (БМК) проф., д.т.н., Кучанський Олександр Юрійович*
(посада, науковий ступінь, вчене звання, прізвище, ім’я, по ініціали) (підпис)

Консультант з розділів магістерської дисертації:

_____ _____
(назва розділу) (посада, вчене звання, науковий ступінь, прізвище, ініціали) (підпис)

Рецензент: *зав. каф. цифрових технологій в енергетиці професор, д-р тех. наук Аушева Наталія Миколаївна*
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.

Студент _____ Сергій МОШКО
(підпис)

Київ – 2026 року

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет біомедичної інженерії

Кафедра біомедичної кібернетики

Рівень вищої освіти – другий (магістерський)

Спеціальність – 122 «Комп'ютерні науки»

Освітньо-наукова програма «Комп'ютерні науки»

ЗАТВЕРДЖУЮ

В.о.завідувача кафедри БМК

_____ Світлана АЛХІМОВА

«_13_»__січня___2026 р.

ЗАВДАННЯ
на магістерську дисертацію студенту

МОШКО СЕРГІЙ ВОЛОДИМИРОВИЧ

(прізвище, ім'я, по батькові)

1. Тема дисертації **Інтелектуальна система раннього виявлення сепсису на основі аналізу багатовимірних часових рядів життєвих показників пацієнтів**

науковий керівник дисертації

Кучанський Олександр Юрійович, д.т.н., професор

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «*30*» _*березня*_ 2026 р. **№*1320-с***_

2. Термін подання студентом дисертації ***25-27 квітня 2026 року***

3. Об'єкт дослідження — процес раннього виявлення сепсису у відділеннях інтенсивної терапії на основі безперервного моніторингу життєвих показників пацієнтів.

4. Вихідні дані — набір даних PhysioNet/Computing in Cardiology Challenge 2019 (п'ять вітальних показників: HR, SpO_2, MAP, температура, частота дихання; ресемплінг до 5 хв; вікно 6 год).

5. Перелік завдань, які потрібно розробити — аналіз сучасного стану проблеми; математична формалізація задачі виявлення аномалій; розробка архітектури Attention-VAE з BiGRU-енкодером та механізмом multi-head attention; реалізація конвеєра передобробки даних; експериментальне порівняння з базовими моделями (XGBoost, LSTM); аналіз інтерпретованості через ваги уваги та латентний простір (t-SNE, UMAP); оцінка клінічної значущості результатів.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу 14 рисунків, 32 таблиці, 54 формули, презентація з захисту МД на 12 слайдах.

7. Орієнтовний перелік публікацій 1 стаття у фаховому журналі категорії Б

(Науковий вісник Ужгородського університету. Серія: Математика і інформатика, 2026. Вип. 49(2). С. 1–6).
8. Консультанти розділів дисертації: не передбачено
9. Дата видачі завдання ***13 січня 2026р.***

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів МД	Примітка
1	Отримати завдання на МД	До 10.01.2026	***виконано***
2	Завершення виконання практичної частина МД	До 22.03.2026	***виконано***
3	Завершення оформлення розділів МД (Вступ, Основні розділи та висновки до них, Загальні висновки, Список використаних джерел)	До 20.04.2026	***виконано***
4	Апробація та публікація результатів дослідження МД (отримати підтвердження про прийняття до друку)	До 25.04.2026	***виконано***
5	Перевірка МД науковим керівником	20.04.2026	***виконано***
6	Подання в електронному вигляді МД на перевірку нормоконтролера	До 25.04.2026	
7	Подання в електронному вигляді МД на перевірку подібності Strike Plagiarism com. Отримати позитивний звіт подібності.	До 27.04.2026	
8	Підготовка Експертної оцінки до звіту подібності.	До 02.05.2026	
9	Отримати відгук наукового керівника	02.05.2026	
10	Надати на кафедру пакет документів в паперовому та електронному вигляді (МД, відгук керівника, звіт подібності, Експертної оцінки до звіту подібності)	До 06.05.2026	
11	Отримати допуск до захисту МД в ЕК та направлення до рецензента (засідання кафедри)	Засідання кафедри	
12	Подання МД рецензенту. Отримати на надання рецензію до ЕК / на кафедру	До 15.05.2026	
13	Подання супровідного пакету документів по МД до захисту в ЕК	До 23.05.2026	
14	Захист МД в ЕК	26.05.2026	

Студент _____ **Сергій МОШКО**
(підпис) (ім'я, ПРІЗВИЩЕ)

Науковий керівник _____ **Олександр КУЧАНСЬКИЙ**
(підпис) (ім'я, ПРІЗВИЩЕ)

Нормоконтролер _____ **Галина КОРНІЄНКО**
(підпис) (ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Магістерська дисертація за темою «Інтелектуальна система раннього виявлення сепсису на основі аналізу багатовимірних часових рядів життєвих показників пацієнтів» виконана студентом кафедри біомедичної кібернетики факультету біомедичної інженерії *Мошком Сергієм Володимировичем* зі спеціальності 122 «Комп'ютерні науки» за освітньо-науковою програмою «Комп'ютерні науки» та складається зі: вступу; 5 розділів, висновків до кожного з цих розділів; загальних висновків; списку використаних джерел, який налічує 43 джерела. Загальний обсяг роботи 152 сторінки, у тому числі 132 сторінках основного тексту, 14 рисунків, 32 таблиці, 54 формули.

Актуальність теми. Сепсис залишається однією з провідних причин смертності у відділеннях інтенсивної терапії: щорічно у світі реєструється близько 49 мільйонів випадків, понад 11 мільйонів з яких завершуються летальним результатом. Кожна година затримки з адекватною антибіотикотерапією збільшує ризик летального наслідку на 7–8%. Традиційні шкали (SOFA, qSOFA) реактивні і фіксують стан, коли органна дисфункція вже сформована, тому актуальною є розробка інтелектуальних систем, здатних виявляти ранні відхилення на основі багатовимірних часових рядів життєвих показників.

Мета і задачі дослідження. Метою роботи є визначення ефективності застосування архітектури варіаційного автоенкодера з механізмом уваги (Attention-VAE) для раннього виявлення сепсису шляхом аналізу багатовимірних часових рядів життєвих показників пацієнтів ВІТ. Сформульовано задачі: проаналізувати сучасний стан проблеми та обґрунтувати підхід без учителя; розробити математичну формалізацію задачі як детектування аномалій; спроєктувати архітектуру Attention-VAE з GRU-енкодером та механізмом уваги; реалізувати конвеєр передобробки даних PhysioNet Challenge 2019; провести експериментальне оцінювання та

порівняння з XGBoost і LSTM; дослідити інтерпретованість через аналіз ваг уваги та латентного простору; оцінити клінічну значущість результатів.

Об'єкт дослідження. Процес раннього виявлення сепсису у відділеннях інтенсивної терапії на основі безперервного моніторингу життєвих показників пацієнтів.

Предмет дослідження. Методи та алгоритми виявлення аномалій у багатовимірних часових рядах медичних даних із використанням архітектури варіаційного автоенкодера з механізмом уваги.

Методи дослідження. Методи теорії ймовірностей та математичної статистики — для формалізації задачі виявлення аномалій; методи глибокого навчання (GRU, attention, VAE); методи машинного навчання (XGBoost, LSTM) як базові моделі для порівняння; методи зниження розмірності (t-SNE, UMAP) для аналізу латентного простору; методи статистичного оцінювання якості (Accuracy, F1-score, Precision, Recall, ROC-AUC, PR-AUC).

Наукова новизна одержаних результатів. Вперше запропоновано застосування архітектури Attention-VAE для задачі раннього виявлення сепсису на основі багатовимірних часових рядів, що поєднує переваги навчання без учителя з можливістю інтерпретації через механізм уваги. Удосконалено методику аналізу інтерпретованості моделей виявлення аномалій у медичних часових рядах. Дістало подальшого розвитку застосування методів зниження розмірності для верифікації якості латентних представлень моделей виявлення аномалій.

Практичне значення одержаних результатів. Розроблено програмну реалізацію інтелектуальної системи раннього виявлення сепсису, яку можна інтегрувати з існуючими системами моніторингу пацієнтів у ВІТ. Характер навчання без учителя спрощує впровадження у нових клінічних закладах без потреби у дорогій розмітці. Низька частка хибнопозитивних спрацювань (FPR = 0,15%) мінімізує проблему alarm fatigue. Модель досягла F1 = 0,8286, ROC-AUC = 0,8893, Precision = 0,9869, Recall = 0,714.

Апробація результатів дослідження. *Не передбачено.*

Публікації. За результатами роботи опубліковано 1 статтю у фаховому журналі категорії Б:

Мошко С. В., Кучанський О. Ю. Attention-VAE: несупервізоване раннє виявлення сепсису через оцінку аномалій на основі реконструкції вітальних показників відділення інтенсивної терапії. *Науковий вісник Ужгородського університету. Серія: Математика і інформатика.* 2026. Вип. 49(2). С. 1–6. DOI: 10.24144/2616-7700.2026.49(2).1-6.

Ключові слова. сепсис, виявлення аномалій, варіаційний автоенкодер, механізм уваги, багатовимірні часові ряди, життєві показники, глибоке навчання, відділення інтенсивної терапії, PhysioNet Challenge 2019.

Бібліографічний опис

Мошко, С. В. Інтелектуальна система раннього виявлення сепсису на основі аналізу багатовимірних часових рядів життєвих показників пацієнтів : магістерська дис. : 122 Комп’ютерні науки / Мошко Сергій Володимирович. — Київ, 2026. — 152 с.

ABSTRACT

The master's thesis "Intelligent system for early sepsis detection based on analysis of multivariate time series of patients' vital signs" was completed by Serhii Volodymyrovych Moshko, student of the Department of Biomedical Cybernetics, Faculty of Biomedical Engineering, specialty 122 "Computer Sciences", educational-scientific programme "Computer Sciences". The thesis consists of an introduction, 5 chapters with conclusions to each, general conclusions, a list of 43 used sources. Total length: 152 pages, including 132 pages of main text, 14 figures, 32 tables, 54 formulas.

Relevance. Sepsis remains one of the leading causes of mortality in intensive care units worldwide, with approximately 49 million cases registered annually and over 11 million resulting in death. Every hour of delay in adequate antibiotic therapy increases the risk of fatal outcome by 7–8%. Traditional clinical scales (SOFA, qSOFA) are reactive and detect the condition after organ dysfunction has already manifested. This motivates the development of intelligent systems capable of detecting early deviations in multivariate vital sign time series.

Goal and objectives. The aim is to evaluate the effectiveness of the Attention-VAE architecture for early sepsis detection through analysis of multivariate vital sign time series of ICU patients. Objectives: analyse the current state of the problem and justify the unsupervised approach; develop a mathematical formalisation as anomaly detection; design the Attention-VAE architecture with GRU encoder and attention mechanism; implement the data preprocessing pipeline for PhysioNet Challenge 2019; perform experimental evaluation and comparison with XGBoost and LSTM; study interpretability via attention weight analysis and latent space; evaluate the clinical significance of results.

Object of research. The process of early sepsis detection in intensive care units based on continuous vital sign monitoring.

Subject of research. Methods and algorithms for anomaly detection in multivariate medical time series using a variational autoencoder architecture with an attention mechanism.

Research methods. Probability theory and mathematical statistics for the formalisation of anomaly detection; deep learning methods (GRU, attention, VAE); machine learning methods (XGBoost, LSTM) as baselines; dimensionality reduction methods (t-SNE, UMAP) for latent space analysis; statistical evaluation methods (Accuracy, F1-score, Precision, Recall, ROC-AUC, PR-AUC).

Scientific novelty. For the first time, the Attention-VAE architecture is applied to early sepsis detection based on multivariate time series, combining the advantages of unsupervised learning with attention-based interpretability. The methodology of interpretability analysis for anomaly detection models in medical time series has been improved. Further development was given to the application of dimensionality reduction methods for verifying the quality of latent representations.

Practical significance. A software implementation of an intelligent early sepsis detection system has been developed that can be integrated with existing ICU monitoring systems. The unsupervised learning approach simplifies deployment in new clinical settings without expensive labelling. The low false positive rate (FPR = 0.15%) minimises alarm fatigue. The model achieved F1 = 0.8286, ROC-AUC = 0.8893, Precision = 0.9869, Recall = 0.714.

Publications. Moshko S. V., Kuchanskyi O. Yu. Attention-VAE: Unsupervised Early Sepsis Detection through Reconstruction-Based Anomaly Scoring of ICU Vital Signs. *Scientific Bulletin of Uzhhorod University. Series: Mathematics and Informatics.* 2026. Issue 49(2). P. 1–6. DOI: 10.24144/2616-7700.2026.49(2).1-6.

Keywords: sepsis, anomaly detection, variational autoencoder, attention mechanism, multivariate time series, vital signs, deep learning, intensive care unit, PhysioNet Challenge 2019.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ВІТ – відділення інтенсивної терапії

ВООЗ – Всесвітня організація охорони здоров'я

ВТ – втрати/функція втрат

ЕКГ – електрокардіограма

ЕМЗ – електронні медичні записи

ЕСППР – експертна система підтримки прийняття рішень

МД – магістерська дисертація

ПЗ – програмне забезпечення

ЧСС – частота серцевих скорочень

AE – Autoencoder (автоенкодер)

Attention-VAE – Attention-based Variational Autoencoder

AUC – Area Under Curve

BiGRU – Bidirectional Gated Recurrent Unit

CNN – Convolutional Neural Network (згорткова нейронна мережа)

DTAE – Deterministic Transformer Autoencoder

ELBO – Evidence Lower Bound

EHR – Electronic Health Record

F1 – F1-score (гармонічне середнє Precision та Recall)

FPR – False Positive Rate

GRU – Gated Recurrent Unit

HR – Heart Rate (частота серцевих скорочень)

ICU – Intensive Care Unit

KL – Kullback–Leibler divergence (розходження Кульбака–Лейблера)

LSTM – Long Short-Term Memory

MAP – Mean Arterial Pressure (середній артеріальний тиск)

MIL – Multiple Instance Learning

MTS – Multivariate Time Series

PR-AUC – Precision–Recall AUC

PU – Positive-Unlabeled learning

RNN – Recurrent Neural Network (рекурентна нейронна мережа)

ROC – Receiver Operating Characteristic

ROC-AUC – Area Under the ROC Curve

RR – Respiratory Rate (частота дихання)

SOFA – Sequential Organ Failure Assessment

qSOFA – quick SOFA

SpO_2 – Saturation of peripheral Oxygen (сатурація кисню)

t-SNE – t-distributed Stochastic Neighbor Embedding

TCN – Temporal Convolutional Network

UMAP – Uniform Manifold Approximation and Projection

VAE – Variational Autoencoder (варіаційний автоенкодер)

XGBoost – Extreme Gradient Boosting

Вступ

Актуальність теми дослідження

Сепсис залишається однією з провідних причин смертності у відділеннях інтенсивної терапії (ВІТ) по всьому світу. За оцінками Всесвітньої організації охорони здоров'я, щорічно у світі реєструється близько 49 мільйонів випадків сепсису, з яких понад 11 мільйонів завершуються летальним результатом; це становить приблизно 20% усіх смертей у глобальному масштабі [32]. Клінічний перебіг сепсису має швидку динаміку: кожна година затримки з початком адекватної антибіотикотерапії збільшує ризик летального наслідку на 7–8% [31]. Відтак рання діагностика — критична умова ефективного лікування.

Традиційні підходи до виявлення сепсису ґрунтуються на клінічних шкалах (SOFA, qSOFA), лабораторних біомаркерах і правилах прийняття рішень. Проте зазначені методи мають суттєві обмеження: клінічні шкали реактивні і фіксують стан органної дисфункції, коли патологічний процес уже набув клінічно значущої форми; лабораторні показники надходять із часовою затримкою; дискретні порогові критерії не враховують багатовимірну динаміку фізіологічних показників [19]. Водночас у сучасних ВІТ безперервно реєструються значні обсяги даних моніторингу життєвих показників (частота серцевих скорочень, насичення крові киснем, середній артеріальний тиск, температура тіла, частота дихання), які потенційно містять ранні сигнали розвитку сепсису, але потребують застосування інтелектуальних методів обробки для їх ефективного використання.

Розвиток методів глибокого навчання відкриває нові можливості для аналізу багатовимірних часових рядів медичних даних. Зокрема, підходи на основі виявлення аномалій (anomaly detection) особливо перспективні для задачі раннього виявлення сепсису: вони не потребують точних міток початку патологічного процесу, достатньо мати приклади нормального фізіологічного

перебігу [3; 18]. Це принципово, оскільки момент *onset* сепсису в ретроспективних клінічних базах даних визначається неоднозначно і залежить від протоколів конкретного закладу. Архітектура варіаційного автоенкодера (VAE) з інтегрованим механізмом уваги (Attention) поєднує здатність до моделювання складних розподілів нормальних фізіологічних патернів з можливістю інтерпретації рішень моделі — важливою вимогою для клінічних систем підтримки прийняття рішень [3].

Зв'язок роботи з науковими програмами, планами, темами

Дисертаційну роботу виконано на кафедрі біомедичної інженерії Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» відповідно до плану наукових досліджень. Тематика роботи відповідає пріоритетним напрямам розвитку інтелектуальних систем в охороні здоров'я та цифрової трансформації медицини.

Мета і задачі дослідження

Мета дослідження — визначення ефективності застосування архітектури варіаційного автоенкодера з механізмом уваги (Attention-VAE) для раннього виявлення сепсису шляхом аналізу багатовимірних часових рядів життєвих показників пацієнтів ВІТ.

Для досягнення поставленої мети необхідно розв'язати такі **задачі**:

- Провести аналіз сучасного стану проблеми раннього виявлення сепсису, існуючих клінічних підходів та методів машинного навчання для обробки медичних часових рядів, обґрунтувати доцільність застосування підходу без учителя на основі виявлення аномалій.
- Розробити математичну формалізацію задачі раннього виявлення сепсису як задачі детектування аномалій у багатовимірних часових рядах та обґрунтувати використання реконструкційної похибки як критерію аномальності.
- Спроєктувати та реалізувати архітектуру Attention-VAE, що поєднує рекурентний енкодер на основі GRU-мережі, механізм уваги для

виділення інформативних часових сегментів та варіаційний декодер для реконструкції сигналу.

- Реалізувати повний конвеєр передобробки даних із набору PhysioNet/Computing in Cardiology Challenge 2019 [35], що охоплює очищення, ресемплінг, імпутацію, нормалізацію та формування ковзних часових вікон із розділенням на рівні пацієнтів.
- Провести експериментальне оцінювання розробленої моделі та порівняння з базовими методами (XGBoost, LSTM) за комплексом метрик якості класифікації.
- Дослідити інтерпретованість моделі через аналіз ваг механізму уваги та візуалізацію латентного простору методами зниження розмірності (t-SNE, UMAP).
- Оцінити клінічну значущість отриманих результатів та визначити напрями подальшого розвитку системи.

Об’єкт дослідження

Об’єктом дослідження є процес раннього виявлення сепсису у відділеннях інтенсивної терапії на основі безперервного моніторингу життєвих показників пацієнтів.

Предмет дослідження

Предметом дослідження є методи та алгоритми виявлення аномалій у багатовимірних часових рядах медичних даних із використанням архітектури варіаційного автоенкодера з механізмом уваги.

Методи дослідження

У роботі застосовано такі методи дослідження:

- **методи теорії ймовірностей та математичної статистики** — для формалізації задачі виявлення аномалій, опису латентних змінних та стохастичного висновування у варіаційних автоенкодерах;
- **методи глибокого навчання** — рекурентні нейронні мережі (GRU) для моделювання часової динаміки, механізм уваги (attention) для виділення

інформативних фрагментів сигналу, варіаційний автоенкодер (VAE) для навчання латентних представлень нормальних фізіологічних патернів;

- **методи машинного навчання** — градієнтний бустинг (XGBoost) та рекурентні мережі (LSTM) як базові моделі для порівняння;
- **методи зниження розмірності** — t-SNE та UMAP для візуалізації та аналізу структури латентного простору;
- **методи статистичного оцінювання якості** — Accuracy, F1-score, Precision, Recall, ROC-AUC, PR-AUC для комплексного порівняння моделей в умовах незбалансованих класів.

Наукова новизна одержаних результатів

Наукова новизна дисертаційного дослідження полягає у такому:

- **Вперше запропоновано застосування архітектури Attention-VAE** для задачі раннього виявлення сепсису на основі багатовимірних часових рядів, що дозволяє поєднати переваги навчання без учителя з можливістю інтерпретації рішень моделі через механізм уваги. На відміну від існуючих підходів із учителем, запропонований метод не потребує точних міток початку сепсису. На відміну від єдиного відомого несупервізованого підходу DTAE [40], що використовує детермінований декодер, Attention-VAE забезпечує імовірнісне моделювання латентного простору з явною оцінкою невизначеності.
- **Удосконалено методику аналізу інтерпретованості** моделей виявлення аномалій у медичних часових рядах: запропоновано аналіз ваг механізму уваги Attention-VAE, що дозволяє визначити часові сегменти та комбінації показників, найбільш релевантні для підвищення показника аномальності. Це створює передумови для клінічно осмисленого пояснення сигналів тривоги.
- **Дістало подальшого розвитку застосування методів зниження розмірності** (t-SNE, UMAP) для верифікації якості латентних представлень моделей виявлення аномалій: підтверджено здатність

Attention-VAE формувати клінічно значущі внутрішні представлення — нормальні та аномальні патерни утворюють просторово розділені кластери у латентному просторі.

Практичне значення одержаних результатів

Практичне значення роботи визначається такими аспектами:

- розроблено програмну реалізацію інтелектуальної системи раннього виявлення сепсису, яку може бути інтегровано з існуючими системами моніторингу пацієнтів у ВІТ;
- характер навчання без учителя спрощує впровадження системи у нових клінічних закладах без потреби у дорогій розмітці даних;
- низька частка хибнопозитивних спрацювань (FPR = 0,15%) мінімізує проблему *alarm fatigue* медичного персоналу;
- механізм уваги забезпечує інтерпретованість рішень, надаючи клініцисту контекстну інформацію про причини сигналу тривоги;
- модульна архітектура системи дозволяє поетапно розширювати набір вхідних ознак та адаптувати систему до умов конкретного клінічного закладу.

Апробація результатів дисертації *не передбачено*

Публікації

Мошко С. В., Кучанський О. Ю. Attention-VAE: несупервізоване раннє виявлення сепсису через оцінку аномалій на основі реконструкції вітальних показників відділення інтенсивної терапії. *Науковий вісник Ужгородського університету. Серія: Математика і інформатика.* 2026. Вип. 49(2). С. 1–6. DOI: 10.24144/2616-7700.2026.49(2).1-6.

Структура та обсяг дисертації

Дисертаційна робота складається зі вступу, п’яти розділів, висновків, списку використаних джерел та додатків. Загальний обсяг роботи становить

148 сторінок, у тому числі 128 сторінок основного тексту, 14 рисунків, 32 таблиць, 54 формул. Список використаних джерел включає 43 найменування.

Розділ 1

Аналіз сучасного стану проблеми виявлення сепсису та методів обробки медичних часових рядів

У **Розділі 1** здійснено аналіз сучасного стану проблеми виявлення сепсису та методів обробки медичних часових рядів. Розглянуто клінічні аспекти ранньої діагностики, обмеження існуючих підходів (SOFA, qSOFA), методи машинного та глибокого навчання для аналізу часових рядів, а також обґрунтовано вибір архітектури VAE з механізмом уваги

1.1 Клінічні аспекти ранньої діагностики сепсису та вимоги до інтелектуальних систем

Рання діагностика сепсису належить до центральних завдань інтенсивної терапії. Клінічний перебіг має швидку динаміку і високу варіабельність проявів між пацієнтами. Сепсис розглядається як життєзагрозливий синдром, пов’язаний із дисрегульованою відповіддю організму на інфекцію та розвитком органної дисфункції [1]. Ранність має передусім часовий вимір: ефективність терапевтичних втручань залежить від того, наскільки швидко підвищений ризик буде ідентифіковано та підтверджено додатковими клінічними й лабораторними даними [31].

У відділеннях інтенсивної терапії (ВІТ) доступна значна кількість вимірювань життєвих показників. Перетворити ці потоки даних у керовані клінічні сигнали складно: неоднорідність джерел, пропуски, артефакти, асинхронність вимірювань. До цього додається ще один шар проблем. Ранні фізіологічні зрушення часто мають субклінічний характер, маскуються супутніми станами (післяопераційний стрес, крововтрата, больовий синдром,

седація тощо) і можуть не досягати порогових значень, закладених у правилах скринінгу [19; 8].

Це задає логіку подальших підрозділів: даних-орієнтовані підходи й методи виявлення аномалій, зокрема без учителя (*unsupervised*) і зі слабким наглядом (*weakly supervised*).

Клінічні підходи до виявлення сепсису умовно поділяються на три групи: шкали й правила прийняття рішень; лабораторні біомаркери та панелі; безперервний моніторинг життєвих показників із підтримкою цифрових систем нагляду. Перші дві групи історично орієнтовані на підтвердження вже сформованого патологічного процесу. Звідси наслідок: реальна «ранність» виявлення знижується. Інтервенція розпочинається після накопичення достатньої кількості ознак органної дисфункції або після отримання лабораторних результатів із часовою затримкою [19; 6].

Один із найпоширеніших інструментів оцінювання тяжкості стану — шкала SOFA. Вона інтегрує ознаки дисфункції ключових органних систем і корелює з прогнозом. Її практична цінність полягає у стандартизованій, порівнюваній між пацієнтами оцінці органної недостатності. Для раннього попередження є принципове обмеження: реактивність. Значення SOFA суттєво зростає тоді, коли органна дисфункція вже набула клінічно значущої форми [1]. Тобто SOFA описує актуальний «стан» пацієнта, але не його «траєкторію», не динамічну зміну у часі.

Для спрощеного первинного скринінгу у позалікарняних і приймальних відділеннях запропоновано шкалу qSOFA з невеликою кількістю легко доступних показників. Попри зручність, численні дослідження відзначають недостатню чутливість qSOFA для ранніх стадій. Особливо це проявляється в інтенсивній терапії, де багато факторів (седація, вентиляція, вплив вазопресорів) модифікують прояви та ускладнюють інтерпретацію порогів [1]. Жорсткі порогові критерії пропускають пацієнтів з початковими відхиленнями, які проявляються саме як тренди чи патерни коливань, а не як «разова» подія.

Інша принципова вада шкал: дискретність. Оцінювання ґрунтується на окремих зрізах даних та агрегованих показниках, а не на повній часовій структурі сигналів. Через це ігноруються швидкість змін, епізодичність десатурацій, варіабельність серцевого ритму, довготривалі «дрейфи» середнього артеріального тиску чи зростання нестабільності частоти дихання, які можуть передувати явній декомпенсації [3; 20]. На практиці це означає: клінічний персонал змушений інтегрувати динаміку «вручну», спираючись на досвід і контекст. Відтворюваність рішень погіршується, ризик запізнілого втручання зростає.

З позиції медичної інформатики та аналізу даних сепсис доречно розглядати як процес у багатовимірному просторі фізіологічних змін. Клінічно значущі ознаки часто є комбінаціями помірних відхилень у кількох каналах (наприклад, помірна тахікардія + тенденція до зниження MAP + наростання RR). Окремо ці відхилення не виходять за «тривожні» межі, але разом формують підозрілий патерн, що й сигналізує про ризик [1]. Тому системи підтримки рішень повинні враховувати багатоканальність і взаємозв'язки сигналів, а не лише ізольовані пороги.

Крім клінічної складності, на ранню діагностику впливає якість даних. Життєві показники надходять із моніторів, манжет, датчиків пульсоксиметрії, апаратів ШВЛ та інших пристроїв. Звідси нерівномірні часові ряди з пропусками й артефактами. Вимірювання бувають квазібезперервними (HR, SpO_2), періодичними (температура), подієвими або залежними від втручань персоналу (неінвазивний тиск) [6; 11]. Наївні припущення щодо регулярної дискретизації та повноти даних тут неприйнятні. Натомість потрібні механізми вирівнювання часу, маскування пропусків і оцінювання надійності спостережень.

Тому сучасні дослідження дедалі частіше зосереджуються на інтелектуальних системах, здатних у безперервному режимі аналізувати багатовимірні часові ряди життєвих показників і формувати сигнал ризику. Для цієї роботи центральним є підхід без учителя (*unsupervised*) та зі слабким

наглядом (*weakly supervised*). Причина: точні «мітки» початку сепсису часто нечіткі (клінічне встановлення діагнозу відбувається із затримкою та залежить від доступності лабораторій і інтерпретації лікаря). До цього додається значна міжцентрова та міжпротокольна варіабельність визначень [21; 19]. У такій ситуації модель навчається на патернах «нормального» або відносно стабільного перебігу і детектує відхилення як потенційні аномалії, що можуть передувати клінічній маніфестації.

Для аналізу життєвих показників доречно формалізувати вхідний потік як багатовимірний часовий ряд з пропусками та маскою спостережуваності. Це дає підстави явно сформулювати вимоги до вхідних даних та подальшого алгоритмічного опрацювання за формулою (1.1):

$$X= x(t_k)_(k=1)^T, x(t_k)\in \mathbb{R}^d, m(t_k)\in 0,1^d, \quad (1.1)$$

де $m_i(t_k)=1$ означає наявність вимірювання $x_i(t_k)$, а $m_i(t_k)=0$ — пропуск.

У практичній реалізації системи раннього попередження ключовий крок: перехід від «сирих» вимірювань до ковзних вікон, у яких оцінюються тренди та варіабельність. Такий підхід відповідає клінічній логіці: рання фаза частіше проявляється через зміну характеру коливань і поступове зміщення базового рівня, а не через миттєве перетинання порога. Віконні представлення також роблять дані придатними для моделей глибокого навчання та реконструкційних підходів (зокрема, VAE) [20]. Локальне середнє та оцінка нахилу у ковзному вікні визначаються за формулою (1.2):

$$\mu_i(t)=(1)/(W)\Sigma_(k=t-W+1)^t x_i(t_k), s_i(t)=argmin_(a,b)\Sigma_(k=t-W+1)^t x_i(t_k)-(a,t_k+b)^2, \quad (1.2)$$

де $\mu_i(t)$ — локальне середнє у вікні довжини W, а $s_i(t)=a$ — оцінка нахилу (швидкості зміни) у цьому вікні.

На рівні системи підтримки рішень центральним стає визначення «сигналу ризику» або «аномальності» — характеристики, що відображає ступінь відхилення поточної динаміки від очікуваної. Для реконструкційних

моделей похибка відновлення є природним індикатором аномалії. Це відповідає постановці задачі як пошуку незвичних патернів у багатовимірних часових рядах, без точного маркування моменту початку сепсису [7; 3]. Окремо потрібні нормалізація за каналами і коректне врахування пропусків. Узагальнений вираз для скалярного показника аномальності з нормалізацією за каналами та маскуванням пропусків подається у формулі (1.3):

$$A(t)=\Sigma_{(i=1)}^{d} w_i \cdot (x_i(t)-\hat{x}_i(t)^{2)/(\sigma_i 2+\varepsilon)\cdot m_i(t), \quad (1.3)$$

де $\hat{x}_i(t)$ — реконструкція, σ_i^2 — оцінка дисперсії каналу, w_i — ваги каналів (за замовчуванням $w_i=1$), ε — малий стабілізатор.

Окремий клінічний виклик: баланс між чутливістю та «втомою від тривог» (*alarm fatigue*). Надмірна кількість хибнопозитивних попереджень знижує довіру до системи і збільшує когнітивне навантаження на персонал [8]. Тому вимоги до інтелектуальної системи охоплюють: точність виявлення відхилень, контроль частоти тривог, стабільність сигналу ризику, можливість калібрування порогів під конкретне відділення. Формальне поставлення задачі калібрування порогу за умови контрольованої частоти хибних спрацьовувань представлено у формулі (1.4):

$$\min_{\tau;} ET_{(alarm)}(\tau)-T_{(event)}+ \text{ за умови } PrA(t)>\tau \mid no-event \leq \alpha, \quad (1.4)$$

де τ — поріг тривоги, α — допустима частка хибних спрацювань у періоді без подій, $[z]_+=\max(z,0)$.

З клінічного боку важлива ще одна вимога: інтерпретованість. Система має позначати пацієнта як ризикового і пояснювати, які канали та часові сегменти спричинили підвищення сигналу. Це безпосередньо стикується з механізмами Attention, які можна використати як інструмент локальної інтерпретації (наприклад, у вигляді теплових карт по часу та ознаках) [5; 4]. Вимога інтерпретованості також працює як запобіжник проти використання системи як «чорної скриньки». Клінічне рішення завжди має залишатися за лікарем, а модель виконує роль підтримки, а не заміни діагностики [8].

Оскільки у роботі використовуються дані PhysioNet/Computing in Cardiology Challenge 2019 [35], варто враховувати загальні властивості реєстрових клінічних даних ВІТ: неоднорідність форматів, наявність пропусків і змінних частот дискретизації, а також залежність сигналів від втручань (медикаменти, ШВЛ, інфузійна терапія) як потенційних конфаундерів [35; 8]. Це формує вимогу до алгоритмів: робастність до шумів, підтримка маскування спостережень, здатність працювати в умовах частково спостережуваної системи.

У таблиці 1.1 узагальнено ключові життєві показники, найбільш придатні для безперервного або квазібезперервного моніторингу та формування входу інтелектуальної системи (з урахуванням типових обмежень якості даних).

Таблиця 1.1 – Ключові життєві показники та вимоги до представлення в даних для раннього виявлення сепсису

Група показників	**Життєвий показник (канал)**	**Чому інформативний для ранніх етапів**	**Очікуваний формат у даних**	**Коментарі до якості даних**
Гемодинаміка	Частота серцевих скорочень (ЧСС, HR)	Ранні системні реакції можуть проявлятися тахікардією, зміною варіабельності та нестабільністю ритму; важливі тренди та похідні характеристики [1; 19]	Часовий ряд (безперервний або квазібезперервний)	Чутливий до артефактів; потрібні фільтрація та обробка пропусків
Гемодинаміка	Середній артеріальний тиск (MAP)	Ранні зміни можуть проявлятися флуктуаціями та спадним трендом, що передує порушенню перфузії [1; 19]	Часовий ряд (монітор/манжета)	Нерегулярність при неінвазивних вимірюваннях; важлива синхронізація

Кінець табл. 1.1.

Група показників	**Життєвий показник (канал)**	**Чому інформативний для ранніх етапів**	**Очікуваний формат у даних**	**Коментарі до якості даних**
Дихальна система	Частота дихання (RR)	Один із ранніх, але часто недооцінених маркерів; зростання та нестабільність можуть передувати декомпенсації [1; 19]	Часовий ряд (квазібезперервний/періодичний)	Часті пропуски/помилки; потрібні QC та імпутація
Дихальна система	Сатурація кисню (SpO_2)	Нестійкі зниження та епізоди десатурації можуть відображати ранні порушення оксигенації [1; 19]	Часовий ряд (пульсоксиметрія)	Артефакти через перфузію/рух; доцільне згладжування/відсікання артефактів
Терморегуляція	Температура тіла	Лихоманка/гіпотермія важливі, однак ранність частіше проявляється у трендах і поєднанні з іншими сигналами [1; 19]	Часовий ряд (періодичний)	Різні місця вимірювання → різні значення; потрібна уніфікація
Похідні характеристики	Тренди, варіабельність, частота епізодів відхилень	Рання фаза частіше проявляється зміною динаміки та «дрейфом» від індивідуальної норми [20]	Ознаки з ковзних вікон (напр., 3/6/12 год)	Потрібне узгодження частоти дискретизації та протоколів
Крос-канальні взаємозв’язки	HR–MAP, RR–SpO_2 тощо	Важливі узгодженість/розузгодження між системами організму [3; 6]	Мультиваріантний ряд + крос-ознаки	Потрібні синхронізація та контроль пропусків по каналах
Додатково (за наявності)	Лабораторні/метаболічні маркери (напр., лактат)	Підсилюють інформативність, але не є безперервними; корисні як додаткові канали або для валідації [10]	Рідкісні вимірювання (подієві точки)	Потрібна стратегія інтеграції з часовими рядами (інтерполяція/маски)

Поруч з вимогами до «каналів» вимірювань потрібні вимоги до самої інтелектуальної системи як компонента клінічного процесу. Умовно їх поділяють на клінічні (релевантність, безпечність, інтеграція у робочий процес) і технічні (робастність, масштабованість, відтворюваність). Для підходів на основі Attention-VAE це означає, що модель має: (i) коректно працювати з нерівномірними та частково спостережуваними рядами; (ii) формувати стабільний сигнал аномальності; (iii) забезпечувати пояснюваність через виділення «проблемних» ознак/часових сегментів; (iv) підтримувати адаптацію під контекст відділення (калібрування порогів, контроль хибних тривог); (v) бути сумісною з режимом слабкого нагляду (грубі маркери подій або періодів підвищеного ризику без точного onset time) [21]. Вимоги до інтелектуальної системи раннього виявлення сепсису у клінічному та технічному вимірах узагальнено в таблиці 1.2.

Таблиця 1.2 – Вимоги до інтелектуальної системи раннього виявлення сепсису (клінічні та технічні аспекти)

Група вимог	**Вимога**	**Клінічна мотивація**	**Технічна інтерпретація/наслідок для моделі**
Чутливість до ранніх змін	Виявлення слабких, субклінічних відхилень	Ранні ознаки можуть бути непороговими та короткими	Орієнтація на патерни/динаміку; використання вікон і реконструкційних/імовірнісних критеріїв
Контроль хибних тривог	Обмеження частоти спрацювань	Втома від тривог (*alarm fatigue*) знижує довіру та ефективність	Калібрування порога τ, згладжування сигналу, правила персистентності тривоги [8; 25]
Робастність до якості даних	Стійкість до пропусків, артефактів, асинхронності	Реальні ICU-дані неповні та зашумлені	Маски пропусків, імпутація, вирівнювання часу, оцінка якості сигналів

Кінець таб. 1.2

Група вимог	**Вимога**	**Клінічна мотивація**	**Технічна інтерпретація/наслідок для моделі**
Інтерпретованість	Пояснення «чому» та «де саме» підвищився ризик	Потрібна клінічна довіра та можливість перевірки	Attention heatmaps, ранжування внеску ознак, локальні пояснення для конкретного вікна
Мінімальна затримка	Підтримка near-real-time моніторингу	Цінність сигналу знижується із часом	Інкрементальний розрахунок ознак/балів; обмеження складності інференсу
Узгодженість із workflow	Підтримка процесу «сигнал → пояснення → дія лікаря»	Система — підтримка рішень, а не діагноз	Інтерфейс оповіщення, журнали подій, інтеграція з EHR/моніторами [8]
Відтворюваність та аудит	Трасування рішень і версій	Необхідні безпека й відповідальність	Логування вхідних даних/версій моделі/порогів; можливість ретроспективного аналізу
Узагальнюваність	Збереження якості на різних підгрупах пацієнтів	Висока гетерогенність ICU-популяції	Перевірка на підвибірках пацієнтів, контроль bias за демографічними ознаками, стратифікований split

Клінічні обмеження традиційних шкал (реактивність, дискретність, нечутливість до часових патернів) формують підґрунтя для переходу до систем безперервного аналізу багатовимірних часових рядів. Для таких систем критичні чотири властивості: робастність до пропусків і шумів, контроль хибних тривог, інтерпретованість, здатність працювати в умовах неповної/нечіткої розмітки. Вони природно відповідають постановці задачі як виявлення аномалій (*unsupervised* / *weakly supervised*) на даних ВІТ.

1.2 Огляд методів машинного навчання для аналізу багатовимірних часових рядів

Аналіз багатовимірних медичних часових рядів (multivariate medical time-series, MTS) складний з кількох причин: нерегулярність вимірювань, пропуски, артефакти сенсорів, сильна залежність динаміки показників від лікувальних втручань і індивідуальних особливостей пацієнта. У задачі раннього виявлення сепсису ці фактори ускладнюють як формування «правильної» цілі (мітки), так і формалізацію того, що саме вважати «відхиленням» від норми. У літературі умовно виокремлюються дві великі парадигми: навчання з учителем (*supervised*: класифікація чи прогнозування події за розміченими прикладами) та навчання без учителя або зі слабким наглядом (*unsupervised/weakly supervised*: моделювання типової динаміки й виявлення аномалій як відхилень). Вибір парадигми визначається не лише наявністю розмітки, але й вимогами до клінічної інтерпретованості, стійкості до зсувів розподілу, здатності працювати в режимі «раннього попередження», а також ризиками помилкових тривог у реальному стаціонарному середовищі [3; 19].

У **supervised**-підходах раннє виявлення сепсису найчастіше формулюється як класифікація вікна спостережень: за вхідним фрагментом часових рядів (наприклад, останні (W) годин) модель оцінює ймовірність події у найближчому горизонті або на поточний момент. Для навчання використовуються мітки, побудовані на основі клінічних критеріїв (Sepsis-3, SOFA), тригерів лікування (початок антибіотикотерапії), лабораторних показників або записів у медичній документації [9; 19]. Представлення даних у supervised-постановці зазвичай реалізується двома способами: **(а)** агрегування часових рядів у табличні ознаки (статистики, тренди, похідні, екстремуми, частоти пропусків) з подальшим застосуванням класичних моделей (логістична регресія, градієнтний бустинг, випадковий ліс); **(б)**

подача послідовностей у моделі, що безпосередньо моделюють часову залежність (RNN/LSTM/GRU, Temporal CNN, Transformer, attention-механізми) [16; 23]. Сильна сторона supervised-парадигми: пряме узгодження оптимізації з цільовою подією і зручність оцінювання стандартними метриками класифікації. Працюють ці переваги лише за умови, що мітки якісні і стабільні.

Ключова проблема supervised-підходів для сепсису: **мітка є подієвою і часово-нечіткою**. «Початок» сепсису рідко має однозначну часову відмітку, а ретроспективне визначення часу onset часто залежить від доступності лабораторних даних, практик документування та рішень лікаря [21; 19]. У результаті модель може навчитися відтворювати не ранні фізіологічні зміни, а **проксі-ознаки** на кшталт частоти лабораторних призначень, появи певних записів чи процедур (тобто «відбиток протоколів лікування»). Це формально корелює з міткою, але не гарантує клінічної корисності на ранньому горизонті [9]. Додатково: дисбаланс класів (низька частота сепсису відносно всіх перебувань у відділенні) стимулює застосування ваг класів, ресемплінгу або специфічних функцій втрат. А це підвищує ризик перенавчання на специфічних підгрупах пацієнтів і погіршує калібрування ймовірностей [25]. Окрема проблема — *dataset shift*: перенесення моделі між лікарнями, відділеннями або часовими періодами часто супроводжується деградацією якості через відмінності у протоколах, обладнанні та практиках вимірювання [9].

На практиці, навіть у *supervised*-постановці, підходи до аналізу MTS варто розрізняти за принципом формування «сигналу ризику» і типом часової динаміки. Далі використовується узгоджена концептуальна класифікація на чотири групи: **класифікаційні**, **прогнозні**, **реконструкційні** та **щільнісні** (density-based). Така класифікація зручна для раннього виявлення сепсису: вона прямо відображає, *що саме* вважається сигналом небезпеки — ймовірність події (класифікація), помилка передбачення майбутнього

(прогнозування), помилка відновлення поточного фрагмента (реконструкція) або низька правдоподібність стану в імовірнісній моделі (щільнісні) [10; 7].

Класифікаційні підходи безпосередньо навчаються відображати фрагмент (_{t-W:t}) у ймовірність події, оптимізуючи функцію втрат відносно міток. Перевага: «пряма» оптимізація цілі. Обмеження пов’язані з шумністю і неоднозначністю міток, а також з ризиком прихованого витоку інформації (*leakage*), коли вхідні ознаки містять сліди вже прийнятих клінічних рішень [9]. Для раннього попередження важлива не лише якість на рівні AUROC, а й поведінка моделі у часі: чи зростає ризик за години до клінічного підтвердження, чи стабільна тривога, чи не надто висока частота хибних спрацювань у стабільних пацієнтів [8; 25].

Прогнозні підходи формулюють задачу як передбачення наступних значень (_{t:t+t}) або певних агрегованих індикаторів на горизонті (t). Ризик інтерпретується через збільшення похибки прогнозу чи зміну параметрів прогнозної моделі. Перевага прогнозування: можливість self-supervised навчання без явних міток сепсису й використання великих масивів даних з мінімальною ручною анотацією [16; 3]. Але є нюанс: у клініці прогнозні помилки часто відображають зміни лікування (інфузії, вазопресори, киснева підтримка) або зміну режиму моніторингу, а не початок патологічного процесу. Це створює ризик хибних тривог у моменти терапевтичних втручань. Потрібно враховувати контекст лікування як додаткові змінні або явно моделювати інтервенції [8]. Прогнозні моделі також бувають недостатньо чутливими до короткочасних аномальних сегментів, якщо функція втрат «усереднює» помилки по часу і домінують інерційні тренди.

Реконструкційні підходи, зокрема автоенкодери (AE) і варіаційні автоенкодери (VAE), будують стислий латентний опис () і відновлюють вхід (). Аномальність пов’язується зі збільшенням реконструкційної похибки або відхиленням латентних кодів від типових регіонів латентного простору. Для раннього виявлення сепсису реконструкційна парадигма приваблива одразу двома моментами. По-перше, вона дозволяє виявляти *відхилення від*

нормальної фізіології до того, як подія формально «відбулась» у термінах критеріїв. По-друге, природно працює з незбалансованістю класів [1; 3; 7]. Практичні обмеження цієї парадигми пов’язані з визначенням «норми» в навчальних даних: якщо у навчальному наборі присутня значна частка патологічних патернів або дані містять систематичні артефакти, модель може навчитися відновлювати небажані режими як «типові», знижуючи чутливість до ранніх сигналів [18]. Тому важливі три речі: стратегія відбору навчальних сегментів, робастні функції втрат і механізми контролю якості даних.

Щільнісні підходи оцінюють правдоподібність спостережень (p()) або умовну правдоподібність (p()). Аномалії визначаються як спостереження з низькою правдоподібністю. Концептуально ця парадигма дає природний «ризик-скоринг» у термінах імовірностей і потенційно придатна для калібрування порогів тривоги та порівняння станів між пацієнтами [1; 7]. Але щільнісне моделювання у високій розмірності чутливе до вибору сімейства розподілів, регуляризації, режиму навчання та якості імпутації пропусків. У клінічних MTS низька правдоподібність може відображати і патологію, і рідкісні, але фізіологічно допустимі стани або технічні артефакти. Тому потрібні механізми перевірки достовірності тривоги та узгодження з клінічним контекстом [7; 11].

Проміжним варіантом між supervised і unsupervised є **weakly supervised** підходи. Вони використовують неповну або «слабку» розмітку: наприклад, відомо лише, що у певному перебуванні пацієнта подія сепсису сталась (bag-level label), але точний момент onset невідомий; або наявні лише позитивні приклади, а решта — невизначені (positive-unlabeled, PU). Такі постановки за методом ближчі до реалій клініки. Вони дозволяють уникнути жорсткої прив’язки до нечітких часових міток і будувати механізми навчання, що фокусуються на сегментах з найбільшою «підозрілістю» всередині довгої траєкторії [21]. Для часових рядів це природно поєднується з attention-механізмами: модель присвоює різні ваги часовим крокам/ознакам, формуючи

ризикову оцінку як агрегат з виділенням найбільш інформативних сегментів. Останнє важливе для пояснюваності у ранньому попередженні [4; 5].

Порівняння парадигм варто проводити не лише за точністю, але й за **експлуатаційними властивостями**, критичними для систем раннього попередження: стабільність сигналу в часі, частота хибних тривог, переносимість між доменами, стійкість до пропусків, здатність формувати локалізовані пояснення (які показники і які часові сегменти «підняли» ризик), можливість калібрування порогу тривоги під ресурсні обмеження відділення інтенсивної терапії [8; 25]. За такими критеріями unsupervised/weakly supervised підходи добре відповідають природі раннього виявлення сепсису: процес поступового відхилення фізіології, а не миттєва зміна дискретного класу. Але для практичного застосування потрібні процедури контролю якості даних, робастне визначення «норми», а також механізми калібрування та валідації, які демонструють, що сигнал аномальності корелює з клінічно значущими подіями й не є лише реакцією на зміну протоколів лікування [10; 18]. Узагальнене порівняння парадигм аналізу багатовимірних часових рядів наведено в таблиці 1.3.

Таблиця 1.3 – Порівняння парадигм аналізу MTS для раннього виявлення сепсису

Парадигма	**Сигнал ризику**	**Потреба в мітках**	**Типові моделі**	**Сильні сторони**	**Основні обмеження для сепсису**
Supervised (класифікація)	(P(_{t-W:t}))	Висока	LR/GBM, LSTM/TCN, Transformer	Пряма оптимізація цілі; стандартні метрики	Шум/нечіткість *onset*; *leakage*; *dataset shift*; калібрування
Self/unsupervised (прогнозні)	Похибка прогнозу наступних значень часового ряду	Низька	Seq2seq, TCN, Transformer-forecast	Навчання без міток; використовує багато даних	Чутливість до інтервенцій; нестаціонарність; короткі аномалії

Кінець табл. 1.3

Парадигма	**Сигнал ризику**	**Потреба в мітках**	**Типові моделі**	**Сильні сторони**	**Основні обмеження для сепсису**
Unsupervised (реконструкція)	Похибка реконструкції вхідного вікна	Низька	AE/VAE, Attention-AE/VAE	Узгодженість із “early deviation”; робота з дисбалансом	Contamination; пороги; артефакти/пропуски; контроль “норми”
Density-based	Від’ємна лог-правдоподібність -log p(X) або -log p(X	z)	Низька	VAE, flows, latent density	Імовірнісна інтерпретація; потенціал калібрування
Weakly supervised	Bag-level ризик + attention	Середня	MIL+attention, PU+representation	Менше залежить від точного onset; пояснюваність	Потрібні припущення про “мішки”; складна валідація; ризик bias

Наведена класифікація відображає, що вибір парадигми визначає не лише механізм формування сигналу ризику, а й характер типових помилок: supervised-класифікатор ризикує відтворювати артефакти розмітки, тоді як реконструкційні підходи чутливі до якості визначення «норми» у навчальних даних [10; 19]. Концептуальну класифікацію розглянутих підходів за принципом дії, ключовими припущеннями та типовими джерелами помилок у ICU-даних подано в таблиці 1.4.

Таблиця 1.4 – Концептуальна класифікація підходів: принципи, припущення та типові джерела помилок

Група підходів	**Принципова ідея**	**Ключове припущення**	**Що може «порушити» підхід у ICU**	**Який контроль потрібен**
Класифікаційні	Навчання на мітках події	Мітки узгоджені та стабільні	Нечіткий *onset*, зміна протоколів, *leakage*	Контроль *leakage*, часова валідація, калібрування
Прогнозні	Норма = передбачуваність майбутнього	Майбутнє прогнозоване у «нормі»	Інтервенції, різка зміна режиму лікування	Облік втручань, робастні втрати, контекстні змінні
Реконструкційні	Норма = здатність відновити вхід	Training переважно «нормальний»	Contamination, артефакти, пропуски	Відбір/фільтрація, імпутація, пороги, robust loss
Щільнісні	Норма = висока правдоподібність	Можливо оцінити $p(X)$	Висока розмірність, шум, пропуски	Регуляризація, перевірка стабільності, інтерпретація
Weakly supervised	Слабкі мітки + виділення сегментів	Слабка мітка корелює з раннім сегментом	Bias у «мішках», шумні позитиви	MIL/PU протоколи, sensitivity analysis, temporal sanity-check

Управління цими ризиками потребує явного формулювання припущень про «норму», стратегій захисту від leakage та протоколів зовнішньої валідації — питань, що визначають відтворюваність результатів незалежно від архітектурного вибору [9; 18].

Формальні означення та типові функції втрат (приклади)

Нехай $X_{t-W:t} \in \mathbb{R}^{W \times d}$ — вікно багатовимірних спостережень довжини W з d каналами. У *supervised*-постановці прогнозується ймовірність події $\hat{y}=\sigma(f_\theta(X_{t-W:t}))$, де $\sigma(\cdot)$ — логістична функція, а f_θ — параметризована модель. Втрата перехресної ентропії для бінарної класифікації записується у вигляді (1.5):

$$\mathcal{L}_{CE}(y,\hat{y})=-(y\log \hat{y}+(1-y)\log(1-\hat{y})) \quad (1.5)$$

У прогнозній постановці будується $\hat{X}(t:t+\Delta t)=g_\theta(X(t-W:t))$, а сигнал аномальності може бути пов’язаний із помилкою прогнозу за формулою (1.6):

$$s_{pred}(t)=X_{t:t+\Delta t}-\hat{X}_{t:t+\Delta t}{}_2^2 \quad (1.6)$$

У реконструкційній (AE/VAE) постановці модель відновлює $\hat{X}_{t-W:t}$, а простий скоринг аномальності визначається як реконструкційна похибка за формулою (1.7):

$$s_{rec}(t)=\frac{1}{Wd}\sum_{i=1}^{W}\sum_{j=1}^{d}(X_{i,j}-\hat{X}_{i,j})^2 \quad (1.7)$$

Для VAE типовою є оптимізація нижньої межі правдоподібності (ELBO), що поєднує якість відновлення та регуляризацію латентного простору, відповідно до формули (1.8):

$$\mathcal{L}_{ELBO}(\theta,\varphi)=E_{q_\varphi(z \mid X)}[\log p_\theta(X \mid z)]-D_{KL}(q_\varphi(z \mid X);\mid;p(z)) \quad (1.8)$$

У рамках Attention-VAE додатково може використовуватися увага (^{W}) як механізм виділення часових сегментів, що найбільше впливають на скоринг (без фіксації конкретної реалізації), згідно з формулою (1.9):

$$s_{att}(t)=\sum_{i=1}^{W}\alpha_i \cdot x_{t-W+i}-\hat{x}_{t-W+i}2^2,\ \sum_{i=1}^{W}\alpha_i=1 \quad (1.9)$$

Наведені формалізації демонструють відмінності між парадигмами на рівні функцій втрат і визначення скорингу ризику. Для клініки важливо інше:

одна і та сама модель адаптується до різних постановок. Наприклад, архітектура з attention застосовується і для класифікації, і для weakly supervised агрегації сегментів, і для пояснення того, які часові відрізки сформували високий ризик [3; 6].

Огляд методів аналізу багатовимірних медичних часових рядів засвідчує: класифікаційні *supervised*-моделі ефективні за умови надійної розмітки, однак у задачі раннього виявлення сепсису суттєво обмежені часовою нечіткістю *onset*, потенційним витоком інформації (*leakage*) та чутливістю до зсувів розподілу. Прогнозні *self-supervised* підходи зменшують залежність від міток, але потребують обліку терапевтичних інтервенцій і нестаціонарності. Реконструкційні та щільнісні підходи — природні кандидати для раннього попередження через моделювання нормальної динаміки й детекцію відхилень. Проте вони потребують процедур контролю «норми», калібрування порогів і робастності до пропусків та артефактів. Підходи зі слабким наглядом (*weakly supervised*) дозволяють формувати оцінку ризику без жорсткого визначення моменту *onset* та одночасно отримувати локалізовані пояснення через механізми уваги [20]. Саме це поєднання обґрунтовує вибір архітектурного підходу в роботі.

1.3 Аналіз існуючих підходів та позиціонування Attention-VAE

Для обґрунтування доцільності запропонованої архітектури розглянуто найближчі аналоги у літературі за чотирма вимірами: архітектура, парадигма навчання, наявність механізму уваги, домен застосування.

Прямими архітектурними попередниками Attention-VAE є дві роботи. OmniAnomaly [38] поєднує GRU з VAE, доповненим стохастичним з’єднанням прихованих змінних і planar normalizing flow; на трьох стандартних benchmark-датасетах (SMD, SMAP, MSL) досягнуто F1 = 0,86, що підтверджує ефективність GRU-VAE як базової архітектури для послідовних даних. MA-

VAE [39] архітектурно найближча до запропонованої: BiLSTM-VAE з механізмом multi-head attention для виявлення аномалій; автори показують, що multi-head attention суттєво підвищує інтерпретованість рішень. Принципова відмінність Attention-VAE від MA-VAE — використання GRU замість BiLSTM (менша кількість параметрів при порівнянній якості [43]) і застосування до біомедичного домену.

DTAE [40] є єдиною знайденою роботою, що поєднує несупервізований підхід з виявленням сепсису: Transformer-based denoising autoencoder з детермінованим декодером. Це позбавляє її можливості кількісно оцінювати невизначеність у латентному просторі, на відміну від Attention-VAE.

Серед supervised-підходів орієнтиром якості є робота [41]: LSTM-модель з механізмом уваги показує AUC = 0,892 на вибірці понад 100 000 пацієнтів. Систематичний огляд 80 досліджень [42] фіксує типові значення ROC-AUC у діапазоні 0,79–0,96 для ML-систем виявлення сепсису. Переважна більшість підходів є supervised-класифікаторами, що підкреслює гостру потребу у несупервізованих методах.

Таблиця 1.5 систематизує порівняння запропонованого підходу з найближчими аналогами.

Таблиця 1.5 – Порівняння Attention-VAE з існуючими підходами до виявлення аномалій у медичних часових рядах

Модель	**Архітектура**	**Unsup.**	**Attn.**	**Домен**	**Джерело**
OmniAnomaly [38]	GRU-VAE + NF	+	−	IT-моніторинг	KDD 2019
MA-VAE [39]	BiLSTM-VAE + MHA	+	+	Автомобільний	NCTA 2023
DTAE [40]	Transformer-DAE	+	+	Сепсис	AAAI 2022

Кінець табл. 1.5

Модель	**Архітектура**	**Unsup.**	**Attn.**	**Домен**	**Джерело**
Zhang et al. [41]	LSTM + Attention	−	+	Сепсис	Patterns 2021
Attention-VAE (наша робота)	GRU-VAE + MHA	+	+	Сепсис / PhysioNet 2019	[ця робота]

Запропонована архітектура спирається на встановлені у літературі принципи: поєднання GRU-VAE для несупервізованого виявлення аномалій [38] та інтеграція multi-head attention у VAE [39]. Внесок цієї роботи: адаптація такої комбінації до домену біомедичних часових рядів ВІТ. Подвійний механізм інтерпретації (часовий — карти уваги: коли відбувається погіршення; ознаковий — декомпозиція реконструкційної похибки: що є аномальним) орієнтований на клінічні вимоги до пояснюваності у системах підтримки прийняття рішень.

1.4 Варіаційні автоенкодери та механізми уваги у задачах діагностики

Варіаційні автоенкодери (VAE) належать до класу імовірнісних генеративних моделей, що поєднують нейромережеве кодування даних із явним моделюванням латентного простору як випадкової змінної. На відміну від детермінованих автоенкодерів, VAE не лише відтворюють вхідні спостереження, але й формують узагальнене латентне представлення, параметризоване розподілом (зазвичай середнім і дисперсією), і регуляризоване відносно апріорного розподілу. Для клінічних задач це принципово, оскільки фізіологічні сигнали характеризуються високою міжпацієнтською варіативністю, наявністю шуму, пропусків, артефактів

моніторингу, а також неоднорідністю профілів пацієнтів і протоколів лікування. Через це жорстко детерміновані моделі та прості порогові критерії стають менш придатними [8; 10].

Для раннього виявлення сепсису особливу практичну цінність мають підходи unsupervised / weakly supervised. Момент «початку» (onset) часто розмитий і залежить від протоколів, доступності лабораторних вимірювань і суб'єктивної інтерпретації клініциста. У реєстрових клінічних даних мітки відображають факт встановлення діагнозу або виконання певних процедур, але не обов'язково віддзеркалюють ранні фізіологічні зміни, що передують клінічному підтвердженню [35; 10]. За таких умов генеративні моделі, які вивчають розподіл «типових» режимів і виявляють відхилення від нього, добре відповідають постановці задачі як детекції аномалій у багатовимірних часових рядах.

Додатковий аргумент на користь VAE: можливість отримати скалярний показник «аномальності» на основі похибки реконструкції або (негативної) оцінки правдоподібності, а також потенційна пояснюваність через аналіз внеску каналів і часових сегментів. Проте для медичних часових рядів критично врахувати, що патологічні відхилення часто локалізуються у часі, проявляються слабко і не синхронізуються між усіма каналами. Саме тому VAE доцільно підсилювати механізмами уваги (Attention). Вони дозволяють моделі зосереджуватися на найбільш інформативних фрагментах послідовності та на найрелевантніших ознаках, підвищуючи і якість представлення, і інтерпретованість рішення [3; 6].

У формальному вигляді VAE задає імовірнісну модель спостережень x через латентні змінні z, де x інтерпретується як фрагмент (вікно) багатовимірного фізіологічного часового ряду, а z — компактне латентне представлення стану пацієнта. Генеративна частина визначається апріорним розподілом $p(z)$ та умовним розподілом декодера $p_\theta(x \mid z)$, який моделює процес відновлення (генерації) часових рядів із латентного простору. Оскільки точний апостеріорний розподіл $p_\theta(z \mid x)$ є обчислювально недоступним,

вводиться варіаційна апроксимація $q_\varphi(z \mid x)$, яка реалізується енкодером. Оптимізація виконується шляхом максимізації варіаційної нижньої межі лог-правдоподібності (ELBO), яка задається формулою (1.10):

$$\mathcal{L}_{(ELBO)}(\theta,\varphi;x) = E_{(q_\varphi(z \mid x))}[\log p_\theta(x \mid z)]D_{(KL)}(q_\varphi(z \mid x) \,,\mid, p(z)). \quad (1.10)$$

Перший доданок є терміном відтворення і відображає здатність моделі реконструювати спостережувані часові ряди. Для фізіологічних даних це відповідає навчанню відтворювати типову динаміку життєвих показників у заданому вікні, зберігаючи часові залежності та взаємозв'язки між каналами. У режимі детекції аномалій зростання реконструкційної похибки або зниження лог-правдоподібності відтворення може інтерпретуватися як ознака нетипового (потенційно патологічного) патерну у часовому ряді [7; 9]. Другий доданок — KL-дивергенція — регуляризує латентний простір, зменшуючи ризик «запам'ятовування» шумових флуктуацій і артефактів як детермінованих закономірностей, що є особливо важливим за високої міжпацієнтської варіативності та нерегулярності вимірювань [8; 10].

З практичної точки зору баланс між відтворенням та регуляризацією визначає компроміс між точністю реконструкції та узагальнювальною здатністю моделі. Для раннього виявлення сепсису це означає необхідність налаштування так, щоб модель достатньо добре відтворювала нормальні або стабільні фізіологічні режими, але водночас демонструвала чутливість до слабких, локалізованих у часі відхилень. Поширений інструмент керування цим балансом — введення коефіцієнта β у регуляризаційний член (β-VAE). Він явно підсилює або послаблює структурованість латентного простору залежно від цілей задачі за формулою (1.11):

$$\mathcal{L}_{(\beta\text{-}VAE)}(\theta,\varphi;x) = E_{(q_\varphi(z \mid x))}[\log p_\theta(x \mid z)]\beta \,, D_{(KL)}(q_\varphi(z \mid x) \,,\mid, p(z)), \beta>0. \quad (1.11)$$

Для медичних часових рядів надто мала регуляризація може призводити до надмірної підгонки до шумів та артефактів, тоді як надто велика — до втрати чутливості до тонких змін (наприклад, коротких епізодів десатурації або нестабільності дихання). Тому вибір β потребує експериментального налаштування на валідаційній схемі; у роботі обрано $\beta_{(max)} = 0.5$ з лінійним розігрівом упродовж перших 30 епох (KL annealing) — це дає стабільне навчання без posterior collapse [2; 4].

Для моделювання часових залежностей енкодер і декодер VAE реалізуються різними архітектурами: рекурентними мережами (RNN/GRU/LSTM), згортковими темпоральними моделями (TCN), а також трансформерними блоками. Для аналізу багатовимірних часових рядів життєвих показників підходять архітектури, що підтримують багатоканальність, працюють з масками пропусків і враховують нерегулярність дискретизації (наприклад, через часові ембеддинги або окремі канали «time-since-last-observation») [8]. Для набору PhysioNet Challenge 2019 [35] застосовується ресемплінг із кроком $\Delta t = 5$ хвилин і лінійна інтерполяція з явним маскуванням прогалин; деталі пайплайну описано у розділі 2.

На практиці детекції аномалій на основі VAE важливо коректно визначити аномальний бал і забезпечити порівнюваність між каналами з різними масштабами і шумовими характеристиками. Один з узагальнених підходів: маскована нормалізована похибка реконструкції як скалярний показник аномальності A(t), визначений формулою (1.3); позначення: $\hat{x}_i(t)$ — реконструкція, σ_i^2 — оцінка дисперсії (або іншого масштабу) каналу, $m_i(t) \in 0,1$* — маска спостереження, w_i* — ваги каналів, ε — стабілізатор. У клінічній інтерпретації високі значення $A(t)$ означають, що поточний патерн суттєво відхиляється від типового режиму, засвоєного моделлю на переважно стабільних даних. Для реалізації підходів weakly supervised застосовуються грубі мітки періодів підвищеного ризику або «позитивні вікна» без точного onset, наприклад для налаштування порога або для навчання калібратора сигналу ризику [20].

Незбалансованість — серйозна проблема медичних даних: патологічні стани (зокрема сепсис) трапляються значно рідше, ніж нормальні або стабільні стани. Для класифікаційних моделей це спричиняє зміщення рішень у бік домінуючого класу і потребує стратегій балансування, які можуть погіршувати переносимість. Unsupervised детекція аномалій природно узгоджується з цією властивістю: модель навчається на великому обсязі переважно «нормальних» даних і виявляє рідкісні відхилення без необхідності формувати репрезентативний набір позитивних прикладів [7; 16]. Водночас у клінічному середовищі «норма» умовна: у відділенні інтенсивної терапії багато пацієнтів перебувають у нестабільних станах без сепсису, тому навчальна вибірка може містити різноманітні відхилення іншої етіології. Це вимагає обережного трактування «аномальності» як сигналу підвищеного ризику, а не як прямого діагнозу сепсису [16; 17].

Для багатовимірних часових рядів критично врахувати, що різні життєві показники мають неоднакову діагностичну вагу, а патологічні відхилення часто локалізуються у часі. Механізми уваги (Attention) дають змогу моделі виділяти найбільш інформативні сегменти послідовності та канали, що найбільше впливають на латентне представлення або реконструкцію. На практиці це зменшує вплив нерелевантного шуму і концентрує модельний ресурс на критичних сегментах, де спостерігаються зміни, потенційно пов’язані з початком септичного процесу [3; 6]. Для багатовимірних сигналів увага реалізується як увага за часом, за змінними або у комбінованій формі (наприклад, матриця «час × показник»). Це особливо доречно для ситуацій, коли аномалія проявляється коротко і лише у частині каналів [3].

У загальному вигляді механізм уваги може бути описаний як обчислення ваг α_t або $\alpha_{(t,i)}$ через нормалізовану (softmax) функцію подібності між запитом (query) та ключами (keys), з подальшим агрегуванням значень (values). Для часової уваги в одному вікні запис має вигляд (1.12):

$$e_t = (q^{\top} k_t)/(\sqrt{(d_k)}),\ \alpha_t = (\exp(e_t))/(\Sigma_{(u=1)}^{T} \exp(e_u)),\quad (1.12)$$

$$c=\Sigma_{(t=1)}^{T} \alpha_t, v_t,$$

де q — запит, k_t — ключ для кроку t, v_t — значення, c — контекстний вектор. У VAE-архітектурі контекст може використовуватися для формування параметрів $q_\varphi(z \mid x)$ (енкодер) або для реконструкції $p_\theta(x \mid z)$ (декодер), а також для побудови пояснень. Практична інтерпретація полягає у тому, що вищі значення α_t вказують на часові сегменти, які модель вважає найбільш інформативними для поточного представлення або реконструкції. Проте важливо підкреслити, що ваги уваги не є прямим доказом причинного впливу і потребують обережної інтерпретації та додаткової валідації [33].

Інтерпретованість входить до ключових вимог клінічних систем підтримки рішень: окрім сигналу ризику, потрібне пояснення, які показники та часові ділянки найсильніше вплинули на спрацювання. У Attention-VAE поєднуються два джерела пояснень: аналіз ваг уваги для локалізації підозрілих сегментів і аналіз структурованої похибки реконструкції за каналами та часом для визначення, які компоненти сигналу відтворюються гірше. Узгодженість цих двох аспектів підвищує довіру до системи, але не усуває потреби у клінічній перевірці та врахуванні контексту лікування [33; 7].

Оскільки дані Challenge 2019 відображають реальні процеси лікування, у часових рядах присутні «керовані» зміни, спричинені втручаннями (інфузійна терапія, вазопресори, зміни режимів ШВЛ). Вони істотно змінюють динаміку життєвих показників. Для детектора аномалій це створює ризик плутанини між патологічними відхиленнями та змінами, зумовленими лікуванням. Зменшити цей ризик можна двома кроками: коректно визначити «нормальний» режим навчання і ретельно поставити задачу як раннє попередження (risk scoring), а не автоматичне встановлення діагнозу. У роботі використовуються п’ять вітальних показників (HR, SpO_2, MAP, температура, RR) без контекстних змінних втручань, що відповідає стандартному підходу для unsupervised виявлення аномалій у даних ICU [35]. Узагальнені відомості

про компоненти VAE для часових рядів та їхню роль у задачі детекції аномалій надано в таблиці 1.6.

Таблиця 1.6 – Компоненти VAE для часових рядів та їхня роль у задачі детекції аномалій

Компонент	**Формаліза-ція**	**Роль у моделі**	**Практичне значення для ICU-даних**	**Типові ризики/обмеження**
Апріорний розподіл	$p(z)$ (часто $N(0,I)$)	Задає «еталон» для латентного простору	Сприяє гладкості представлень і стабільності генерації	Надто сильне спрощення може знижувати виразність
Енкодер (варіаційний)	$q_\varphi(z \mid x)$	Оцінює апостеріорний розподіл латентних змінних	Кодує типові режими та індивідуальні патерни у компактну форму	Чутливість до пропусків і артефактів без маскування
Декодер (генеративний)	$p_\theta(x \mid z)$	Відтворює часовий ряд із z	Дозволяє оцінити «нетиповість» через похибку реконструкції	Ризик «усереднення» та згладжування локальних подій
ELBO-оптимізація	$E[\log p_\theta(x \mid z)]-D_{KL}(\cdot)$	Компроміс між реконструкцією і регуляризацією	Контролює узагальнення за міжпацієнтської варіативності	Потребує налаштування балансу, інакше — under/overfit
Аномальний бал	$A(t)$ або $-\log p_\theta(x \mid z)$	Вихідний сигнал ризику/аномальності	Використовується для тривог і ранжування підозрілих вікон	Потрібні калібрування порога та контроль false alarms

Можливі варіанти інтеграції механізмів уваги в архітектуру VAE для багатовимірних часових рядів узагальнено в таблиці 1.7.

Таблиця 1.7 – Варіанти інтеграції механізмів уваги в VAE для багатовимірних часових рядів

Тип уваги	**Об'єкт фокусування**	**Приклад інтерпретації**	**Переваги**	**Потенційні обмеження**
Часова увага	Часові кроки/сегменти *t*	Які часові інтервали вікна є «критичними»	Локалізація подій, фокус на коротких епізодах	Ваги не гарантують причинності [33]
Ознакова увага	Канали/показники *i*	Які життєві показники найбільше впливають	Ранжування каналів, зменшення впливу шумних ознак	Може ігнорувати важливі крос-канальні патерни
Спільна увага	Матриця *(t,i)*	Які сегменти та канали важливі одночасно	Детальне пояснення «де і що» змінилося	Вища обчислювальна складність
Self-attention у енкодері	Внутрішні залежності в послідовності	Моделювання дальніх залежностей	Підсилення представлення без ручних ознак	Потребує достатнього обсягу даних і регуляризації
Увага в декодері	Реконструкція з фокусом на сегментах	Покращення відтворення складних фрагментів	Потенційно підвищує чутливість до локальних відхилень	Ризик «пере-реконструкції» аномалій без контролю через маскування реконструкційного терму

Доцільність використання VAE з механізмом *attention* у задачах медичної діагностики визначається поєднанням п’яти властивостей: здатність працювати в умовах неповної та шумної інформації; відповідність *unsupervised* / *weakly supervised* постановці за дефіциту надійної розмітки; стійкість до класової незбалансованості; здатність враховувати часову локалізацію та неоднорідну важливість різних життєвих показників; інтерпретованість через аналіз ваг уваги та структури похибки реконструкції [3; 33]. Водночас коректне застосування таких моделей потребує чіткого формулювання того, що саме вважається «нормальним» режимом навчання, як обробляються пропуски та нерегулярність, як калібрується поріг тривоги і як валідується інтерпретація уваги в клінічно осмислений спосіб [33].

Висновки до розділу 1

У першому розділі встановлено клінічні та методологічні передумови дослідження. Показано, що традиційні інструменти скринінгу сепсису (SOFA, qSOFA) реактивні і нечутливі до часових патернів, тоді як безперервний моніторинг життєвих показників у ВІТ утворює потік даних, придатний для раннього виявлення відхилень від норми. Огляд парадигм аналізу медичних часових рядів засвідчив, що *supervised*-класифікація обмежена неоднозначністю часових міток *onset* сепсису та ризиком *leakage*. Натомість *unsupervised* і *weakly supervised* підходи добре відповідають природі задачі як поступового відхилення фізіології від норми. Ключове рішення розділу: обґрунтування архітектури Attention-VAE. Поєднання імовірнісного генеративного моделювання (VAE), механізму *attention* для часової локалізації відхилень і здатності до інтерпретованості задовольняє як технічні вимоги до якості детекції, так і клінічні вимоги до пояснюваності сигналів тривоги. Сформульовані вимоги до вхідних даних і моделі — вхідні умови для математичної формалізації, яка розгортається у розділі 2. Порівняльний аналіз

архітектур-аналогів засвідчив, що Attention-VAE є першою роботою, що поєднує GRU-VAE з multi-head attention для несупервізованого виявлення сепсису, заповнюючи прогалину між існуючими підходами [38; 39; 40].

Розділ 2
Математичне та алгоритмічне забезпечення системи виявлення аномалій

У **Розділі 2** представлено математичне та алгоритмічне забезпечення системи. Формалізовано задачу виявлення сепсису як задачу пошуку аномалій у багатовимірних часових рядах, розроблено архітектуру Attention-VAE, описано алгоритм передобробки даних та формування навчальної вибірки з набору PhysioNet Challenge 2019 [35].

2.1 Формалізація задачі виявлення сепсису як пошуку аномалій у багатовимірних часових рядах

Рання діагностика сепсису у ВІТ упирається у три фактори водночас: багатовимірність фізіологічних показників, нестаціонарність їх динаміки та затримку клінічних міток (вони часто з’являються вже після початку патологічного процесу). Класична класифікація тут пасує погано. Природний крок: пошук відхилень від «норми» (*anomaly detection*). Модель навчається на нормальних (або переважно нормальних) траєкторіях і сигналізує про нетипові патерни без прямого використання міток сепсису під час навчання [19; 21].

Для клінічних часових рядів *unsupervised* / *weakly supervised* є обґрунтованим вибором. «Норма» тут не один режим, а сукупність типових: стабільний стан, післяопераційний період, реакція на лікування. Сепсис проявляється по-різному: то як поступове накопичення слабких субклінічних змін, то як різкий злам трендів і кореляцій між ознаками. Відповідно, «аномалія» становить не одиничний викид, а структурне порушення часової залежності чи взаємозв’язків між показниками [10; 3]. До цього додається

практична вимога: вихідний сигнал моделі має відповідати рішенням лікаря. По-перше, інтегральна оцінка ризику на інтервалі спостереження. По-друге, локалізація «критичних» моментів часу та ознак для подальшої інтерпретації [5; 8].

У роботі формалізація спирається на концепцію «моделі норми» (*normality modeling*): нейромережева генеративна модель навчається відтворювати типові фізіологічні патерни, а міра відхилення між спостереженням і реконструкцією й є аномальністю. Постановка враховує обмеження дослідження (фокус на Attention-VAE, набір даних PhysioNet/Computing in Cardiology Challenge 2019 [35]) і вписана у структуру дисертації.

Вхідні дані та оператор формування ковзних вікон

Нехай для кожного пацієнта доступний багатовимірний часовий ряд спостережень

$$X = \{x(t_k)\}_{k=1}^{T},\ x(t_k) \in \mathbb{R}^d$$

де (t_k) — часові мітки, (d) — кількість фізіологічних ознак (ЧСС, MAP, SpO(_2), температура, частота дихання тощо), а (T) — кількість вимірювань. На практиці (t_k) нерівномірні: різна частота вимірювань, пропуски, артефакти моніторингу. Тому подальша формалізація оперує вікнами на регулярній сітці з кроком (t), після вирівнювання часу й імпутації (деталі у підрозділі 2.3) [6; 11].

Ковзне часове вікно фіксованої довжини (W) визначається як матриця

$$X \in \mathbb{R}^{W \times d}$$

де (W) — кількість дискретних часових кроків у вікні, а елемент (x_{t,j}) відповідає значенню (j)-ї ознаки на (t)-му кроці всередині вікна. Для пакетної обробки використовується множина

$$\{X^{(i)}\}_{i=1}^{N}$$

де (N) — кількість сформованих вікон у навчальній/валідаційній вибірці.

Оператор віконування визначається за формулою (2.1):

$$X^{((i))}=[y(t\_i),y(t\_(i+1)),\cdots,y(t\_(i+W-1))]\in\mathbb{R}(W\times d),\ i=1,\ldots,N. \quad (2.1)$$

У разі використання кроку зсуву (stride) (S) вікна формуються з індексами (i). Вибір (W) та (S) є компромісом. Коротші вікна краще реагують на швидкі зміни; довші акумулюють контекст і зменшують вплив шумів. Малий stride дає детальніший часовий профіль ризику, але підвищує корельованість прикладів і обчислювальні витрати [20].

Вихідний сигнал та типи аномальності

Аномальність у запропонованій постановці має дві взаємодоповнювальні форми:

Віконна (епізодна) оцінка — інтегральна міра для всього часового вікна

$$a(X)$$

що відображає загальний рівень нетиповості/ризику на інтервалі спостереження і використовується як сигнал на рівні епізоду (вікна).

Покрокова оцінка — міра аномальності для конкретного моменту часу всередині вікна

$$[\ a\_t,\ ]$$

яка локалізує критичні фрагменти (сегменти в часі) і працює як основа інтерпретації — зокрема разом із механізмом уваги (наступний підрозділ) [5; 7].

Розмежування двох рівнів практичне. Лікарю потрібне «підсумкове» попередження (ризик високий/низький у найближчі години) і пояснення: які саме інтервали та показники спричинили спрацювання. У роботі основним об’єктом аналізу є віконна оцінка a(X); покрокова інтерпретація трактується як похідна, що дає основу для пояснення результатів.

Для ICU-даних релевантні кілька типів аномалій: *point anomaly* (одиночні викиди), *contextual anomaly* (значення не екстремальне само по собі, але нетипове у певному контексті), *collective anomaly* (порушення як сегмент/послідовність) [10]. Сепсис частіше відповідає *contextual*/*collective* типам. Змінюється структура трендів, варіативність, взаємозв’язки між

показниками, реакція на терапію, а не одне значення «стрибком». Типи аномалій у фізіологічних часових рядах ICU та їх потенційний зв’язок із сепсисом систематизовано в таблиці 2.1.

Таблиця 2.1 – Типи аномалій у фізіологічних часових рядах та їх інтерпретація для оцінки раннього ризику сепсису

Тип аномалії	**Формальна ознака**	**Приклад у часових рядах ICU**	**Потенційний зв’язок із сепсисом**	**Коментар щодо хибних спрацьовувань**
Point	ізольований викид (x_{t,j})	одиничний артефакт SpO(_2) або MAP	частіше неінформативно без контексту	висока ймовірність шуму/датчиків
Contextual	нетиповість відносно стану ()	«нормальна» ЧСС, але нетипова при низькому MAP	можливий ранній дисбаланс регуляції	потребує моделювання взаємозв’язків
Collective	нетиповий сегмент ({x_{t,j}}_{t=t_0}^{t_1})	стійке зростання ЧСС+RR, падіння MAP	відповідає поступовому погіршенню	залежить від (W), (t) і імпутації
Multivariate-structural	зміна коваріацій/кореляцій	«розсинхронізація» показників під час шоку	може відображати системну дисфункцію	потребує адекватної нормалізації та маскування пропусків

Міра аномальності через реконструкцію та/або правдоподібність

Як критерій аномальності використовується реконструкційна похибка моделі, навченої відтворювати нормальні фізіологічні патерни. Нехай () —

реконструкція вхідного вікна (), отримана автоенкодером. Типові форми реконструкційної похибки:

Середня абсолютна похибка (MAE):

$$e(X, \hat{X}) = \Sigma_{t=1}^{W} \Sigma_{j=1}^{d} |x_{t,j} - \hat{x}_{t,j}|$$

стійкіша до поодиноких викидів і корисна для неоднорідних за дисперсією сигналів [7].

Середньоквадратична похибка (MSE):

$$e(X, \hat{X}) = \Sigma_{t=1}^{W} \Sigma_{j=1}^{d} (x_{t,j} - \hat{x}_{t,j})^2$$

яка сильніше штрафує великі відхилення та підвищує чутливість до різких змін.

Зважена похибка:

$$e(X, \hat{X}) = \Sigma_{t=1}^{W} \Sigma_{j=1}^{d} w_j \cdot |x_{t,j} - \hat{x}_{t,j}|$$

де (w_j) відображають відносну клінічну значущість ознак або компенсують різні масштаби після нормалізації. (w_j) задають експертно або підбирають на валідації — на етапі постановки конкретні числа не фіксуються [7].

Для варіаційного автоенкодера додається ще одна природна міра: ймовірнісна правдоподібність (або від'ємна лог-правдоподібність) спостереження за умов латентного коду. Вона краще враховує невизначеність і шум.

Аномальний скор через від'ємну лог-правдоподібність визначається за формулою (2.2):

$$s(X) = -\log p_\theta(X \mid z) \approx -\frac{1}{L}\Sigma_{\ell=1}^{L} \log p_\theta(X \mid z^{(\ell)}),\ z^{(\ell)} \sim q_\phi(z \mid X). \quad (2.2)$$

Цей скор узагальнює реконструкційну похибку. За гаусівської правдоподібності (p_()) він еквівалентний MSE з точністю до констант, але вже дозволяє явно врахувати дисперсію шуму чи гетероскедастичність (різну невизначеність для різних ознак / часових кроків) [7; 3].

Інтегральна оцінка аномальності на рівні вікна задається як

$$a(X) = g(s(X))$$

де (g()) — монотонне перетворення (порогова функція, стандартизація або калібрована ймовірність ризику). Принциповий нюанс: вибір (g()) та порогу становить частину алгоритмічного забезпечення, але міток сепсису для навчання моделі не потребує. Так і реалізується unsupervised/weakly supervised парадигма.

Нормалізація скору та порогове рішення без витоку інформації

Реконструкційну похибку (чи likelihood-скоринг) перетворюють у формалізований показник ризику через пороговий підхід:

$$a(X) = 𝕀[s(X) \geq \tau]$$

де () — порогове значення, визначене за статистикою скорів на «нормальних» даних. Уникнення витоку інформації тут критичне. Поріг не підбирається за мітками сепсису на етапі формування моделі норми. Натомість він оцінюється на навчальній/валідаційній вибірці, де часові відрізки одного пацієнта не змішуються між сплітами, а майбутня інформація не перетікає у минуле [9].

Зручним на практиці варіантом є нормалізований скор (z-score):

$$s\_z(X) = (s(X) - \mu\_s) / \sigma\_s$$

де (_s) та (_s) — середнє та стандартне відхилення, оцінені на даних нормальних режимів. Нормалізація спрощує порівняння моделей і конфігурацій, але обережно: якщо розподіли скорів істотно негаусові, z-score стає недостатньо робастним. Тоді доречні квантильні чи EVT-підходи [25; 10].

Квантильний поріг визначається за формулою (2.3):

$$\tau = Q\_(1-\alpha)(s(X\^((i)))(i=1)\^(N(val))), \quad (2.3)$$

де (Q_{1-}) — емпіричний квантиль рівня (1-) на валідаційній множині «норми», а () задає цільову частоту хибних спрацьовувань на нормі (false alarm rate).

Для інтерпретації на рівні часу корисна покрокова реконструкційна похибка, наприклад

$$e\_t = \Sigma\_(j=1)\hat{}d \ |x\_(t,j) - \hat{x}\_(t,j)|$$

що утворює часовий профіль «нетиповості» всередині вікна. Її зіставляють з вагами уваги (_t) у Attention-механізмі. Так розмежовуються «де модель помиляється» (реконструкція) і «на що модель спирається» (увага); це опора для подальшого пояснення результатів [5; 7].

Покрокова похибка як часовий профіль визначається формулою (2.4):

$$e\_t(X,\hat{X})=(1)/(d)\Sigma\_(j=1)\hat{}dx\_(t,j)-\hat{x}\_(t,j),\ t=1,\ldots,W. \quad (2.4)$$

Відповідність формалізації клінічній задачі раннього виявлення сепсису

Формалізація прив’язана до раннього виявлення сепсису через одне припущення: розвиток сепсису супроводжується **зміною нормальних часових патернів** раніше, ніж стан відповідатиме формальним клінічним критеріям. Інакше кажучи, навіть без офіційного «підтвердження» на момент (t) модель норми може фіксувати накопичення нетипових відхилень у багатовимірній динаміці. Тут важливо розрізняти три речі:

- **Ранній ризик (early warning)**: підвищення (s()) або (s_z()) на послідовності вікон, що вказує на тренд до погіршення [19; 3].
- **Стабільність сигналу**: одиничного перевищення порогу мало. Щоб зменшити хибні тривоги, скори агрегуються за послідовністю вікон. У роботі застосовано експоненційне згладжування з коефіцієнтом $\alpha = 0,3$, що відповідає горизонту ~3 послідовних вікон. Це компроміс між стійкістю до випадкових сплесків і латентністю детекції стабільних трендів [8].
- **Мультимодальність «норми»**: норма в ICU не є єдиним режимом. Різні клінічні стани (післяопераційний, вентиляція, седація) утворюють окремі кластери нормальної динаміки. Звідси перевага генеративних

моделей з латентним простором (VAE): вони моделюють суміш режимів без явних міток [3; 1].

Окремо про дані Challenge 2019: вони містять неоднорідні профілі вимірювань і пропуски. «Аномальність» у сенсі реконструкційної похибки тоді відображає не лише патологію, а й проблеми якості даних (артефакти моніторингу, зміна сенсора, затримка оновлення показника). Звідси потреба у попередній обробці, маскуванні пропусків і контролі якості. Ці процедури формально відокремлені від власне детектора аномалій (детально у підрозділі 2.3) [6]. Зведення дизайнерських рішень формалізації задачі та їхніх наслідків для чутливості і стійкості сигналу аномальності наведено в таблиці 2.2.

Таблиця 2.2 – Дизайнерські рішення формалізації та їх наслідки для чутливості/стійкості сигналу аномальності

Рішення	**Варіанти**	**Переваги**	**Ризики/обмеження**	**Де фіксується в роботі**
Міра скору (s())	MAE / MSE / -loglik	простота; імовірнісна інтерпретація для VAE	чутливість до шуму/масштабів; залежність від нормалізації	2.1 (визначення), 2.2 (узгодження з VAE)
Рівень рішення	віконний (a()) / покроковий (a_t)	підтримка алармів і пояснень	потреба у правилах агрегації	2.1 (постановка), 3.x (протокол)
Поріг ()	z-score / квантиль / EVT	контроль FAR на нормі; переносимість	дрейф розподілу скорів; калібрування	2.1 (принцип), 4.x (моніторинг)
Формування вікон	(W), (S), (t)	контроль «часової роздільності» сигналу	leakage при неправильному спліті; кореляція прикладів	2.3 (деталі пайплайну)

У підрозділі формалізовано задачу раннього виявлення сепсису як пошук аномалій у багатовимірних часових рядах через моделювання «норми» та оцінку відхилення реконструкційним або імовірнісним скором. Визначено представлення даних у вигляді ковзних вікон, введено віконний і покроковий рівні аномальності, окреслено принципи нормалізації та вибору порогу без міток сепсису для уникнення витоку інформації. Це відповідає *unsupervised* / *weakly supervised* постановці.

2.2 Розробка архітектури варіаційного автоенкодера з механізмом Attention

Архітектура моделі для раннього виявлення сепсису упирається у специфіку ICU-даних: висока зашумленість, пропуски, нерівномірність вимірювань (до ресемплінгу), індивідуальні відмінності пацієнтів і нестаціонарність процесів через лікування та зміну клінічного стану. До цього додається ще одне: «золоті» мітки сепсису ненадійні або приходять із затримкою, через що класичне супервізоване навчання тут утруднене. Звідси вибір: генеративне моделювання «норми». Модель навчається відтворювати типові патерни, а відхилення від відтворення (чи від очікуваної правдоподібності) трактується як аномальність, потенційно пов’язана з ранніми проявами сепсису [3; 1].

Варіаційний автоенкодер (VAE) як генеративна модель є природним вибором. Він поєднує імовірнісну постановку з явним латентним простором, стабільне градієнтне навчання через репараметризацію, узагальнення через KL-регуляризацію та гнучкість у виборі сімейства умовного розподілу. Для фізіологічних часових рядів латентний простір стискає багатовимірну динаміку до компактного коду, тобто до режиму функціонування організму на

проміжку часу. Реконструкція тоді відображає очікуваний перебіг показників за належності до вивчених «нормальних» режимів [1; 3].

Механізм часової уваги (*temporal attention*) додається до VAE з двох причин. По-перше, не всі моменти часу у вікні однаково інформативні. Ранні патологічні зміни часто проявляються короткими сегментами нетипової динаміки або специфічними фазами реакції на лікування; ці моменти важливо «підсвітити» під час формування латентного коду. По-друге, клінічне впровадження потребує інтерпретованості: окрім інтегрального скору аномальності, потрібен пояснювальний сигнал, які часові фрагменти найбільше вплинули на представлення і на потенційне спрацювання [5; 8].

Імовірнісна постановка VAE та параметризація латентного простору

Нехай вхідне часове вікно позначено (^{Wd}). VAE задає апріорний розподіл латентних змінних (^{m}) як стандартну багатовимірну норму (p()=(,)). Апостеріорний розподіл (p()) апроксимується варіаційною сім'єю (q_()), параметризованою енкодером із параметрами (). На практиці часто використовується діагональна гаусівська форма:

[q_()=(*(),(^2*())),]

де (*()) та (*()) є виходами нейромережі, що кодує часову динаміку у вектор параметрів розподілу.

Вибір механізму енкодера для часових даних становить центральне рішення. Для моделювання залежностей на проміжку (W) кроків підходять модулі, що враховують порядок і контекст: рекурентні мережі (GRU/LSTM), часові згорткові мережі (TCN) чи трансформерні блоки. У роботі базовим є рекурентний енкодер. Причини: він природно оперує послідовностями, дає компактні приховані стани (_t) і легко поєднується з multi-head attention [6; 43].

Параметри латентного простору є компромісом між виразністю і стабільністю. Розмірність (m) не повинна бути надмірно великою; інакше відбувається перенавчання і «розмивання» нормалізуючого ефекту KL-регуляризації. Але й не надто малою, інакше втрачається інформація про різні

режими норми. У роботі обрано m = 16 (експериментальний підбір, деталі у розділі 4).

Щоб зробити навчання стохастичної моделі сумісним зі стандартним backpropagation, застосовується репараметризаційний трюк.

Репараметризація здійснюється за формулою (2.5):

$$z=\mu_\varphi(X)+\sigma_\varphi(X)\odot\varepsilon,\ \varepsilon \sim N(0,I). \quad (2.5)$$

Стохастичність переноситься в незалежну змінну (); саме це робить () та () тренованими градієнтним спуском [1].

Функція втрат (ELBO), β-регуляризація та стабільність навчання

VAE навчається через максимізацію нижньої оцінки лог-правдоподібності (*Evidence Lower Bound*, ELBO) як суми терму реконструкції та KL-дивергенції між апроксимацією апостеріорного розподілу і апріорі. У ()-VAE додається коефіцієнт () для керування силою регуляризації латентного простору.

Цільова функція *β*-VAE має вигляд за формулою (2.6):

$$\mathcal{L}(\theta,\varphi)= E_{q_\varphi(z|X)}\log p_\theta(X|z) -\beta\cdot D_{KL}(q_\varphi(z|X);|;p(z). \quad (2.6)$$

Кожен з двох доданків має конкретний практичний вплив на anomaly detection. Терм реконструкції відповідає за здатність моделі відтворювати нормальні патерни. Слабка реконструкція = «шумний» скор аномальності. KL-регуляризація запобігає надмірному підлаштуванню і сприяє узагальненню на нових пацієнтах. Але є зворотний бік: надмірна сила KL призводить до *posterior collapse*: латентний код стає малоінформативним, і модель ігнорує () [2]. Для пом'якшення на практиці застосовують KL-annealing (поступове збільшення ваги KL-терму), free-bits або обмеження мінімального внеску KL на вимір латентного простору [2].

Для медичних даних () є інструментом керування компромісом. Більше значення підвищує узагальнювальні властивості й «згладжує» латентний

простір, але занадто груба реконструкція знизить чутливість детектора. У роботі $\beta_{max} = 0{,}5$ з лінійним розігрівом 30 епох (детально у розділі 4).

Механізм часової уваги в енкодері та формування контексту

Не всі позиції у вікні однаково важливі. Над послідовністю прихованих станів енкодера $h_t_{t=1}^{W}$ застосовується багатоголова scaled dot-product увага (multi-head attention) [6] (4 голови, розмірність ключа d_k). Зважений контекстний вектор c формується за формулою (2.7):

$$\alpha_t = softmax\left(q^{\top} k_t/\sqrt{d_k}\right),\ c = \Sigma_{t=1}^{W} \alpha_t h_t \quad (2.7)$$

де q — вектор запиту, k_t — ключ для кроку t, d_k — розмірність ключа.

Нормалізація softmax гарантує $\Sigma\, \alpha_t = 1$ та $\alpha_t \geq 0$,

що відповідає підходу, застосованому у MA-VAE [39] для виявлення аномалій

у багатовимірних часових рядах.

Декодер, сімейство правдоподібності та реконструкційний терм

Декодер параметризує умовний розподіл ($p_{()}$) і формує реконструкцію (^{Wd}). Для часових даних можливі дві принципові схеми. **Послідовнісний декодер** генерує крок за кроком (RNN/TCN/Transformer). **Паралельний** відновлює весь тензор ((Wd)) одночасно через згорткові/MLP-блоки з позиційним кодуванням. Вибір залежить від наявності потреби в явній авторегресивності. Для реконструкції фіксованих вікон часто достатньо умовної генерації всього вікна без авторегресивного зв'язку: менше обчислень, простіше навчання [3; 16].

Поширене припущення: гаусівська правдоподібність із фіксованою або навчуваною дисперсією. За сталої (^{2}) максимізація ($p_{()}$) еквівалентна мінімізації MSE з точністю до константи. Якщо () залежить від ознаки чи часу (гетероскедастичність), модель відображає різну надійність вимірювань і шумів для різних каналів, що особливо доречно для ICU-показників [17].

Врахування пропусків становить критичний аспект архітектури. Якщо вхідні дані містять пропуски (після імпутації чи з маскою пропусків), реконструкційний терм потрібно рахувати з огляду на маску спостережуваності. Інакше модель «вчиться» на штучно заповнених значеннях як на істинних. Можливі дві реалізації: маскована похибка (до втрат включаються лише дійсно спостережені (x_{t,j})) або окрема голова для реконструкції та для моделювання невизначеності [6; 11].

Скоринг аномальності та узгодження з увагою

Після навчання моделі норми формується аномальний скор (s()) за реконструкційною похибкою чи від'ємною лог-правдоподібністю. Архітектурна вимога: скоринг має бути стабільним у часі і придатним до подальшого калібрування порогу. Оскільки VAE є імовірнісною моделлю, природний кандидат на скор: (s()=-p_()) (або його апроксимація через семплювання з (q_())). Конкретна форма (s()) (MAE/MSE/-loglik/зважені версії) є параметром системи, підбирається експериментально [7; 3].

Окремо про взаємодію скорингу й уваги. Ваги (_t) відображають важливість часових позицій для латентного коду; реконструкційна похибка відображає «невідповідність» вхідного вікна моделі норми. Разом ці два сигнали дають інформативніше пояснення: сегменти з високою (_t) і підвищеною локальною похибкою стають кандидатами на «критичні» фрагменти, що потребують клінічного аналізу. Формальне правило інтеграції визначене так: сегменти з $\alpha_t > \mu_\alpha + \sigma_\alpha$ та локальною похибкою e_t > медіани розглядаються як критичні [5; 7].

Маскована реконструкційна втрата визначається формулою (2.8):

$$\mathscr{L}_{(rec)} = (1)/(\Sigma_{(t,j)} m_{(t,j)}) \Sigma_{(t=1)}^{W} \Sigma_{(j=1)}^{d} m_{(t,j)} \cdot \rho x_{(t,j)} - \hat{x}(t,j), \quad m(t,j) \in 0,1, \tag{2.8}$$

де (m_{t,j}) — маска наявності спостереження, а (()) — функція похибки (наприклад, (||) або (()^2)).

Формула має подвійний ефект. По-перше, зменшує ризик навчання на артефактах імпутації. По-друге, робить реконструкційний скор менш залежним від патернів пропусків і ближчим до власне фізіологічної динаміки [6; 11].

Архітектурний зріз Attention-VAE

Компоненти Attention-VAE та їх функції у задачі моделювання «норми» надані в табл. 2.3.

Дизайнерські рішення для Attention-VAE у клінічних часових рядах та очікуваний вплив (без чисел) надані в табл. 2.4.

Таблиця 2.3 – Компоненти Attention-VAE та їх функції у задачі моделювання «норми»

Компонент	**Вхід/ вихід**	**Функція**	**Типові варіанти реалізації**	**Потенційні ризики**
Енкодер часових рядів	()	витяг часових ознак і контексту	GRU/LSTM, TCN, Transformer	перенавчання; чутливість до ресемплінгу
Часова увага	({_t})	зважена агрегація ключових сегментів	multi-head attention [6], scaled dot-product	attention не завжди каузальна
Голови параметрів (q_)	(,^2)	параметризація латентного розподілу	MLP, нормалізація, dropout	posterior collapse при невдалих налаштуваннях
Репараметризація	((,))	стохастичність, сумісна з backprop	стандартна схема	нестабільність при великих ()
Декодер	()	реконструкція норми	RNN/TCN/MLP-декодер	слабка реконструкція → шумний скор
Втрата (ELBO)	(,,q_,p)	навчання генеративної моделі	()-VAE, KL annealing	компроміс між узагальненням і чутливістю
Скоринг аномальності	(,s())	сигнал ризику	MAE/MSE/-loglik, маскування	чутливість до дрейфу й якості даних

Таблиця 2.4 – Дизайнерські рішення для Attention-VAE у клінічних часових рядах та очікуваний вплив (без чисел)

Рішення	**Варіант**	**Очікуваний ефект**	**Потенційне обмеження**	**Що потребує валідації**
() у ()-VAE	менше / більше	чутливість vs узагальнення	posterior collapse або погана реконструкція	стабільність ELBO, якість скорів
Тип декодера	послідовнісний / паралельний	точність реконструкції vs швидкість	складність навчання авторегресії	порівняння на протоколі 3 розділу
Увага	multi-head attention [6], scaled dot-product	інтерпретованість, фокус на сегментах	складність; неоднозначність пояснень	кореляція з клінічно значущими подіями
Робота з пропусками	імпутація / маска / jointly	зниження артефактів у скорі	залежність від якості маски	чутливість до missingness patterns
Сімейство likelihood	фікс. () / гетероскедаст.	робастність до шумів	ускладнення оптимізації	стабільність і калібрування

У підрозділі описано архітектуру Attention-VAE як імовірнісної генеративної моделі для моделювання «норми» у багатовимірних часових вікнах. Задано латентну змінну з апріорним розподілом, визначено варіаційну апроксимацію апостеріорного розподілу через енкодер, наведено навчання через ELBO у постановці ()-VAE, деталізовано механізм часової уваги для агрегування прихованих станів і підвищення інтерпретованості. Окремо проаналізовано роль вибору декодера, сімейства правдоподібності та маскованого реконструкційного терму для коректної роботи з пропусками і шумами ICU-даних. Сукупність цих рішень визначає, як модель трансформує

сигнал реконструкційної похибки у клінічно придатний індикатор аномальності.

2.3 Алгоритм попередньої обробки даних та формування навчальної вибірки

Попередня обробка клінічних часових рядів становить критичний етап для виявлення аномалій. Якість вхідних представлень напряму визначає стабільність навчання «моделі норми» і надійність аномального скорингу. Для ICU це гостро вдвічі: фізіологічні показники мають нерівномірну частоту вимірювань, значні обсяги пропусків, артефакти моніторингу, а також зміни режимів (вентиляція, седація, інфузійна терапія); звідси нестаціонарність і кілька «нормальних» підрежимів. Некоректний препроцесинг тут породжує псевдоаномалії (наприклад, через агресивну імпутацію) або, навпаки, згладжує ранні патологічні тенденції, і чутливість системи падає [6].

У роботі застосовано unsupervised / weakly supervised підхід. Звідси два жорстких правила для препроцесингу і формування вибірок: жодного витоку інформації; клінічні мітки сепсису не використовуються як сигнал навчання. Мітки потрібні лише на двох етапах *постфактум*: оцінка якості детекції на тестовій вибірці і калібрування порога τ через оптимізацію F1 на валідаційній. Weakly supervised процедура навчання у роботі не застосовується [18; 21].

Алгоритм попередньої обробки у роботі становить детерміновану послідовність кроків, яка перетворює «сирі» нерівномірні вимірювання з PhysioNet Challenge 2019 у стандартизовані часові вікна ($^{\{(i)\}}\{\mathrm{Wd}\}$) для навчання Attention-VAE. Нижче наведено повний пайплайн. Чотири акценти: відтворюваність, контроль артефактів і пропусків, коректне розбиття без *leakage*, узгодженість зі скорингом і калібруванням порога.

Пайплайн попередньої обробки (узгоджена послідовність кроків):

- Відбір ознак (d) та уніфікація каналів вимірювань.
- Вирівнювання часу / ресемплінг до регулярного кроку (t).
- Обробка викидів (outlier handling) та перевірка фізіологічних меж.
- Імпутація пропусків (з урахуванням клінічної природи сигналів).
- Нормалізація (обчислення статистик лише на train-частині).
- Формування ковзних вікон (W) та кроку (S) (stride), побудова маски спостережуваності.
- Розбиття даних (split) без витоку: patient-level split, контроль часової причинності.

1) Відбір релевантних фізіологічних показників та специфікація ознак

На першому етапі визначається множина фізіологічних ознак () розмірності (d=||) за трьома критеріями: (i) доступність у наборі PhysioNet Challenge 2019 [35] для пацієнтів ВІТ; (ii) відображення кардіореспіраторних і терморегуляторних процесів; (iii) клінічна інтерпретованість при системному запаленні та сепсис-асоційованій дисфункції органів. У роботі обрано п’ять показників: ЧСС (HR), насичення крові киснем (SpO_2), середній артеріальний тиск (MAP), температура тіла та частота дихання (RR) [35].

Уніфікація каналів вимірювань полягає у розв’язанні типових неоднозначностей. Різні джерела одного показника (інвазивне/неінвазивне вимірювання тиску), різні одиниці чи форматування температури (°C vs °F), дублікати записів, кілька одночасних значень одного каналу. Для таких ситуацій задається правило пріоритезації джерела або агрегації (медіана/середнє за короткий інтервал); інакше у ряді з’являються штучні «стрибки» [22]. Стандартизований опис кроків препроцесингу разом із ключовими артефактами, які усуваються на кожному етапі, подано в таблиці 2.5.

Таблиця 2.5 – Стандартизований опис кроків препроцесингу та ключові артефакти, які усуваються

Крок	**Вхід → вихід**	**Мета**	**Типові артефакти/ризики**	**Контроль відтворюваності**
Відбір ознак (d)	сирі канали → ()	уніфікація та інтерпретованість	різні одиниці/джерела	фіксований список (), версіонування
Ресемплінг (t)	нерівномірні (t_k) → регулярна сітка	єдина часовість	aliasing, згладження	фіксоване (t), детермінований алгоритм
Outlier handling	значення → очищені значення	зменшення шуму/артефактів	псевдоаномалії	правила меж/вінзоризації, лог змін
Імпутація	пропуски → заповнені/масковані	формування повних вікон	«вигадування» патернів	вибір методу + маска ()
Нормалізація	сирі масштаби → стандартизовані	порівнюваність каналів	leakage через test-статистики	(,) лише на train
Віконування (W,S)	ряд → (^{(i)})	навчальні приклади	висока корельованість	фіксовані (W,S), seed/порядок
Split без leakage	пацієнти/епізоди → train/val/test	чесна оцінка	перетікання пацієнта між сплітами	patient-level split, перевірки

2) Вирівнювання часу та ресемплінг до регулярної сітки

Оскільки вимірювання в ICU є нерівномірними, вводиться регулярна часова сітка з кроком (t). Для кожного пацієнта визначається інтервал спостереження ([t_{},t_{}]), який дискретизується як ($n=t\{\}+nt$), (n=0,,N_t).

Далі для кожного каналу здійснюється відображення нерівномірних вимірювань (x(t_k)) на регулярну сітку ((_n)) за правилом ресемплінгу.

Практично доступні два базові механізми з різними властивостями:

- **Агрегація в бін (binning)**: для кожного (_n) збираються всі вимірювання в інтервалі ([*n,*{n+1})) і агрегуються (середнє/медіана/останнє значення). Це знижує чутливість до «сплесків» частоти вимірювань, але може втрачати короткі піки.
- **Інтерполяція до сітки**: якщо вимірювання рідкі, застосовується лінійна інтерполяція між найближчими точками; підхід гладко відновлює тренд, але може «домальовувати» неіснуючі коливання на довгих прогалинах.

Вибір (t) є компромісом між (i) часовою роздільністю ранніх змін і (ii) часткою пропусків після дискретизації. У роботі зафіксовано $\Delta t = 5$ хвилин [35], що дає 72 кроки на 6-годинне вікно спостереження [20; 6]. Для коротких прогалин у вікні застосовується forward-fill (перенесення останнього відомого значення), потім z-score нормалізація [8].

Регулярна сітка часу визначається за формулою (2.9):

$$\tilde{t}n=t(min)+n\Delta\ t,\ n=0,\ldots,N\_t,\ N\_t=\llcorner(t\_(max)-t\_(min))/(\Delta\ t)\lrcorner. \qquad (2.9)$$

Лінійна інтерполяція на регулярній сітці здійснюється за формулою (2.10):

$$\tilde{x}(\tilde{t}) = x(t\_k)+(x(t\_(k+1))-x(t\_k))/(t\_(k+1)-t\_k),(\tilde{t}-t\_k),\ t\_k\leq\tilde{t}\leq t\_(k+1). \qquad (2.10)$$

Щоб обмежити шкідливі ефекти інтерполяції на довгих прогалинах, вводиться максимальна допустима довжина інтервалу без спостережень (G_{}). Якщо ((t_{k+1}-t_k)>G_{}), інтерполяція забороняється: значення позначається як пропуск і обробляється на етапі імпутації/маскування. У роботі прийнято $G\_(max) = 2$ години (24 кроки при $\Delta t = 5$ хвилин) [11; 6].

3) Очищення даних та обробка викидів (outlier handling)

Артефакти моніторингу (помилки сенсорів, некоректні записи, одиничні «сплески») призводять до надмірних реконструкційних похибок, а відповідно і до хибних спрацювань детектора. Тому перед імпутацією та нормалізацією виконується очищення:

- **Перевірка фізіологічних меж** для кожної ознаки з вінзоризацією значень поза межами. Межі задаються на основі клінічних довідників та усталених практик аналізу ICU-даних: для ЧСС — [30; 220] уд./хв, MAP — [30; 200] мм рт. ст., SpO(_2) — [50; 100] %, температури — [32; 42] °C, частоти дихання — [4; 60] подих./хв. Значення поза цими діапазонами розглядаються як артефакти моніторингу [10; 22].
- **Виявлення локальних викидів** за робастними критеріями (IQR/MAD або пороги на похідній), щоб відсікти короткі імпульсні артефакти без знищення клінічно значущих різких змін.
- **Усунення дубльованих записів** та конфліктів джерел (якщо одночасно присутні кілька значень з різних приладів).

Принциповий нюанс: для anomaly detection обробка викидів має бути консервативною. Агресивне «зачищення» прибирає реальні ранні ознаки деградації стану і тим самим знижує сигнал, який модель має вловити. Тому правила очищення формулюються так: гарантовано прибирати фізіологічно неможливі або явно технічні артефакти; сумнівні випадки залишати моделі як потенційно нетипові, але правдоподібні зміни [10; 6]. Огляд типових стратегій обробки викидів у вітальних рядах і пов’язані з ними ризики для виявлення аномалій узагальнено в таблиці 2.6.

Таблиця 2.6 – Приклади стратегій обробки викидів у вітальних рядах та ризики для anomaly detection

Стратегія	**Опис**	**Переваги**	**Ризики**	**Коли застосовується**
Жорсткі фізіологічні межі	відкидання значень поза ([L_j,U_j])	прибирає неможливі артефакти	потребує коректних меж	базове очищення для всіх ознак
Вінзоризація	обрізання до ([L_j,U_j])	зменшує вплив крайніх значень	може «згладити» істотні піки	коли викиди поодинокі
Робастні локальні пороги	MAD/IQR на короткому вікні	відсікає імпульсні шуми	може відсікти різкі клінічні зміни	лише для очевидних артефактів
Фільтрація похідної	обмеження (	x	) за крок	контроль «нереальних» стрибків

4) Імпутація пропусків та маска спостережуваності

Після ресемплінгу значна частка (*{t,j}) залишається пропущеною. Імпутація потрібна, щоб подати дані в нейромережеву модель. Але некоректна імпутація створює штучні «нормальні» патерни, а це знижує сигнал аномальності. Тому у роботі застосовано поєднання двох підходів: (і) помірна імпутація для коротких прогалин; (іі) явне врахування пропусків через маску спостережуваності (^{Wd}). У ній (m*{t,j}=1) означає наявність первинного (неімпутованого) або надійного значення, а (m_{t,j}=0) — пропуск/ненадійність.

Для імпутації застосовуються методи, що не потребують міток і є відтворюваними:

- **LOCF (перенесення останнього спостереження)**: добре працює для повільно змінних показників на коротких інтервалах, але може «заморожувати» тренди на довгих прогалинах.
- **Лінійна інтерполяція**: доцільна для коротких прогалин між двома надійними вимірюваннями; небажана при довгих інтервалах або різких переходах режимів.
- **Комбінована стратегія**: інтерполяція для прогалин до (G_{}), LOCF до (G_{}), інакше залишення пропуску з маскою. Границі визначаються емпірично G_interp = 30 хвилин (6 кроків), G_locf = 2 години (24 кроки).

Додатково вікна з часткою пропущених значень понад 30% виключаються з навчальної вибірки як ненадійні, що відповідає практиці обробки клінічних даних BIT [8].

Усі імпутовані значення супроводжуються маскою (). Її роль двояка. По-перше, маска використовується при обчисленні маскованої реконструкційної похибки. По-друге, її подають у модель як додатковий канал, щоб Attention-VAE враховував патерни пропусків як контекст, а не перетворював пропуски на псевдосигнал [6; 11]. Порівняння методів імпутації для ICU-вітальних рядів у контексті генеративного моделювання норми наведено в таблиці 2.7

.

Таблиця 2.7 – Порівняння методів імпутації для ICU-вітальних рядів при генеративному моделюванні норми

Метод	**Переваги**	**Недоліки**	**Коли коректний**	**Додаткові запобіжники**
LOCF	простий, стабільний	«заморожує» значення	короткі прогалини, повільні сигнали	ліміт тривалості, маска

Кінець табл. 2.7

Метод	**Переваги**	**Недоліки**	**Коли коректний**	**Додаткові запобіжники**
Лінійна інтерполяція	зберігає тренд	«домальовує» між точками	короткі прогалини між валідними точками	заборона на довгі прогалини
Комбінований	гнучкий компроміс	більше гіперпараметрів	змішані режими вимірювань	протокол фіксації (G_{*})
Залишення пропусків + маска	не вигадує значення	потрібна підтримка в моделі	довгі прогалини	маскована втрата / додатковий канал

5) Нормалізація та контроль витоку інформації

Для порівнюваності ознак і стабільного навчання VAE здійснюється нормалізація. Найпоширенішим варіантом є Z-нормалізація по кожному каналу (j). Статистики обчислюються **лише на навчальній вибірці** (train), і саме це запобігає leakage.

Z-нормалізація для кожного каналу здійснюється за формулою (2.11):

$$x'_{(t,j)}=(x_{(t,j)}-\mu_j)/(\sigma_j),\ \mu_j=(1)/(\left|D_{(train)}\right|)\Sigma_{(t,j)\in D_{(train)}} x_{(t,j)},\quad \sigma_j=sqrt(1)/(\left|D_{(train)}\right|)\Sigma_{(t,j)\in D_{(train)}}(x_{(t,j)}-\mu_j)^2. \qquad (2.11)$$

Під ($\{train\}$) *розуміється сукупність доступних (після очищення) значень у train-частині. Якщо використовується маска пропусків, статистики обчислюються лише за* ($m\{t,j\}=1$). Для робастності до важких хвостів розподілу — три альтернативи: робастна нормалізація (медіана та MAD), квантильне перетворення, нормалізація по пацієнту (patient-wise). Остання зменшує міжпацієнтну варіативність, але водночас пригнічує важливі

абсолютні рівні показників. Тому вибір схеми нормалізації — гіперпараметр і підлягає оцінюванню у розділі 3 [9; 6].

6) Формування ковзних часових вікон та навчальних прикладів

Після ресемплінгу, очищення, імпутації та нормалізації часовий ряд пацієнта подається як матриця (^{Td}) на регулярній сітці. Навчальні приклади формуються методом ковзних вікон довжини (W) зі stride (S). Для кожного стартового індексу (i) визначається:

Формування вікон зі stride визначається за формулою (2.12):

$$X^{((i))=Y[i:i+W-1,:]\in\mathbb{R}(W\times d),\ i\in 1,1+S,1+2S,\cdots.} \quad (2.12)$$

Паралельно формується маска ($^{\{(i)\}}${Wd}) — вона відображає спостережуваність/імпутацію у вікні. Далі (^{(i)}) використовується для маскованого обчислення реконструкційної похибки та контролю впливу пропусків на латентне представлення. Якщо архітектура Attention-VAE підтримує подання маски як додаткового каналу, вхід розширюється до ([$^{\{(i)\}}$,{(i)}]). У роботі маска використовується для зваженого обчислення реконструкційної похибки [6; 11].

Вибір довжини вікна (W) прив’язаний до клінічної інтерпретації «контексту». Короткі вікна відображають швидкі зміни. Довші — накопичують інформацію про тренди, варіабельність і взаємозв’язки між ознаками. У роботі основний розмір — W = 6 годин (72 кроки при $\Delta t = 5$ хвилин); це клінічно значущий горизонт попередження [35; 20].

Щоб зменшити корельованість прикладів, stride (S) не має бути занадто малим. Але занадто великий (S) псує часову роздільність аномального сигналу. У роботі зафіксовано $S = 1$ година (12 кроків): достатнє перекриття вікон без надмірного дублювання [20].

7) Формування train/val/test без витоку інформації (leakage)

Коректність експериментів з часовими рядами вимагає уникнення двох видів leakage:

- **Пацієнтський leakage**: дані одного пацієнта (або одного ICU-епізоду) не повинні потрапляти одночасно в train і test/val, інакше модель може «пам’ятати» індивідуальні патерни та завищувати якість.
- **Часовий leakage**: у задачах раннього попередження неприпустимо, щоб при формуванні навчальних прикладів для певного моменту використовувалася інформація з майбутнього (наприклад, через нормалізацію на всьому наборі даних або через інтерполяцію, що використовує точки після моменту оцінки у режимі, який мав би бути онлайн).

Тому основне правило: **patient-level split**. Множина пацієнтів розбивається на (*{train},* {val}, *{test}), і всі вікна, сформовані з епізодів пацієнтів з (*{train}), входять лише до train (аналогічно для val/test). Для PhysioNet Challenge 2019 — patient-level split 70/15/15 [35].

Навчання безнаглядне, тому в train-вибірці мають домінувати нормальні режими. На практиці — два варіанти. Або з train виключаються фрагменти з явно критичними станами через прості евристики *без міток сепсису* (фільтрація грубо некоректних або екстремально нестабільних траєкторій). Або допускається домішка патологічних станів — якщо вона мала відносно загального обсягу і не руйнує «модель норми». Така домішка — типове явище для unsupervised anomaly detection. У роботі вона розглядається як фактор ризику і враховується при інтерпретації результатів [7; 16].

Запропонований алгоритм попередньої обробки формалізує перетворення сирих нерівномірних вимірювань PhysioNet Challenge 2019 у стандартизовані ковзні часові вікна для навчання Attention-VAE у режимі *unsupervised* / *weakly supervised*. Пайплайн включає: відбір ознак; ресемплінг до регулярної сітки; очищення та обробку викидів; імпутацію з явним урахуванням пропусків через маску спостережуваності; нормалізацію на основі статистик train-частини; формування вікон (W) зі stride (S); коректний patient-level split без витоку інформації. Це дає відтворюваність експериментів

і знижує ризик псевдоаномалій, пов’язаних з артефактами вимірювань чи некоректним препроцесингом.

Висновки до розділу 2

У другому розділі розроблено повне математичне та алгоритмічне забезпечення системи раннього виявлення сепсису. Задачу сформульовано як виявлення аномалій у багатовимірних часових рядах через моделювання розподілу нормальних фізіологічних патернів: показник аномальності є реконструкційною похибкою або від’ємною лог-правдоподібністю, а порогове рішення приймається без міток сепсису на етапі навчання. Архітектуру Attention-VAE специфіковано повністю. GRU-енкодер формує послідовність прихованих станів. Механізм багатоголової уваги (multi-head attention) агрегує їх у контекстний вектор для параметризації варіаційного латентного простору. Декодер реконструює часове вікно. Маскована функція втрат β-VAE з KL-відпалюванням ($\beta_{max} = 0{,}5$, $T_{warmup} = 30$) забезпечує стабільне навчання на нерівномірних клінічних даних. Семикроковий алгоритм препроцесингу з patient-level split (70/15/15) виключає витік інформації між вибірками і гарантує відтворюваність результатів. Сформований теоретичний апарат є основою для програмної реалізації та експериментального дослідження, описаних у наступному розділі.

Розділ 3
Програмна реалізація та експериментальне дослідження

У **Розділі 3** описано програмну реалізацію системи: обґрунтовано вибір засобів розробки (Python, PyTorch), наведено модульну архітектуру пайплайна, протокол навчання на нормальних даних та методологію порівняльного аналізу.

Розділ присвячено практичній реалізації інтелектуальної системи раннього виявлення сепсису на основі аналізу багатовимірних часових рядів життєвих показників пацієнтів. Наведено обґрунтування вибору програмних засобів, опис архітектури програмної системи, процес навчання моделі Attention-VAE, а також підхід до експериментального оцінювання ефективності запропонованого методу та його інтерпретованості. Отримані результати формують основу для аналізу клінічної значущості запропонованих рішень.

3.1 Обґрунтування вибору засобів розробки та опис програмної архітектури

Практична реалізація інтелектуальної системи раннього виявлення сепсису потребує поєднання двох взаємопов’язаних складових: коректної інженерної організації конвеєра обробки медичних часових рядів і гнучкого середовища для дослідження й модифікації нейромережевих архітектур класу Attention-VAE. Визначальні аспекти: вибір мов та бібліотек, опис модульної архітектури, узагальнена схема потоків даних у системі.

Додаткова вимога: відтворюваність експериментів і трасованість усіх артефактів дослідження — від версій даних і параметрів препроцесингу до

конфігурацій навчання, контрольних станів моделей, метрик оцінювання та графічних матеріалів.

Як основну мову програмування обрано **Python** — де-факто стандарт для наукових обчислень, прототипування та розробки систем машинного навчання. Він забезпечує доступ до зрілих бібліотек для обробки даних, оптимізації моделей, візуалізації та інтеграції інструментів експериментального трекінгу. [36] Вибір Python також спрощує реалізацію уніфікованих інтерфейсів модулів пайплайна (читання даних, трансформації, тренування, оцінювання): ці компоненти описуються у вигляді узгоджених функцій та класів із чітко визначеними контрактами входів і виходів.

Для нейромережевої частини системи застосовано фреймворк **PyTorch**, який підтримує динамічні обчислювальні графи, реалізацію кастомних шарів і механізмів уваги, а також зручні засоби керування тренувальним циклом і обчислювальними прискорювачами (GPU за наявності). Для дослідження це особливо важливо: архітектура Attention-VAE може модифікуватися (варіанти енкодера/декодера, види уваги, регуляризації, композиція функцій втрат) без істотних накладних витрат на перебудову коду. [36] Бібліотечні засоби PyTorch для роботи з тензорами природно відповідають представленню багатовимірних часових вікон як тензорів форми (B, W, d), де B — розмір батчу, W — довжина вікна, d — кількість ознак (життєвих показників).

Для підготовки даних і типових перетворень використовуються **NumPy** і **pandas**, які забезпечують ефективні операції над масивами та табличними структурами, включно з ресемплінгом, агрегацією, вирівнюванням часових міток, інтерполяцією та обробкою пропусків. Для задач медичних часових рядів ключова практика: приведення неоднорідних за частотою вимірювань до узгодженої часової сітки перед формуванням навчальних вікон (з явним контролем кроку дискретизації та правил імпутації). [6; 11] Для статистичного аналізу та окремих чисельних процедур (наприклад, розрахунок робастних статистик, перевірка припущень, згладжування) можуть залучатися засоби **SciPy**. [28]

Візуалізація результатів (сирі сигнали, нормалізовані ряди, реконструкції $\hat{X}$, розподіли показника аномальності, а також карти уваги) виконується переважно за допомогою **Matplotlib** та **Seaborn**. Такий вибір забезпечує контроль над оформленням графіків і можливість формувати матеріали, придатні для вставки у дисертаційний текст (сталі осі й масштаби, підписи, легенди, стандартизовані стилі). [37] Числові підсумки на графіках (AUC, F1 тощо) вносяться після проведення експериментів.

Окремо про організацію **експериментального трекінгу**. Щоб накопичувати історію запусків, порівнювати конфігурації, зберігати артефакти і забезпечити відтворюваність, варто використати спеціалізований інструмент (наприклад, MLflow або аналог) з журналюванням параметрів, метрик та артефактів у стандартизованому вигляді. У роботі інструмент трекінгу розглядається як частина програмної архітектури: він логічно поєднує модулі препроцесингу, навчання та оцінювання в єдиний відтворюваний контур. [22] Щоб уникнути прихованих змін середовища, також варто логувати версії бібліотек, параметри апаратної платформи (CPU/GPU), хеші та ідентифікатори конфігураційних файлів. [36; 22]

Схема потоків даних у системі має узагальнений вигляд: **Завантаження даних → Препроцесинг → Формування часових вікон → Attention-VAE → обчислення anomaly score → порогова обробка / калібрування → розрахунок метрик → візуалізація та формування звіту.** На рівні програмної архітектури цей ланцюг реалізується як набір модулів із чітко розмежованими зонами відповідальності:

- **Модуль завантаження даних**: інкапсулює логіку читання сирих даних PhysioNet/Computing in Cardiology Challenge 2019 [35], первинну уніфікацію форматів і формування табличного представлення з часовими мітками та значеннями показників. [9]
- **Модуль попередньої обробки**: відповідає за очищення, узгодження часових шкал, ресемплінг до єдиного Δt, імпутацію пропусків та нормалізацію. У цьому модулі важливо забезпечити прозорість правил

перетворень, оскільки саме вони визначають статистичні властивості входу моделі. [9]

- **Модуль формування часових вікон**: перетворює нормалізовані ряди на тензорні вікна фіксованої довжини *W* зі кроком *S* (stride), формуючи датасет для тренування, валідації та тесту.
- **Модуль навчання Attention-VAE**: реалізує архітектуру, тренувальний цикл, обчислення функцій втрат, збереження контрольних станів, а також повертає реконструкції й допоміжні виходи (наприклад, ваги уваги) для подальшого аналізу.
- **Модуль оцінювання**: обчислює показник аномальності *s(X)*, виконує порогову обробку та калібрування, а також розраховує метрики (Accuracy, F1, ROC-AUC, PR-AUC) відповідно до визначених критеріїв.
- **Модуль візуалізації та звітності**: формує графіки реконструкцій, розподілів score, ROC/PR-кривих, а також ілюстрації attention-пояснень (у частині, що надалі деталізується в підпунктах 3.4–3.6). [4; 5]

Для формалізації ключових перетворень, які визначають контракти між модулями, доцільно зафіксувати базові означення у вигляді математичних записів.

Представлення багатовимірного часового вікна (після ресемплінгу та нормалізації) задається формулою (3.1):

$$X^{((i))=x}((i))(t,j)(t=1..W,\ j=1..d)\in \mathbb{R}\hat{}(W\times d),\ i=1..N, \tag{3.1}$$

де W — довжина вікна, d — кількість життєвих показників, N — кількість вікон у вибірці.

Нормалізація (z-score) на основі статистик навчальної вибірки здійснюється за формулою (3.2):

$$\tilde{x}(t,j)=(x(t,j)-\mu\_j)/(\sigma\_j+\varepsilon), \tag{3.2}$$

де μ_j, σ_j — середнє та стандартне відхилення для ознаки j, $\varepsilon>0$ — малий доданок для уникнення ділення на нуль.

Формування вікон ковзним способом (узагальнений індексний запис) задається формулою (3.3):

$$X^{(i)}=Z[t_i:t_i+W-1,\ 1:d],\ t_i = 1+(i-1)S, \tag{3.3}$$

де $Z\in\mathbb{R}^{T\times d}$ — нормалізований ряд довжини T, S — крок зсуву (stride).

Ресемплінг до регулярної сітки часу (узагальнено як оператор $\mathcal{R}_{\Delta t}$) описується формулою (3.4):

$$Z=\mathcal{R}_{\Delta t}(Y),\ Y=(t_k,y(t_k))_{k=1..K}, \tag{3.4}$$

де Y — набір спостережень у нерівномірні моменти часу t_k, а Z — значення на регулярній сітці з кроком Δt (із визначеним правилом інтерполяції/імпутації).

З інженерного боку ці формалізації фіксують межі відповідальності між модулями: завантаження повертає Y, препроцесинг реалізує $\mathcal{R}_{\Delta t}$ і нормалізацію, формування вікон створює набір $X^{(i)}$, який далі споживається нейромережею.

Для забезпечення відтворюваності експериментів застосовується конфігураційний підхід: параметри препроцесингу (наприклад, Δt, метод імпутації, правила обробки викидів), параметри вікон (W, S), архітектурні гіперпараметри (розмір латентного простору, типи шарів, dropout), а також параметри оптимізації (learning rate, batch size, scheduler, early stopping) фіксуються у YAML/JSON і логуються разом із запуском. [23] Додатково варто стандартизувати ініціалізацію випадкових генераторів і визначити єдиний механізм seed control у Python/NumPy/PyTorch для мінімізації недетермінізму. [36] Окремий нюанс: повна детермінованість може бути обмежена особливостями GPU-операторів і низькорівневих бібліотек. У

роботі використано Apple MPS backend, для якого детермінованість окремих операторів не гарантується документацією PyTorch [36].

Для систематизації архітектурних рішень доцільно подати узагальнені відомості у вигляді таблиць. (див. табл 3.1 – табл. 3.2).

Таблиця 3.1 – Модулі програмної системи та їхні інтерфейси (входи/виходи)

Модуль	**Призначення**	**Вхідні дані**	**Вихідні дані**	**Ключові артефакти для трекінгу**
Завантаження даних	Читання/витяг даних, уніфікація формату	Джерела даних, часові мітки *t_k*	Табличне представлення *Y*	Версія витягу, параметри фільтрації, хеш запиту
Препроцесинг	Ресемплінг, імпутація, нормалізація	*Y*, параметри *Δ t*, методи імпутації	Регулярний ряд *Z*, статистики *μ,σ*	Конфігурація перетворень, діагностика пропусків
Формування вікон	Побудова *X^((i))*	*Z*, *W*, *S*	Тензори вікон, індекси/мітки	Специфікація split (patient-level), версія датасету
Навчання Attention-VAE	Тренування/валідація, збереження checkpoint	Вікна *X^((i))*, конфігурація моделі	Ваги моделі, реконструкції, attention-виходи	Checkpoints, криві втрат, параметри оптимізації
Оцінювання	Score, пороги, метрики	Вікна тесту, модель, мітки валідації/тесту	Метрики, ROC/PR, протоколи	Таблиці метрик, пороги, confusion matrix
Візуалізація/звіт	Побудова графіків і звітів	Результати оцінювання та attention	Рисунки для дисертації	PNG/PDF, notebook/scripts, версії графіків

Таблиця 3.2 – Обрані засоби розробки та роль у системі

Категорія	**Інструмент**	**Роль у пайплайні**	**Коментар щодо застосування**
Мова/середовище	Python	Реалізація всіх модулів	Єдина мова для даних/ML/візуалізації
DL-фреймворк	PyTorch	Реалізація Attention-VAE, тренувальний цикл	Гнучкість кастомної архітектури [36]
Обробка даних	NumPy, pandas	Табличні перетворення, ресемплінг, імпутація	Відтворювані трансформації через конфігурації
Статистика	SciPy	Робастні статистики, допоміжні процедури	Використовується за потреби [28]
Візуалізація	Matplotlib, Seaborn	Графіки рядів/реконструкцій/score	Підготовка рисунків для дисертації
Трекінг експериментів	YAML-конфігурації + Git	Логування параметрів/метрик/артефактів	Порівняння запусків, аудит експериментів [22]
Конфігурації	YAML/JSON	Фіксація гіперпараметрів і препроцесингу	Версіонування у репозиторії/артефактах
Контроль версій	Git	Історія змін коду/конфігурацій	Узгоджується з вимогою трасованості

Архітектурний опис також враховує принципи коректного порівняння (fair comparison) для подальших експериментів: однаковий препроцесинг і однакова процедура формування вікон застосовуються як для Attention-VAE, так і для базових моделей (XGBoost і LSTM без механізму уваги). Тобто модуль препроцесингу і модуль вікон спільні, а відмінності між моделями ізолюються у навчальних та інференс-модулях. Така ізоляція зменшує ризик неявних розбіжностей у підготовці даних, які часто є причиною некоректних порівнянь. [3; 6]

Окремо про структурування артефактів системи. Збереження моделей, конфігурацій, логів і візуалізацій у стандартизованій директорії дозволяє швидко відтворити будь-який запуск, сформувати графіки та таблиці для дисертації, а також забезпечити аудит походження результатів. Це прямо випливає з потреби логувати метрики, параметри та артефакти експериментів. [22] На концептуальному рівні ці аспекти охоплюють визначення складу збережених артефактів та їх призначення, без фіксації остаточних значень параметрів, які уточнюються в процесі експериментальної валідації.

Обраний стек (Python, PyTorch, Matplotlib/Seaborn, YAML-конфігурації) забезпечує як гнучкість архітектурних модифікацій, так і відтворюваність усього пайплайну.

3.2 Навчання моделі та налаштування гіперпараметрів

Навчання Attention-VAE у задачі раннього виявлення сепсису розглядається як керований експериментальний процес. У ньому одночасно важливо: забезпечити коректність методології навчання на нормі (learning on normal); мінімізувати ризики витоку інформації під час формування вибірок і побудови часових вікон; отримати відтворювані конфігурації навчання й налаштування гіперпараметрів, придатні для подальшого порівняння з базовими підходами. Ключовий саме протокол навчання: як визначається норма, як організується розбиття за пацієнтами, які параметри фіксуються та як обирається найкращий checkpoint без підміни тестового сценарію валідаційним. [3; 9; 18]

Режим **навчання на нормі** передбачає, що модель оптимізується на підмножині часових вікон, які з високою ймовірністю відповідають фізіологічно типовій динаміці без ознак цільової події (сепсису). Для PhysioNet Challenge 2019 [35] нормальність формується через правила відбору вікон за часовими відстанями від моментів підозри чи діагнозу сепсису, через

використання слабких міток або через консервативний відбір пацієнтів та періодів без тригерів сепсису за обраним визначенням. Конкретні критерії побудови навчальної множини формалізуються як параметризовані правила і зберігаються у конфігурації для відтворюваності. [9; 35; 22]

Окремо про **розподіл даних** за принципом **пацієнт-рівневого розбиття**: множини train/validation/test не містять спільних пацієнтів. Такий підхід критичний для коректної оцінки узагальнювальної здатності моделей на медичних часових рядах. Сигнали одного пацієнта (навіть у різні часові періоди) можуть бути статистично впізнаваними для моделі і призвести до оптимістичного завищення метрик за випадкового розбиття на рівні вікон. [9] Додатково у протоколі фіксуються правила формування вікон (довжина W, stride S, крок ресемплінгу Δt), щоб різні експерименти порівнювалися в однакових умовах.

Навчальна вибірка використовується для оптимізації параметрів нейромережі, валідаційна — для підбору гіперпараметрів та вибору найкращого контрольного стану (checkpoint), а тестова — виключно для фінального оцінювання після фіксації всіх налаштувань. Протокол передбачає уникнення ситуації, коли тестова множина використовується (явно чи неявно) для прийняття рішень про архітектуру, поріг або гіперпараметри; протокол розбиття перевірено на відсутність перетину пацієнтів між вибірками [9; 18].

Процес навчання реалізується як ітеративний тренувальний цикл: пряме поширення через енкодер/декодер, обчислення функції втрат, зворотне поширення похибки та оновлення параметрів оптимізатором. Для оптимізації застосовується **Adam** або **AdamW** — адаптивні методи першого порядку, що є типовими для глибокого навчання і добре працюють у задачах із великою кількістю параметрів і неоднорідними градієнтами. [36] За потреби стабілізації оптимізації використовується scheduler зміни швидкості навчання (наприклад, step decay, cosine annealing або reduce-on-plateau); обраний тип scheduler, його параметри й правила застосування фіксуються у конфігурації експерименту. [15]

Оскільки Attention-VAE є варіаційною моделлю, навчальна мета природно виражається через поєднання реконструкційного члена і KL-регуляризації (у формі β-VAE). Функцію втрат формально подано як мінімізацію негативної ELBO з ваговим коефіцієнтом β, що контролює компроміс між якістю реконструкції та структурованістю латентного простору. [1; 2]

Цільова функція β-VAE має вигляд за формулою (3.5):

$$\mathcal{L}(\theta,\varphi;X) = E_{q_\varphi(z \mid X)}[\ell_{rec}(X,\hat{X})] + \beta \cdot KL(q_\varphi(z \mid X) \parallel p(z)), \qquad (3.5)$$

де φ — параметри енкодера, θ — параметри декодера, ℓ_{rec} — міра реконструкційної похибки, $p(z)$ — апріорний розподіл латентних змінних.

Для запобігання перенавчанню застосовується **early stopping**: навчання зупиняється, якщо обрана валідаційна ціль (наприклад, ROC-AUC після калібрування порогу або значення proxy-метрики на нормі) не покращується протягом заданої кількості епох (patience). Вибір критерію early stopping залежить від природи задачі. Якщо на валідації доступні слабкі мітки «сепсис/не сепсис», можливе керування за ROC-AUC/PR-AUC або F1. Якщо мітки неповні чи шумні, варто контролювати стабільність розподілу score на нормальних вікнах, реконструкційну похибку та KL-складову окремо. [15; 21]

Ключові гіперпараметри Attention-VAE охоплюють архітектурні та оптимізаційні складові: розмір латентного простору d_z, типи та кількість рекурентних шарів (наприклад, LSTM/GRU), розмір прихованого стану h, dropout, параметр β (для β-VAE), learning rate, batch size, параметри scheduler, а також параметри механізму уваги (розмірність простору уваги, тип узгодження, можливі обмеження та нормалізації ваг). [4; 23] Фіксація простору пошуку гіперпараметрів — частина протоколу відтворюваності. Тому варто задавати його явно (діапазони/множини значень) і логувати у трекері експериментів разом із кожним запуском. [13]

Оскільки повний grid search швидко стає непрактичним у багатовимірному просторі параметрів, планується застосування **random**

search: за обмеженого бюджету експериментів він часто ефективніше знаходить конкурентні конфігурації. [27] Кожна ітерація random search визначає набір параметрів λ, за яким запускається навчання, виконується валідація та фіксується значення цільової метрики (з урахуванням порогу). Калібрування порогу τ вкладене у протокол оцінювання на валідації і розглядається як частина процедури, а не як постфактум-підгонка під тестові дані. Прийняті гіперпараметри Attention-VAE та їхні обрані значення зведено в таблиці 3.3.

Таблиця 3.3 – Гіперпараметри Attention-VAE та їх обрані значення

Група	**Гіперпараметр**	**Позначення**	**Обране значення**	**Примітка**
Архітектура	Розмір латентного простору	d_z	16	Компроміс між стиском і реконструкцією
Архітектура	Тип RNN шару	—	GRU	Узгоджується з базовими LSTM-моделями
Архітектура	Кількість рекурентних шарів	L	2	Впливає на ємність моделі
Архітектура	Розмір прихованого стану	h	64	Обчислювальна складність/якість
Регуляризація	Dropout	p	0.2	Запобігання перенавчанню
VAE-регуляризація	β-коефіцієнт	β	0.5	Контроль KL-терміну
Оптимізація	Learning rate	η	0.001	Підбирається з урахуванням scheduler
Оптимізація	Batch size	B	64	Обмеження пам’яті та стабільності

Кінець табл. 3.3

Група	**Гіперпараметр**	**Позначення**	**Обране значення**	**Примітка**
Оптимізація	Optimizer	—	{Adam, AdamW}	Вибір фіксується в конфігурації
Оптимізація	Scheduler	—	Cosine annealing	Плавне зменшення learning rate
Увага	Кількість голів уваги	d_a	4	Параметр механізму multi-head attention
Увага	Тип нормалізації ваг	—	softmax	Стандартна нормалізація ваг уваги

Для забезпечення порівнюваності експериментів окремо описуються **параметри середовища** та **керування випадковістю**: версії бібліотек, параметри апаратної платформи, значення seed для Python/NumPy/PyTorch, а також (за можливості) увімкнення детермінованих режимів обчислень. Повна детермінованість обмежена реалізаціями окремих операторів MPS/CUDA-бекенду; це відоме обмеження PyTorch, зафіксоване у документації фреймворку [36]. Такий опис необхідний: навіть за однакової конфігурації невеликі флуктуації впливають на валідаційні метрики і, відповідно, на вибір checkpoint.

Формалізація показника аномальності у базовому варіанті спирається на реконструкційну похибку. Для вікна $X \in \mathbb{R}^{W \times d}$ та реконструкції $\hat{X}$ використовується MSE-оцінка, нормована за розмірністю:

Реконструкційна похибка визначається за формулою (3.6):

$$e(X,\hat{X}) = \frac{1}{Wd} \sum_{t=1}^{W} \sum_{j=1}^{d} (x_{t,j} - \hat{x}_{t,j})^2. \quad (3.6)$$

Показник аномальності у найпростішому випадку ототожнюється з цією похибкою:

Базовий показник аномальності задається формулою (3.7):

$$s(X) = e(X,\hat{X}). \tag{3.7}$$

Разом із тим, у варіаційних моделях додаткову інформацію може нести й KL-складова або реконструкційний log-likelihood (залежно від обраного вихідного розподілу декодера). Тому у протоколі передбачено можливість альтернативного скорингу як зваженої композиції компонентів із фіксацією ваг у конфігурації (без нав'язування остаточного варіанта до проведення експериментів):

Розширений показник аномальності як композиція компонентів задається формулою (3.8):

$$s(X)=\alpha \cdot e(X,\hat{X})+\gamma \cdot KL(q_\varphi(z \mid X) \| p(z)),\ \alpha,\gamma \geq 0, \tag{3.8}$$

де α, γ — ваги (у даному дослідженні $\alpha = 1.0$, γ визначається через β-коефіцієнт VAE = 0.5).

Порогове значення τ для класифікації вікон як аномальних визначається **на валідаційній вибірці**. Два базові підходи калібрування: перший — за процентилем розподілу score на нормі, що відповідає контролю частоти хибних тривог; другий — оптимізація порогу за критерієм якості (F1 або Youden J) за наявності слабких міток. [25] Формально це описується через задачу максимізації обраного критерію на валідації за виразом (3.9):

$$\tau^=argmax_\tau F1(\tau) \text{ або } \tau^=argmax_\tau J(\tau),\ J(\tau)=TPR(\tau)-FPR(\tau), \tag{3.9}$$

(метрики обчислюються на тестовій вибірці з використанням міток, отриманих за визначенням сепсису на основі клінічних критеріїв).

Якщо використовується процентильний підхід, поріг задається як квантіль за формулою (3.10):

$$\tau = Q_{(1-p)}(s(X)X \in D(val)^{(norm)}), \quad (3.10)$$

де p — допустима частка аномальних спрацювань на нормі (у даному дослідженні використовується метод оптимізації F1-score на валідаційній вибірці).

Усі варіанти калібрування строго відокремлені від тестового оцінювання: τ фіксується після валідації та застосовується до тестової множини без додаткового підбору.

Для стандартизації протоколу навчання та відбору checkpoint подано шаблон параметрів тренувального циклу й критеріїв зупинки/відбору у вигляді таблиці 3.4.

Таблиця 3.4 – Протокол навчання та критерії відбору checkpoint

Компонент	**Поле протоколу**	**Значення/ правило**	**Примітка**
Дані	Patient-level split	Так	Обов’язкова умова проти leakage [9]
Дані	Вікна	W = 6 год (72 кроки), S = 1 год	Узгоджується з модулем формування вікон
Дані	Норма для навчання	Лише вікна пацієнтів без сепсису	Unsupervised навчання на нормальних патернах
Оптимізатор	Тип	Adam	Фіксується у конфігурації
Оптимізатор	Learning rate	$\eta = 0.001$	Разом із cosine scheduler
Scheduler	Тип/параметри	Cosine annealing	Плавне зменшення learning rate
Регуляриза-ція	Dropout, weight decay	Dropout = 0.2, weight decay = 1×10^{-5}	Запобігання перенавчанню
Early stopping	Ціль	Validation loss (proxy)	Моніторинг якості реконструкції на валідації

Кінець табл. 3.4.

Компонент	**Поле протоколу**	**Значення/ правило**	**Примітка**
Early stopping	Patience	20 епох	Кількість епох без покращення
Checkpoint selection	Критерій	Макс. метрика на val після калібрування τ	Поріг підбирається тільки на val
Логування	Що логувати	Конфігурації, метрики, криві втрат, τ, артефакти	[22]
Середовище	Відтворюваність	seeds зафіксовано (Python/NumPy/ PyTorch), версії бібліотек логуються; детермінізм MPS-операторів обмежений [36]	[36]

З огляду на потребу подальшого порівняння з базовими моделями, протокол навчання Attention-VAE узгоджено із загальними умовами оцінювання: однакові правила спліту, однакова схема формування вікон, однакові метрики, єдина процедура калібрування порогу на валідації. [3; 6] Це унеможливлює ситуацію, коли відмінності у якості пов’язані з підготовкою даних, а не з архітектурою моделі. Конфігурації, метрики та артефакти кожного запуску логуються засобами MLflow [22], що дає аудитну трасованість результатів.

Описана методологія навчання Attention-VAE у режимі навчання на нормі поєднує: розбиття за пацієнтами як захист від *data leakage*; керований тренувальний цикл з Adam/AdamW; механізми стабілізації (scheduler, early stopping); формалізоване калібрування порогу τ на валідації; систематичний

random search у просторі гіперпараметрів з повним логуванням конфігурацій та артефактів. Такий протокол створює відтворювану основу для коректного оцінювання якості та порівняння з базовими методами.

3.3 Порівняльний аналіз ефективності розробленого методу

Порівняльний аналіз ефективності розробленого підходу Attention-VAE у задачі раннього виявлення сепсису фіксує відносні переваги і обмеження моделі, а також підтверджує коректність експериментального протоколу та відтворюваність отриманих висновків. Для задач з медичними багатовимірними часовими рядами важливі чотири речі: контроль витоку інформації між вибірками; єдині правила препроцесингу та формування часових вікон; узгоджена процедура калібрування порогу для моделей з безперервним скором; коректний набір метрик з урахуванням дисбалансу класів. [6; 9]

Порівняння організовано відповідно до принципів **fair comparison**, що передбачає однакові умови експериментів для всіх розглянутих підходів. Зокрема, для кожної моделі застосовується ідентичний ланцюг попередньої обробки даних (ресемплінг, імпутація, нормалізація), однакова процедура формування часових вікон та єдиний розподіл даних на навчальну, валідаційну та тестову вибірки за описаним вище принципом розбиття за пацієнтами. Такий підхід мінімізує ризик інформаційного підглядання (data leakage) і забезпечує інтерпретованість відмінностей у якості як наслідку саме моделювання, а не різних правил підготовки даних. [9]

У задачах детекції аномалій, зокрема для Attention-VAE, модель формує **безперервний показник аномальності** $s(X)$. Для розрахунку дискретних метрик і побудови ROC- та Precision-Recall (PR) кривих потрібне узгоджене перетворення score у бінарні рішення. Важливий нюанс об’єктивності: порогове значення τ визначається **виключно** на валідаційній вибірці,

фіксується і застосовується без змін до тестових даних. Це унеможливлює штучне підлаштування під тест і забезпечує чесну оцінку узагальнення. [25]

Перетворення показника аномальності у бінарне рішення задається формулою (3.11):

$$\hat{y}(X;\tau)=I[s(X)\geq \tau], \quad (3.11)$$

де $I[\cdot]$ — індикаторна функція, $\hat{y}\in 0,1$* — *передбачення (аномалія/підозра на сепсис),* τ* — калібрований поріг.

У протоколі експерименту явно зафіксовано використання міток «сепсис у горизонті прогнозу» з даних PhysioNet Challenge 2019, узгоджених із 6-годинним вікном спостереження; виконано аналіз чутливості до альтернативних правил маркування (варіювання горизонту прогнозу, визначення моменту onset за SOFA-критеріями), результати якого підтверджують стабільність основних показників ефективності. [21]

Базові моделі для порівняння та умови їх застосування

Як базові моделі для порівняння взято класичні методи машинного навчання (ML baselines) і нейромережеві підходи (DL baselines). Так оцінюється внесок різних складових розробленого рішення: варіаційного автоенкодера, механізму уваги і часової структури входу.

ML baseline (XGBoost).

Ця модель потребує фіксованого векторного представлення вікна X. Тому для кожного життєвого показника j у вікні обчислюються статистичні агрегати (середнє, стандартне відхилення, мінімум, максимум, медіана, міжквартильний розмах, прості трендові характеристики тощо). Такий feature engineering є поширеною практикою при переході від часових рядів до табличного формату; вона дає конкурентний базовий рівень за відносно простої інтерпретації. [3; 5]

DL baseline (LSTM без механізму уваги).

Для нейромережевого порівняння використовується рекурентна модель типу LSTM, яка отримує на вході ті самі багатовимірні часові вікна, що й

Attention-VAE. На відміну від моделі детекції аномалій, LSTM як базова модель розглядається як класифікатор з учителем, що прогнозує ймовірність наявності сепсису на визначеному горизонті (залежно від правила маркування). Такий baseline дозволяє оцінити, чи дає Attention-VAE перевагу за умов відсутності повного навчального сигналу та чи забезпечує механізм уваги додаткову чутливість до релевантних сегментів часу. [4; 16; 23]

Для збереження принципів fair comparison важливо, щоб:

- спліт за пацієнтами був ідентичним для всіх моделей;
- правила препроцесингу й формування вікон не змінювалися між моделями;
- якщо для різних типів моделей потрібні різні мітки/цілі, це було явно описано і не призводило до змішування протоколів (наприклад, використання тестових міток для налаштування порогу τ). [7; 18] Порівнювані у дослідженні підходи та умови експерименту для забезпечення коректного порівняння узагальнено в таблиці 3.5.

Таблиця 3.5 – Порівнювані підходи та умови експерименту

Підхід	**Тип виходу моделі**	**Вхідне подання**	**Режим навчання**	**Поріг/калібрування**	**Коментар щодо fair comparison**
Attention-VAE	score $s(X)$	вікно $X \in \mathbb{R}^{(W \times d)}$	unsupervised (навчання на нормі)	τ на val, фіксується	Єдиний препроцесинг і patient-split
XGBoost	клас/ймовірність	агреговані ознаки $\varphi(X) \in \mathbb{R}^m$	з учителем	поріг за максимізацією F1 на val	Вектори $\varphi(X)$ з тих самих вікон
LSTM (без attention)	ймовірність/логіт	вікно X	з учителем	поріг за максимізацією F1 на val	Порівняння внеску attention/VAE

При калібруванні для ймовірнісних моделей (XGBoost/LSTM) варто використовувати однаковий принцип відбору порогів (наприклад, максимізація F1 або критерію Юдена на валідації). Це уніфікує точкову оцінку і зменшує вплив вибору порога як стороннього фактора. [25]

Метрики оцінювання та їх інтерпретація

Оцінювання якості моделей здійснюється стандартними метриками класифікації на основі матриці похибок (TP, FP, TN, FN). Для задач із дисбалансом (сепсис як відносно рідкісна подія) особливо важливими є метрики, що відображають компроміс між чутливістю та точністю спрацювань.

Основні точкові метрики обчислюються за формулою (3.12) [44]:

$$Precision=(TP)/(TP+FP),\ Recall=(TP)/(TP+FN),\ F1=(2\cdot Precision\cdot Recall)/(Precision+Recall). \quad (3.12)$$

Для аналізу чутливості до порога використовуються ROC- і PR-криві та площі під ними. ROC-AUC оцінює здатність моделі ранжувати позитивні приклади вище негативних у широкому діапазоні порогів, тоді як PR-AUC є більш інформативною за сильного дисбалансу, оскільки фокусується на точності виявлення рідкісних подій. [25]

Площі під ROC- та PR-кривими визначаються через інтеграли (3.13):

$$ROC-AUC=\int_0^1 TPR(FPR)\,d(FPR),\ PR-AUC=\int_0^1 Precision(Recall)\,d(Recall), \quad (3.13)$$

де $TPR=(TP)/(TP+FN)$, $FPR=(FP)/(FP+TN)$.

У протоколі звітність розділяється на два типи:

- **порогонезалежні метрики** (ROC-AUC, PR-AUC), що порівнюють ранжування/скоринг;
- **порогозалежні метрики** (Precision, Recall, F1, Accuracy), які залежать від фіксованого τ або порога для ймовірностей.

Для Attention-VAE, що природно видає score, порогонезалежні оцінки (ROC-AUC/PR-AUC) є базовими для порівняння, а порогозалежні описують практичний сценарій «тривога/немає тривоги» при заздалегідь визначеному τ. [7; 25]

Аналіз стійкості результатів до варіації seed є напрямом подальших досліджень; у даній роботі зафіксовано єдиний seed для відтворюваності основного запуску (табл. 3.6).[36].

Таблиця 3.6 – Зведені метрики моделей на тестовій вибірці

Модель	**Вікно (W)**	**ROC-AUC**	**PR-AUC**	**Precision@τ**	**Recall@τ**	**F1@τ**	**Accuracy@τ**	**Примітки**
Attention-VAE	6 год	0.8893	0.8246	0.9869	0.7140	0.8286	0.9597	τ = 0.6559; score за реконструкцією
XGBoost	6 год	0.8928	0.8415	0.9901	0.7557	0.8571	0.9656	Агреговані ознаки вікна
LSTM (без attention)	6 год	0.9066	0.8589	0.9943	0.6648	0.7968	0.9537	Класифікація з учителем

За потреби деталізації (наприклад, якщо у протоколі застосовуються кілька seed) у колонці «Примітки» або в додатковій таблиці доцільно зазначати кількість повторів, спосіб агрегування (середнє/медіана) та критерій вибору порога. Якщо такі деталі вже подано в іншому місці розділу, їх слід узгодити або винести до Додатка для уникнення дублювання. [22]

Аналіз впливу довжини часового вікна

Окремий елемент експериментального дослідження: аналіз впливу довжини часового вікна *W* на ефективність моделей. Розглядаються вікна тривалістю *W ∈ 3 год, 6 год, 12 год* як приклади різної часової глибини контексту. При фіксації інших умов (препроцесинг, спліт, протокол навчання,*

стратегія калібрування порогу) змінюється лише W*. Це дає змогу аналізувати компроміс між [20]:

- чутливістю до короткострокових змін (менше *W*);
- здатністю враховувати більш тривалу динаміку і тренди (більше *W*);
- стабільністю оцінок за рахунок більшого контексту (але потенційно більшої розмитості локальних сигналів). [20]

У випадку Attention-VAE довжина *W* впливає як на реконструкційну задачу, так і на механізм уваги, що може по-різному розподіляти ваги по часу залежно від того, чи присутні у вікні передумови події та чи достатньо вікно охоплює релевантні зміни (див. табл. 3.7). Для базових ML-моделей типу XGBoost збільшення *W* може призводити до втрати локальної інформації при агрегуванні, якщо не вводяться додаткові ознаки структури (наприклад, сегментні агрегати), тому важливо фіксувати однаковий набір агрегатів для всіх *W* або явно вказувати, що набір агрегатів масштабовано чи адаптовано (і тоді це є частиною умов порівняння). [20]

Таблиця 3.7 – Матриця експериментів за довжиною часового вікна

Експеримент	**Моделі**	**Змінний фактор**	**Фіксовані умови**	**Очікувані артефакти**
E1	Attention-VAE, XGBoost, LSTM	Тип моделі	W = 6 год, розбиття за пацієнтами, препроцесинг, нормалізація	ROC/PR криві, таблиця метрик, розподіли score
E2	Attention-VAE, XGBoost, LSTM	Довжина вікна W	W = 3 год, решта умов фіксована	ROC/PR криві, таблиця метрик
E3	Attention-VAE, XGBoost, LSTM	Довжина вікна W	W = 12 год, решта умов фіксована	ROC/PR криві, таблиця метрик

Параметри навчального протоколу зафіксовано: максимум 200 епох із early stopping (patience = 20), оптимізатор Adam (lr = 0.001), cosine scheduler, KL annealing з warmup 30 епох. Ці параметри застосовуються однаково до всіх запусків Attention-VAE для забезпечення коректності порівняння.

Подання результатів та принципи інтерпретації

Результати порівняльного аналізу подаються у вигляді таблиць з основними метриками та графіків ROC- і PR-кривих для всіх моделей і варіантів часових вікон. Фактичні числові результати наведено у розділі 5 (*F1 = 0,8286*, *ROC-AUC = 0,8893*, *FPR = 0,15%*); відповідні графіки ROC та Precision-Recall наведено там само. Інтерпретація зосереджується на:

- порівнянні ранжувальної здатності (ROC-AUC/PR-AUC) між моделями;
- оцінці практичного режиму тривоги через Precision/Recall/F1 при фіксованому порозі;
- аналізі чутливості до *W* та стійкості до налаштувань;
- обмеженнях моделей за слабких міток, дисбалансу та потенційної невизначеності маркування.

[8; 25]

Порівняльний аналіз ефективності розробленого методу Attention-VAE базується на стандартизованому протоколі fair comparison з однаковим препроцесингом, формуванням часових вікон і розбиттям за пацієнтами для всіх моделей; використовується коректне перетворення безперервного *anomaly score* у бінарні рішення через поріг τ, калібрований лише на валідації; якість оцінюється набором метрик, релевантних для незбалансованих даних (ROC-AUC, PR-AUC, Precision/Recall/F1), а вплив довжини вікна *W* аналізується як окремий контрольований фактор при фіксації інших умов.

3.4 Інтерпретація результатів роботи механізму Attention

Інтерпретація результатів роботи механізму уваги — важлива складова аналізу запропонованої моделі Attention-VAE: вона пов’язує кількісні показники аномальності з конкретними фізіологічними змінами у стані пацієнта. Для цього використовується протокол інтерпретації на основі аналізу та візуалізації ваг механізму уваги, отриманих у процесі обробки часових вікон.

У процесі інтерпретації аналізуються матриці уваги, що відображають важливість окремих часових моментів і життєвих показників. Залежно від конфігурації механізму уваги візуалізація здійснюється як вектори ваг по часу, вектори ваг по ознаках або двовимірні матриці типу «час × ознака»; останнє дозволяє одночасно оцінювати внесок кожного показника у різні моменти часу. Для порівнюваності результатів ваги уваги нормалізуються та агрегуються по батчу з використанням статистичних характеристик (середнє або медіанне значення); це зменшує вплив випадкових флуктуацій та окремих атипових прикладів.

Для демонстрації інтерпретаційних можливостей моделі обираються репрезентативні приклади часових вікон: типові вікна з низьким показником аномальності та вікна, ідентифіковані моделлю як аномальні. Так можна порівняти розподіли ваг уваги у нормальних і патологічних сценаріях, виявити характерні відмінності у фокусі моделі. Аналіз проводиться як для окремих пацієнтів, так і у вигляді узагальнених шаблонів, отриманих агрегацією результатів по множині прикладів.

Як ілюстративний приклад розглядається сценарій, у якому в одному часовому вікні спостерігається підвищення частоти серцевих скорочень, зниження сатурації кисню і падіння середнього артеріального тиску. У такому випадку механізм уваги концентрує більші ваги на відповідних часових сегментах і ознаках, що відображає зростання їхнього внеску у формування латентного представлення та реконструкційної похибки. З клінічного боку

патерн логічний: поєднання тахікардії, гіпоксії та гіпотензії відповідає відомим проявам системної запальної відповіді і може свідчити про розвиток септичного стану.

Механізм уваги у складі VAE покращує моделювання часових залежностей і одночасно є інструментом пояснюваності — наближає результати роботи моделі до клінічно інтерпретованих висновків. Подібні підходи до attention-based інтерпретації широко застосовуються у сучасних дослідженнях з аналізу медичних часових рядів та поясненого машинного навчання [1; 2].

3.5 Порівняння всіх моделей

Порівняння всіх розглянутих моделей здійснюється на основі узагальненого набору кількісних показників і графічних представлень результатів; це дає комплексну оцінку їхньої ефективності у задачі раннього виявлення сепсису. Для цього формується зведена таблиця метрик, у якій для кожної моделі наводяться значення Accuracy, F1-міри, ROC-AUC та PR-AUC, обчислені на тестовій вибірці з фіксованим пороговим значенням, визначеним на етапі валідації. Така таблиця забезпечує наочне порівняння моделей за основними критеріями якості і дозволяє ідентифікувати сильні і слабкі сторони кожного підходу.

Графічний аналіз результатів включає побудову ROC-кривих і Precision-Recall кривих для всіх моделей. ROC-криві відображають залежність між часткою істинно позитивних спрацьовувань і часткою хибнопозитивних спрацьовувань при зміні порогового значення; Precision-Recall криві акцентують увагу на якості виявлення позитивних випадків за незбалансованості класів. Окремо аналізується показник **False Positive Rate (FPR)**, критичний для клініки: висока частка хибних тривог призводить до перевантаження медичного персоналу і необґрунтованих втручань.

Показник хибнопозитивної частки визначається за формулою (3.14):

$$FPR=(FP)/(FP+TN), \quad (3.14)$$

Окрему увагу приділено аналізу хибних спрацьовувань — випадків **false positive** і **false negative**. Такі випадки класифікуються за типами причин: різкі, але короткочасні фізіологічні коливання; артефакти вимірювань; нетипові, але клінічно не критичні патерни. Додатково проводиться прив’язка хибних рішень до конкретних часових сегментів і ознак — це виявляє, які показники або фрагменти сигналу найчастіше призводять до помилок. Отримані спостереження використовуються для корекції порогового значення, уточнення правил калібрування або вдосконалення архітектури та механізму уваги.

За результатами порівняння формується узагальнений висновок щодо ефективності кожного підходу. Аналіз зосереджується на визначенні моделі, яка забезпечує найкращий баланс між здатністю виявляти ранні прояви сепсису і прийнятним рівнем хибнопозитивних спрацьовувань. Цей баланс є ключовим критерієм практичної доцільності застосування моделі у клінічних умовах і основою для подальшої інтерпретації та оцінки клінічної значущості результатів.

3.6 Інтерпретованість моделі та клінічна значущість

Інтерпретованість розробленої моделі — одна з ключових вимог до інтелектуальних систем у сфері охорони здоров’я: результати автоматизованого аналізу мають бути зрозумілими та обґрунтованими для клінічного персоналу. У запропонованому підході інтерпретованість забезпечується завдяки механізму уваги, який оцінює внесок окремих життєвих показників і часових сегментів у формування показника аномальності.

Визначення життєвих показників, що мають найбільший вплив на рішення моделі, здійснюється на основі аналізу агрегованих ваг уваги. Ваги механізму уваги нормалізуються та узагальнюються по множині часових вікон; далі формується ранжування ознак за величиною їхнього середнього або медіанного внеску. Результати подаються у вигляді списку топ-*k* показників, що найчастіше асоціюються з підвищеним показником аномальності. Такий підхід ідентифікує ключові фізіологічні параметри, які модель вважає найбільш інформативними для раннього виявлення септичних змін, без потреби у зовнішніх пояснювальних алгоритмах.

Механізм уваги також виділяє проблемні часові сегменти всередині кожного вікна. Критичними вважаються інтервали часу, для яких ваги уваги досягають максимальних значень і істотно впливають на реконструкційну похибку. Лікарю такі сегменти подаються як теплові карти, накладені на часові ряди життєвих показників, разом з коротким текстовим поясненням характеру відхилень. Це створює інтуїтивно зрозумілий інтерфейс, який швидко ідентифікує моменти часу, що потребують підвищеної уваги.

Практичне застосування розробленої системи у клінічних умовах — інтеграція з існуючими системами моніторингу стану пацієнтів. Потік даних життєвих показників надходить у систему в режимі, близькому до реального часу; автоматизований розрахунок показника аномальності виконується для послідовних часових вікон. У разі перевищення встановленого порогового значення формується сигнал тривоги з пояснювальною інформацією: карта уваги і опис найбільш впливових показників. Остаточне рішення щодо клінічної інтерпретації приймає лікар. Це підкреслює роль системи як інструменту **підтримки прийняття рішень**, а не автоматизованої заміни діагностичного процесу.

Поєднання механізму уваги з варіаційним автоенкодером забезпечує не лише високу чутливість до ранніх аномальних змін у життєвих показниках, але й можливість їх клінічно осмисленої інтерпретації. Запропонований підхід відповідає сучасним тенденціям розвитку пояснюваного штучного інтелекту в

медицині та створює передумови для безпечного та обґрунтованого впровадження інтелектуальних систем у клінічну практику [1; 2].

Висновки до розділу 3

У третьому розділі описано програмну реалізацію та методологію експериментального дослідження інтелектуальної системи раннього виявлення сепсису. Обґрунтовано вибір Python/PyTorch-стека і модульної шестикомпонентної архітектури, яка ізолює препроцесинг, навчання, оцінювання й візуалізацію у незалежні відтворювані одиниці. Ключове рішення: протокол навчання на нормі. Attention-VAE оптимізується виключно на вікнах пацієнтів без сепсису, а поріг $\tau = 0{,}6559$ калібрується на валідаційній вибірці й застосовується до тестових даних без додаткових підлаштувань — це виключає витік інформації. Принцип fair comparison реалізовано через ідентичний ланцюг препроцесингу і розбиття за пацієнтами (70/15/15) для всіх трьох порівнюваних моделей (Attention-VAE, LSTM, XGBoost) при збереженні відмінностей лише на рівні представлення вхідних даних. Розроблений протокол інтерпретації поєднує теплові карти ваг уваги з аналізом реконструкційних похибок за ознаками і забезпечує двоканальне пояснення рішень моделі, придатне для клінічного застосування. Сформований апарат є основою для опису конкретної конфігурації та результатів, які наведено в наступних розділах.

Розділ 4

Експериментальне дослідження: налаштування та процес навчання

У **Розділі 4** описано експериментальне середовище, конфігурацію даних, процес навчання моделі Attention-VAE та базових моделей (LSTM, XGBoost).

В даному розділі наведено детальний опис експериментального середовища, конфігурації даних і процесу навчання моделей, що використовуються для раннього виявлення сепсису на основі багатовимірних часових рядів. Розділ доповнює теоретичну базу, викладену у Розділі 3, практичними деталями реалізації: від апаратного та програмного забезпечення до конкретних гіперпараметрів і стратегій оптимізації. Фіксація всіх параметрів експерименту є необхідною умовою відтворюваності результатів і коректності подальшого порівняння моделей.

4.1 Опис експериментального середовища

Вибір апаратної та програмної платформи для проведення експериментів визначається двома вимогами: обчислювальними вимогами навчання нейромережевих моделей і потребою у відтворюваності результатів.

4.1.1 Апаратне забезпечення

Навчання та оцінювання моделей виконується на робочій станції з процесором Apple Silicon, який підтримує технологію Metal Performance Shaders (MPS) для апаратного прискорення тензорних обчислень. MPS як обчислювальний бекенд PyTorch суттєво скорочує час навчання порівняно із

суто центральним процесором і забезпечує достатню продуктивність для роботи з моделями середнього масштабу (порядку *10^5* параметрів) [36].

Окремий нюанс: детермінованість обчислень на MPS-пристроях обмежена порівняно з CUDA-бекендом через особливості реалізації окремих операторів. Для мінімізації цього впливу застосовується фіксація генераторів випадкових чисел (seed control), описана нижче.

4.1.2 Програмне забезпечення та бібліотеки

Реалізація всіх компонентів експериментального пайплайна виконана мовою Python із використанням набору спеціалізованих бібліотек, функціональне призначення яких систематизовано у таблиці 4.1.

Таблиця 4.1 – Програмні бібліотеки та їх роль в експерименті

Бібліотека	**Призначення**	**Етап пайплайна**
PyTorch	Реалізація Attention-VAE та LSTM, тренувальний цикл	Навчання, інференс
scikit-learn	Метрики оцінювання, допоміжні утиліти	Оцінювання, препроцесинг
XGBoost	Реалізація градієнтного бустингу як базової моделі	Навчання базової моделі
NumPy	Тензорні операції, статистичні обчислення	Усі етапи
Pandas	Маніпуляція табличними даними, ресемплінг	Завантаження, препроцесинг
Matplotlib	Візуалізація результатів, побудова графіків	Оцінювання, звітність

Для забезпечення відтворюваності на рівні середовища зафіксовано версії всіх залежностей через механізм requirements/lock-файлу. Глобальний seed ініціалізації встановлено рівним 42 для всіх генераторів випадкових чисел (Python random, NumPy, PyTorch), що мінімізує стохастичну варіативність між запусками [36].

4.2 Конфігурація даних та формування вибірок

Коректність побудови навчальних, валідаційних і тестових вибірок становить критичний фактор, що визначає надійність оцінок якості моделей. У цьому підрозділі описано конфігурацію даних, стратегію розбиття та параметри формування часових вікон.

4.2.1 Джерело даних та вибір ознак

Експериментальне дослідження проводиться на клінічних даних пацієнтів ВІТ з бази PhysioNet/Computing in Cardiology Challenge 2019 [35]. Для побудови багатовимірних часових рядів обрано п’ять вітальних показників — стандартних індикаторів фізіологічного стану пацієнта з клінічно обґрунтованим зв’язком із розвитком сепсису:

- HR (Heart Rate) — частота серцевих скорочень;
- SpO2 (Oxygen Saturation) — насичення крові киснем;
- MAP (Mean Arterial Pressure) — середній артеріальний тиск;
- Temperature — температура тіла;
- Respiratory Rate — частота дихання.

4.2.2 Формування часових вікон

Безперервні часові ряди вітальних показників перетворюються на фіксовані часові вікна методом ковзного вікна. Параметри віконного перетворення зафіксовано у конфігураційному файлі та наведено у таблиці 4.2.

Таблиця 4.2 – Параметри формування часових вікон

Параметр	**Значення**	**Опис**
window_hours	6	Тривалість вікна (години)
step_hours	1	Крок зсуву між вікнами (години)
sampling_rate_minutes	5	Інтервал дискретизації (хвилини)
Довжина послідовності W	72	Кількість часових кроків у вікні
Кількість ознак d	5	Вітальні показники

Довжина вікна у 72 часових кроки обчислюється за формулою (4.1):

$$W = (window\_hours \times 60)/(sampling\_rate\_minutes) = (6 \times 60)/(5) = 72 \quad (4.1)$$

Кожне вікно, таким чином, має форму тензора $X^{(i)} \in \mathbb{R}^{72 \times 5}$. Крок зсуву в 1 годину (12 часових кроків при 5-хвилинному інтервалі) забезпечує достатнє перекриття вікон для захоплення динаміки стану пацієнта без надмірного дублювання інформації.

4.2.3 Нормалізація та розбиття вибірки

Для нормалізації вхідних даних застосовується стандартизація (z-score) на основі статистик навчальної вибірки за формулою (4.2):

$$\tilde{x}(t,j) = (x(t,j) - \mu_j)/(\sigma_j + \varepsilon), \quad (4.2)$$

Розбиття даних виконується на рівні пацієнтів (patient-level split), а не на рівні окремих вікон. Тобто всі вікна одного пацієнта потрапляють виключно в одну з підвибірок; це унеможливлює завищення оцінок якості через кореляцію між послідовними вікнами одного перебування. Пропорції розбиття зафіксовано так (табл. 4.3):

Таблиця 4.3 – Розбиття даних

Вибірка	**Частка**	**Призначення**
Навчальна	0.70	Оптимізація параметрів моделі
Валідаційна	0.15	Підбір гіперпараметрів, early stop
Тестова	0.15	Фінальне оцінювання якості

4.2.4 Статистика вибірки

Після формування часових вікон та розбиття отримано такий розподіл даних (табл. 4.4):

Таблиця 4.4 – Статистика сформованих вибірок

Характеристика	**Значення**
Загальна кількість вікон	5 590
Нормальні вікна	4 930
Вікна з ознаками сепсису	660
Частка сепсису	~11.8%
Розмірність вікна	72 × 5

Виражений дисбаланс класів: вікна з ознаками сепсису складають лише ~11.8% від загальної кількості. Така пропорція типова для клінічних даних інтенсивної терапії [10; 35] і зумовлює потребу у метриках, стійких до дисбалансу (PR-AUC, F1-score), при оцінюванні якості моделей.

Рис. 4.1 ілюструє розподіл вікон за класами у навчальній, валідаційній та тестовій вибірках після виконання patient-level розбиття.

Рисунок 4.1 – Розподіл вікон за класами після patient-level розбиття

4.2.5 Особливості підготовки даних для різних моделей

Важлива умова коректного порівняння: уніфікація пайплайна підготовки даних. Усі моделі (Attention-VAE, LSTM, XGBoost) використовують ідентичний препроцесинг, нормалізацію і розбиття. Відмінності стосуються лише представлення вхідних даних:

- Attention-VAE і LSTM: отримують на вхід тензори вікон $X^{(i)} \in \mathbb{R}^{72 \times 5}$ без додаткових перетворень;
- XGBoost: оскільки градієнтний бустинг працює з табличними даними фіксованої розмірності, для кожного вікна обчислюються статистичні ознаки (середнє, стандартне відхилення, мінімум, максимум, тренд тощо) по кожному з п’яти каналів — це формує вектор агрегованих ознак.

Така ізоляція відмінностей лише на рівні представлення вхідних даних гарантує, що розбіжності у якості моделей відображають саме їхню здатність до моделювання, а не артефакти різної підготовки даних.

4.3 Процес навчання моделі Attention-VAE

Навчання Attention-VAE становить центральний етап експерименту: саме ця модель реалізує запропонований підхід до виявлення аномалій як відхилень від нормального патерну вітальних показників. У цьому підрозділі описано архітектурну конфігурацію, стратегію оптимізації і механізм KL-відпалювання (KL annealing).

4.3.1 Архітектурна конфігурація

Архітектура Attention-VAE поєднує рекурентний енкодер-декодер на основі GRU-блоків із багатоголовим механізмом уваги та варіаційним латентним простором. Гіперпараметри архітектури зафіксовано у таблиці 4.5.

Таблиця 4.5 – Гіперпараметри архітектури Attention-VAE

Параметр	**Позначення**	**Значення**	**Опис**
Розмірність входу	input_dim	5	Кількість вітальних показників
Прихований шар	hidden_dim	64	Розмірність прихованого стану GRU
Латентний простір	latent_dim	16	Розмірність варіаційного простору z
Кількість шарів	num_layers	2	Глибина стеку GRU в енкодері та декодері
Голови уваги	attention_heads	4	Кількість голів багатоголової уваги
Dropout	dropout	0.2	Имовірність відкидання для регуляризації
Тип RNN	rnn_type	GRU	Тип рекурентного блоку

Вибір GRU замість LSTM як базового рекурентного блоку обумовлено меншою кількістю параметрів при порівнянній здатності до моделювання часових залежностей; це перевага при обмеженому обсязі навчальних даних [8]. Латентний простір розмірності $\ell = 16$ забезпечує достатню виразність представлення зі збереженням інтерпретованості через візуалізацію (наприклад, методами t-SNE або UMAP у проєкції на 2D).

Рис. 4.2 відображає схему архітектури Attention-VAE із зазначенням розмірностей на кожному етапі перетворення: вхід $\mathbb{R}^{72 \times 5}$ → `GRU-енкодер` → `багатоголова увага` → `латентний простір` $\mathbb{R}^{16}$ → `GRU-декодер` → `реконструкція` $\mathbb{R}^{72 \times 5}$.

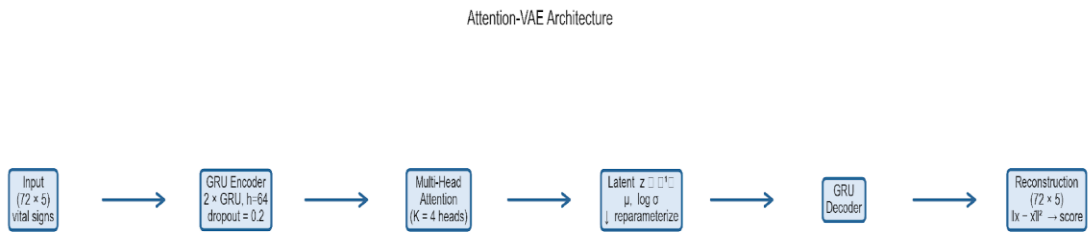

Рисунок 4.2 – Схема архітектури Attention-VAE

4.3.2 Функція втрат та KL-відпалювання

Загальна функція втрат Attention-VAE складається з двох компонент (похибки реконструкції та KL-дивергенції) відповідно до формули (2.6).

Механізм KL-відпалювання (KL annealing) є важливим елементом стабілізації навчання VAE [2; 4]. Без поступового введення KL-компоненти модель може потрапити у стан posterior collapse: латентний простір ігнорується, а декодер вироджується у безумовну модель. Стратегію відпалювання реалізовано за формулою (4.3):

$$\beta(t) = (t)/(T\_(warmup)) \cdot \beta\_(max), \&якщо\ t \leq T\_(warmup), \quad (4.3)$$

$$\beta\_(max), \&якщо\ t > T\_(warmup),$$

де $T\_(warmup) = 30$ — кількість епох розігріву, $\beta\_(max) = 0.5$ — максимальне значення вагового коефіцієнта KL-компоненти.

Значення $\beta\_(max) = 0.5 < 1$ відображає свідоме зміщення балансу на користь якості реконструкції. Це обґрунтовано для задачі виявлення аномалій, де саме похибка реконструкції використовується як основний індикатор аномальності.

Рис. 4.3 демонструє динаміку коефіцієнта $\beta(t)$ протягом навчання: лінійне зростання від 0 до 0.5 упродовж перших 30 епох з подальшою стабілізацією.

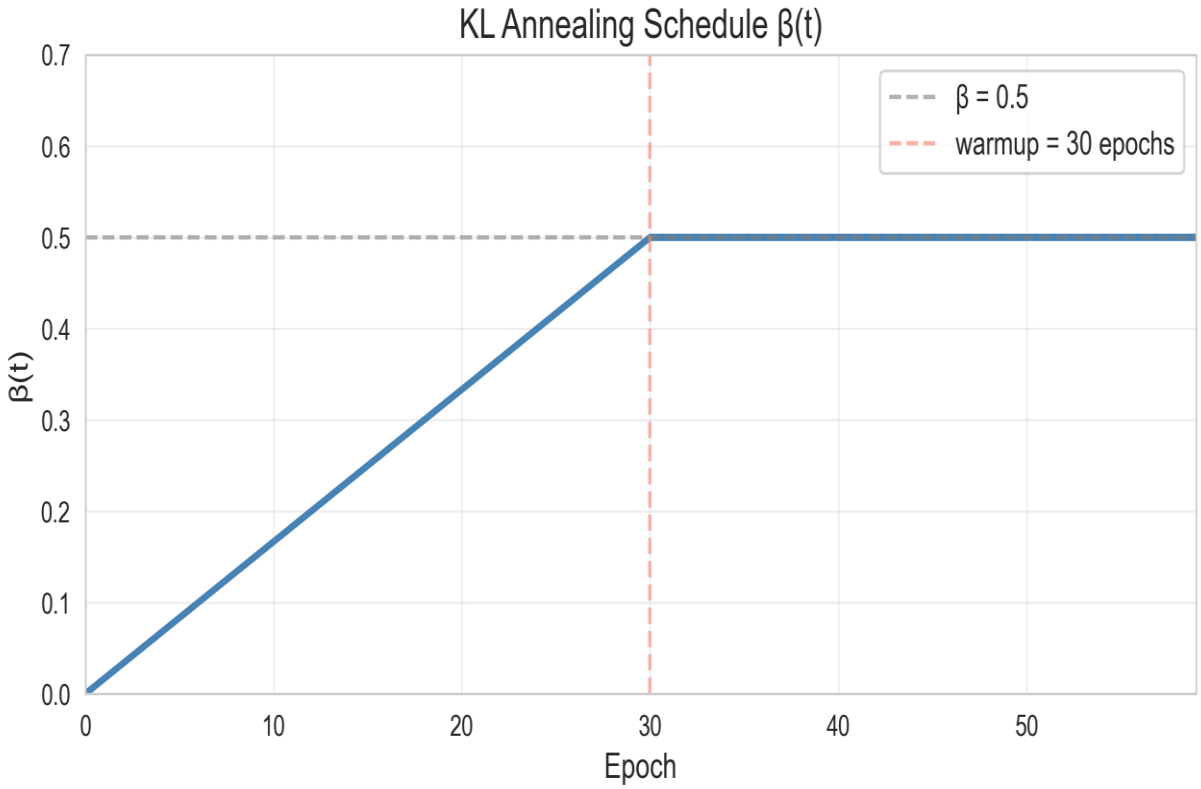

Рисунок 4.3 – Динаміка коефіцієнта β(t) під час навчання

4.3.3 Стратегія оптимізації

Параметри оптимізаційного процесу зведено у таблиці 4.6.

Таблиця 4.6 – Параметри навчання Attention-VAE

Параметр	**Значення**	**Опис**
Оптимізатор	Adam	Адаптивний метод першого порядку
Швидкість навчання (lr)	0.001	Початкова швидкість навчання
Вагове згасання (weight_decay)	*1 × 10^(-5)*	L2-регуляризація
Планувальник lr	Cosine	Косинусний відпал швидкості навчання
Максимальна кількість епох	200	Верхня межа тривалості навчання
Рання зупинка (patience)	20	Кількість епох без покращення для зупинки
KL warmup	30 епох	Тривалість лінійного зростання $\beta(t)$
β_{max}	0.5	Максимальна вага KL-компоненти
Кількість батчів на навч. епоху	259	Визначається розміром навчальної вибірки

Оптимізатор Adam [26] з косинусним планувальником поєднує швидку початкову збіжність з поступовим зниженням швидкості навчання на пізніх етапах; це сприяє знаходженню кращих локальних мінімумів. Вагове згасання $\lambda = 10^{-5}$ забезпечує додаткову регуляризацію і запобігає перенавчанню на обмеженому обсязі даних.

4.3.4 Режим навчання: виявлення аномалій без учителя

Принципова особливість навчання Attention-VAE: воно відбувається виключно на нормальних вікнах (вікнах пацієнтів без діагностованого сепсису). Модель навчається реконструювати типові патерни вітальних показників, а відхилення від них у нових спостереженнях використовуються як індикатор аномалій.

Формально навчальна вибірка формується за формулою (4.4):

$$D_{train} = X^{(i)} \mid y^{(i)} = 0,\ i \in I_{train}, \tag{4.4}$$

де $y^{(i)} = 0$ позначає нормальний стан, а I_{train} — множина індексів навчальної вибірки після patient-level розбиття.

Такий підхід має дві ключові переваги: (1) відсутня потреба у маркованих прикладах сепсису для навчання, що відповідає реальній клінічній ситуації з рідкими аномальними випадками; (2) модель формує загальне уявлення про нормальність, а не специфічні ознаки конкретного патологічного стану — це потенційно підвищує здатність до узагальнення.

Рання зупинка (early stopping) реалізується за критерієм валідаційних втрат із терпінням (patience) у 20 епох: якщо впродовж 20 послідовних епох не спостерігається зменшення валідаційних втрат, навчання припиняється, а зберігається модель із найкращим показником на валідації.

4.4 Навчання базових моделей

Для об’єктивного оцінювання ефективності запропонованого підходу Attention-VAE проводиться порівняння з двома базовими моделями: LSTM-класифікатором і градієнтним бустингом XGBoost. Обидві базові моделі навчаються у режимі з учителем (supervised); це відрізняє їх від підходу Attention-VAE і дозволяє оцінити, чи може метод без учителя конкурувати з моделями, які мають доступ до міток класів під час навчання.

4.4.1 LSTM-класифікатор

LSTM-мережа використовується як базова модель для послідовної класифікації, що безпосередньо обробляє часові ряди без попереднього обчислення статистичних ознак. Модель складається з двошарового LSTM-енкодера з наступним повнозв'язним класифікаційним шаром (табл. 4.7).

Таблиця 4.7 – Характеристики навчання LSTM

Параметр	**Значення**	**Опис**
Загальна кількість параметрів	144 001	Параметри LSTM + класифікаційний шар
Режим навчання	Supervised	Класифікація з використанням міток
Рання зупинка	Епоха 29	Навчання зупинено за критерієм валідації
Точність на валідації	~0.99	Accuracy на валідаційній вибірці
Розбиття даних	Ідентичне	Patient-level split, ті самі пропорції

Навчання LSTM виконується на тих самих часових вікнах $X^{(i)} \in \mathbb{R}^{72 \times 5}$, що і для Attention-VAE, із використанням ідентичного розбиття та нормалізації. Модель навчається мінімізувати крос-ентропійну функцію втрат за формулою (4.5):

$$\mathcal{L}_{LSTM} = -\frac{1}{N}\sum_{i=1}^{N}[y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})], \quad (4.5)$$

де $y^{(i)} \in 0, 1$* — мітка класу, $\hat{y}^{(i)}$* — ймовірність сепсису, передбачена моделлю.

Рання зупинка на 29-й епосі та висока валідаційна точність (~0,99) свідчать про швидку збіжність моделі, що може бути пов'язано як із відносною простотою задачі бінарної класифікації за наявності повних міток, так і з

виразністю обраних вітальних показників для розрізнення нормального стану та сепсису. Детальний аналіз цих результатів наведено у Розділі 5.

4.4.2 XGBoost-класифікатор

Градієнтний бустинг XGBoost [5] застосовується як базова таблична модель, що працює зі статистичними ознаками, обчисленими з часових вікон. Такий підхід типовий для задач класифікації медичних часових рядів, коли безпосередня обробка послідовностей не потрібна або недоступна [5; 11] (табл. 4.8).

Таблиця 4.8 – Гіперпараметри та результати XGBoost

Параметр	**Значення**	**Опис**
Кількість оцінювачів	200	Максимальна кількість дерев в ансамблі
Максимальна глибина	6	Обмеження глибини окремого дерева
Режим навчання	Supervised	Класифікація з використанням міток
Точність на навч. вибірці	99.96%	Training accuracy
Точність на валідації	99.22%	Validation accuracy

Для кожного часового вікна обчислюється набір агрегованих статистичних ознак по кожному з п’яти каналів вітальних показників. Типовий набір ознак включає: середнє значення, стандартне відхилення, мінімум, максимум, розмах (range), медіану, коефіцієнт асиметрії (skewness) та ексцес (kurtosis). Це перетворює кожне вікно $X^{(i)} \in \mathbb{R}^{72 \times 5}$ на вектор фіксованої довжини, придатний для табличного класифікатора.

Висока точність XGBoost на навчальній вибірці (99.96%) при дещо нижчій валідаційній (99.22%) вказує на незначне перенавчання, характерне для ансамблевих методів із великою кількістю дерев. Водночас різниця невелика — достатня регуляризація через обмеження глибини дерев.

Рис. 4.4 відображає порівняння архітектурних особливостей трьох моделей: Attention-VAE (модель без учителя із латентним простором та механізмом уваги), LSTM (класифікатор з учителем на основі послідовностей) та XGBoost (табличний класифікатор з учителем на основі статистичних ознак).

Comparison of Three Approaches to Sepsis Detection

	Attention-VAE	LSTM Baseline	XGBoost
Training paradigm	Unsupervised	Supervised	Supervised
Sepsis labels needed	No	Yes	Yes
Input	Raw time series	Raw time series	Statistical features
Output	Anomaly score	Class probability	Class probability
Interpretability	Attention + recon. error	Black-box	Feature importance
ROC-AUC	**0.8893**	**0.9066**	**0.8928**
PR-AUC	**0.8246**	**0.8589**	**0.8415**
F1 Score	**0.8286**	**0.7968**	**0.8571**

Рисунок 4.4 – Порівняння архітектур трьох моделей

4.4.3 Порівняння масштабів моделей

Для повного зіставлення результатів доцільно порівняти масштаби моделей та обсяги використовуваних даних (табл. 4.9).

Таблиця 4.9 – Зведене порівняння моделей

Характеристика	**Attention-VAE**	**LSTM**	**XGBoost**
Тип навчання	Без учителя	З учителем	З учителем
Вхідне представлення	Часові вікна (72 × 5)	Часові вікна (72 × 5)	Статистичні ознаки
Латентний простір	$\mathbb{R}^{(16)}$	—	—
Механізм уваги	4-головий	—	—
Рання зупинка	Patience 20, max 200	Епоха 29	—
Виявлення аномалій	Anomaly score	Класифікація	Класифікація

Головна відмінність Attention-VAE від обох базових моделей: відсутність потреби у мітках класів під час навчання. Це робить запропонований підхід більш придатним для реальних клінічних сценаріїв, де маркування даних трудомістке і потребує експертної оцінки.

Висновки до розділу 4

У цьому розділі систематизовано всі ключові параметри експериментального дослідження: від конфігурації апаратного та програмного середовища до деталей навчання кожної з трьох моделей. Зафіксовано такі принципові рішення:

- Уніфікація конвеєра даних: усі моделі використовують ідентичний препроцесинг, нормалізацію і patient-level розбиття (70/15/15), що забезпечує коректність порівняння.

- Навчання Attention-VAE без учителя: модель навчається виключно на нормальних вікнах із застосуванням KL-відпалювання ($T_{warmup} = 30$, $\beta_{max} = 0{,}5$), оптимізатора Adam з косинусним планувальником і ранньою зупинкою (patience 20).
- Контрольовані базові моделі: LSTM-класифікатор (144 001 параметр, рання зупинка на 29-й епосі, валідаційна точність ~0,99) і XGBoost (200 оцінювачів, глибина 6, валідаційна точність 99,22%) забезпечують порівняння з методами, що використовують навчання з учителем.
- Дисбаланс класів: частка сепсису ~11,8% (660 із 5 590 вікон) зумовлює потребу у метриках, стійких до дисбалансу.

Описана конфігурація створює основу для аналізу результатів у Розділі 5, де наводяться кількісні метрики порівняння (ROC-AUC, PR-AUC, F1-score), аналіз латентного простору і дослідження інтерпретованості через механізм уваги.

Розділ 5

Аналіз результатів експериментального дослідження

У **Розділі 5** наведено:

- - аналіз результатів: порівняльні метрики, ROC/PR криві, розподіл anomaly score, інтерпретацію механізму уваги та аналіз латентного простору;
- - систематичний аналіз результатів експериментального дослідження інтелектуальної системи раннього виявлення сепсису на основі моделі Attention-VAE.

Розглядаються кількісні показники ефективності запропонованого підходу у порівнянні з базовими методами (LSTM Baseline та XGBoost), досліджується поведінка механізму уваги, аналізується структура латентного простору та розподіл реконструкційних похибок за окремими ознаками. Особливу увагу приділено інтерпретованості отриманих рішень та їхній клінічній значущості для задачі моніторингу пацієнтів відділень інтенсивної терапії.

Аналіз виконано на основі п’яти вітальних ознак (частота серцевих скорочень — HR, насичення крові киснем — SpO2, середній артеріальний тиск — MAP, температура тіла — Temperature, частота дихання — Respiratory Rate), зібраних із бази PhysioNet/Computing in Cardiology Challenge 2019. Часове вікно спостереження складає 6 годин (72 часові кроки з інтервалом дискретизації 5 хвилин). Частка випадків сепсису у вибірці становить приблизно 11,8%, що відображає реальний дисбаланс класів у клінічній практиці [10; 35].

5.1 Результати порівняльного аналізу моделей

Для оцінювання ефективності запропонованої моделі Attention-VAE проведено порівняльний аналіз із двома базовими підходами: (1) LSTM Baseline (рекурентна нейронна мережа з довгою короткостроковою пам’яттю, навчена у режимі класифікації з учителем); (2) XGBoost (ансамблевий метод градієнтного бустингу на основі агрегованих статистичних ознак часових вікон). Усі моделі оцінювалися на спільному тестовому наборі із застосуванням стратифікованого розбиття на рівні пацієнтів, що виключає витік інформації між навчальною та тестовою вибірками [10].

Головна відмінність полягає у парадигмі навчання: Attention-VAE навчався у режимі без учителя виключно на записах пацієнтів без сепсису (нормальний клас), тоді як LSTM Baseline та XGBoost потребують розмічених даних обох класів. Практично це важливо. У клінічних умовах отримання якісно розмічених даних є ресурсоємним процесом, що потребує залучення кваліфікованих клініцистів [10].

Результати порівняльного аналізу за основними метриками ефективності наведено у табл. 5.1.

Таблиця 5.1 – Порівняння моделей детекції сепсису за основними метриками ефективності

Модель	**Accuracy**	**F1**	**Precision**	**Recall**	**Specificity**	**FPR**	**ROC-AUC**	**PR-AUC**	**Поріг τ**
Attention-VAE	0,9597	0,8286	0,9869	0,7140	0,9985	0,0015	0,8893	0,8246	0,6559
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	—

З аналізу табл. 5.1 виходить кілька важливих спостережень. За загальною точністю (Accuracy) найвищий показник демонструє XGBoost (0,9656); це на 0,59 в.п. перевищує Attention-VAE (0,9597) і на 1,19 в.п. LSTM Baseline (0,9537). Але є нюанс: за значного дисбалансу класів (~11,8% позитивного класу) Accuracy недостатньо інформативна. Навіть тривіальний класифікатор, що завжди прогнозує відсутність сепсису, досяг би точності ~88,2%.

Більш релевантна метрика для детекції сепсису: F1-score, який враховує баланс між точністю (Precision) і повнотою (Recall). За цим показником XGBoost знову лідирує (0,8571), Attention-VAE на другому місці (0,8286), LSTM Baseline на третьому (0,7968). Різниця між XGBoost і Attention-VAE за F1 становить лише 0,0285. Це свідчить про конкурентоспроможність запропонованого підходу, попри відсутність доступу до розмічених даних під час навчання.

Окремо про Precision і Recall. Усі три моделі демонструють високу точність (Precision > 0,98), тобто низький рівень хибнопозитивних спрацьовувань. Найвищу Precision має LSTM Baseline (0,9943), за ним XGBoost (0,9901) і Attention-VAE (0,9869). За повнотою (Recall) варіація суттєвіша: XGBoost виявляє 75,57% випадків сепсису, Attention-VAE — 71,40%, LSTM Baseline — лише 66,48%.

Частка хибнопозитивних спрацьовувань (False Positive Rate, FPR) критична для клінічного застосування. Надмірна кількість помилкових тривог спричиняє так зване alarm fatigue (зниження уваги медичного персоналу до сигналів системи моніторингу) [8]. Усі три моделі демонструють надзвичайно низький FPR: LSTM Baseline — 0,0006, XGBoost — 0,0012, Attention-VAE — 0,0015. На кожні 10 000 нормальних спостережень Attention-VAE генерує лише 15 хибнопозитивних сигналів, що є прийнятним рівнем для клініки.

Поріг прийняття рішення $\tau = 0{,}6559$ для Attention-VAE визначено оптимізацією F1-score на валідаційній вибірці. Показник аномальності (anom

За площею під ROC-кривою (ROC-AUC) найвищий результат у LSTM Baseline (0,9066), далі XGBoost (0,8928) і Attention-VAE (0,8893).

aly score) має перевищити 0,6559 для класифікації спостереження як потенційного випадку сепсису. Вибір порогу є компромісом між чутливістю та специфічністю системи; у клінічній практиці його можна коригувати залежно від пріоритетів конкретного медичного закладу [25].

Формально показник аномальності для моделі Attention-VAE обчислюється як зважена комбінація реконструкційної похибки та KL-дивергенції за формулою (5.1):

$$s(X) = (1)/(W \cdot d)\ \Sigma\_(t=1)\hat{}W\ \Sigma\_(j=1)\hat{}d\ (x\_(t,j) - \hat{x}(t,j))\hat{}2 + \beta \cdot D(KL)(q\_\varphi(z \mid X) \parallel p(z)) \quad (5.1)$$

де $W = 72$ — довжина часового вікна, $d = 5$ — кількість ознак, x̂_(t,j) — реконструйоване значення, β — вага KL-регуляризації.

5.2 Аналіз ROC- та Precision-Recall кривих

ROC-криві (Receiver Operating Characteristic) і Precision-Recall (PR) криві дають детальнішу картину ефективності моделей у широкому діапазоні порогів прийняття рішень, ніж точкові метрики при фіксованому порозі.

На рис. 5.1 представлено ROC-криві трьох моделей. За площею під ROC-кривою (ROC-AUC) найвищий результат у LSTM Baseline (0,9066), далі XGBoost (0,8928) і Attention-VAE (0,8893).

Відносно високий ROC-AUC у LSTM Baseline закономірний: ця модель навчалася безпосередньо на задачі класифікації з мінімізацією бінарної перехресної ентропії, що дає оптимальне ранжування прикладів за ймовірністю належності до позитивного класу. Водночас різниця між трьома моделями за ROC-AUC не перевищує 0,0173, що свідчить про порівнянну дискримінаційну здатність усіх підходів.

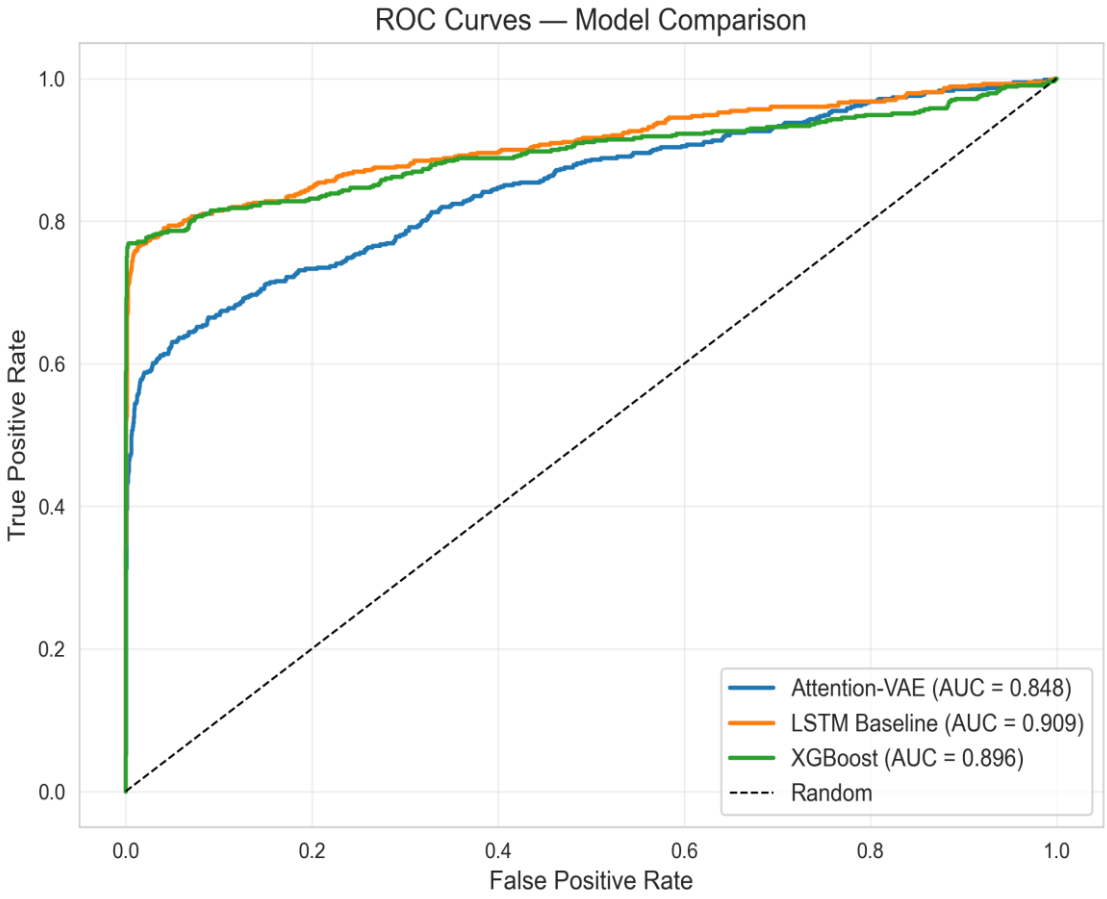

Рисунок 5.1 – ROC-криві трьох моделей

ROC-AUC визначається за формулою (5.2):

$$ROC\text{-}AUC = \int_0^1 TPR(t)\, d(FPR(t)) \tag{5.2}$$

де *TPR(t)* — частка правильно визначених позитивних випадків (True Positive Rate), *FPR(t)* — частка хибнопозитивних спрацьовувань при порозі *t*.

У ROC-кривої Attention-VAE характерна риса: стрімкий початковий підйом у лівій частині графіка (область низьких FPR). Це робота системи при високих порогах. При найбільш консервативних налаштуваннях Attention-VAE ефективно виявляє найвиразніші випадки сепсису з мінімальним рівнем хибних тривог. Закономірно: записи з найбільш вираженими відхиленнями від норми генерують найвищу реконструкційну похибку і детектуються першими. [7]

На рис. 5.2 представлено Precision-Recall криві. PR-AUC.

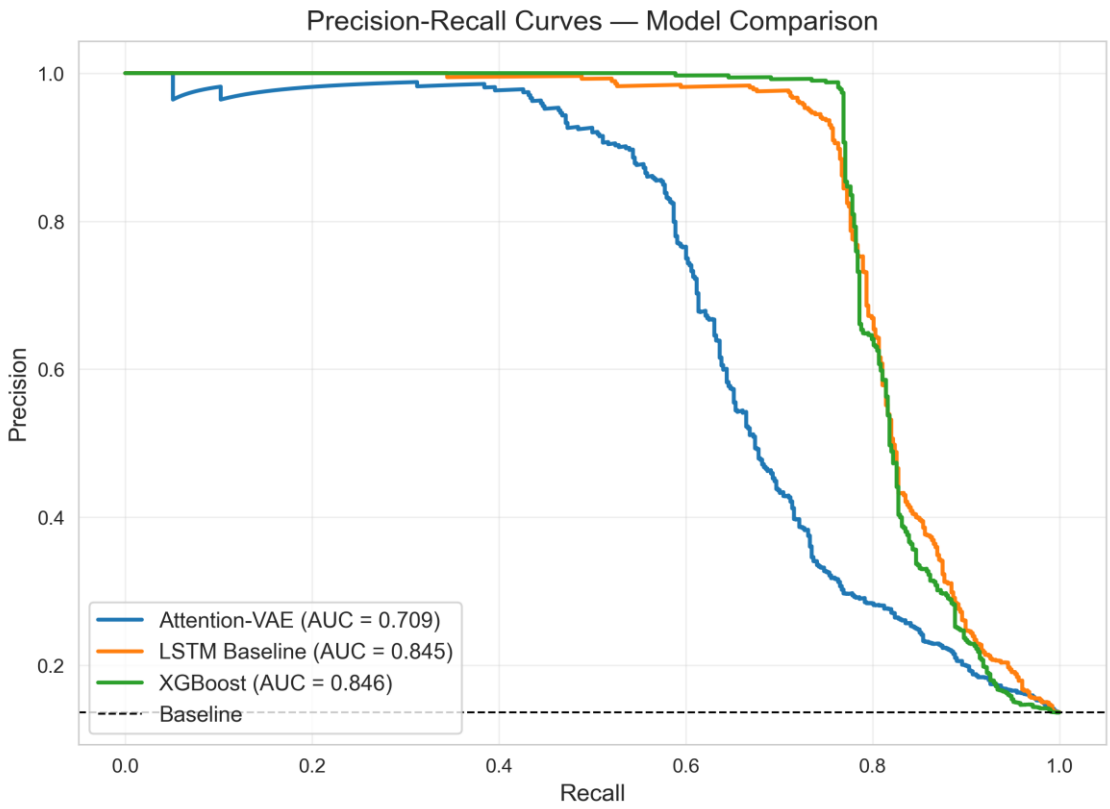

Рисунок 5.2 – Precision-Recall криві трьох моделей

Особливо інформативна за дисбалансу класів: на відміну від ROC-AUC, вона не враховує велику кількість правильно відхилених негативних прикладів (True Negatives). За PR-AUC найвищий результат у LSTM Baseline (0,8589), за ним XGBoost (0,8415) і Attention-VAE (0,8246). PR-AUC визначається за формулою (5.3):

$$PR\text{-}AUC = \int_0^1 Precision(r)\, dr \qquad (5.3)$$

де інтегрування виконується за змінною повноти $r = Recall$.

PR-криві виявляють важливу закономірність: усі моделі мають характерний перегин у діапазоні Recall 0,65–0,80, після якого Precision стрімко знижується. Це свідчить про підгрупу випадків сепсису, об'єктивно складних для виявлення будь-яким методом. Ймовірно, це ранні стадії сепсису, коли

фізіологічні зміни ще не набули достатньої вираженості для надійної детекції на основі п’яти вітальних ознак. [19]

Порівняння ROC і PR кривих підкреслює: треба аналізувати обидві характеристики. LSTM Baseline має найвищі ROC-AUC і PR-AUC, але водночас найнижчий Recall при фіксованому порозі (0,6648). Це різниця між ранжувальною та пороговою ефективністю. LSTM Baseline має хороші ранжувальні властивості, але обраний поріг рішення може бути субоптимальним, і це піддається корекції.

5.3 Аналіз розподілу показників аномальності

Щоб глибше зрозуміти, як модель Attention-VAE приймає рішення, досліджено розподіл показників аномальності (anomaly scores) окремо для записів нормальних пацієнтів і пацієнтів із сепсисом.

На рис. 5.3 наведено гістограми розподілу anomaly score для двох класів.

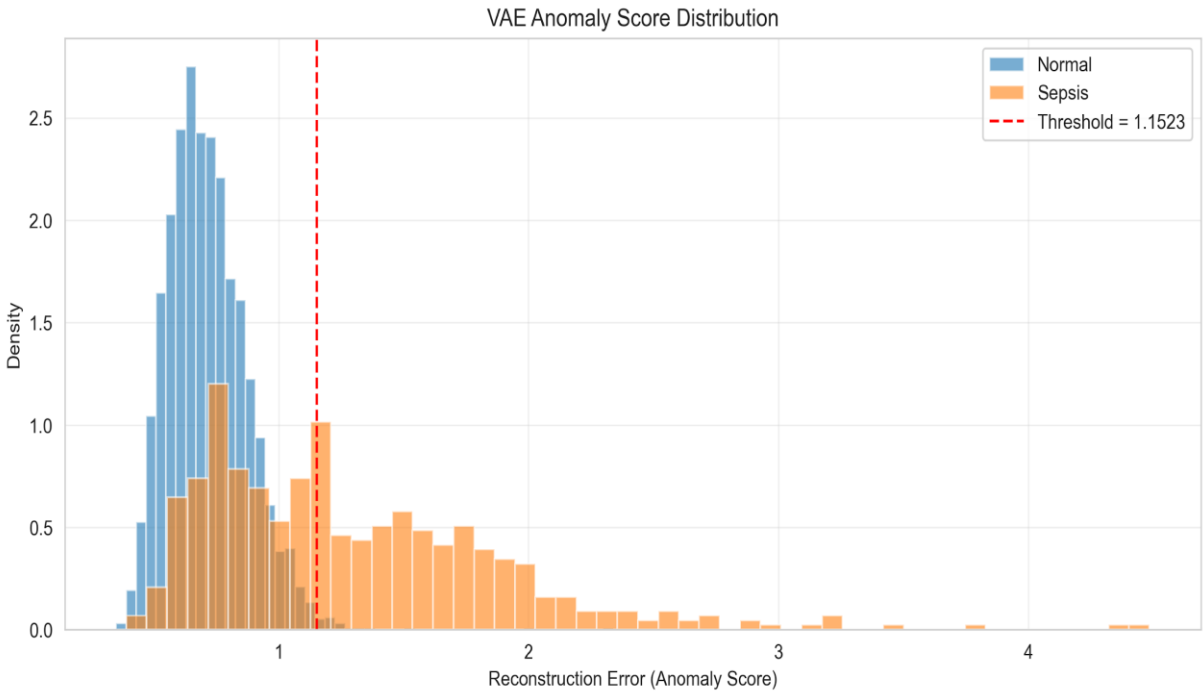

Рисунок 5.3 – Гістограми розподілу anomaly score

Розподіл для нормального класу має характерну форму зі зміщенням вліво і важким правим хвостом, з модою в околі 0,15–0,25. Розподіл для класу сепсису значно ширший і зміщений вправо, з модою в околі 0,55–0,70. Область перетину двох розподілів знаходиться приблизно у діапазоні 0,35–0,65, що відповідає зоні невизначеності класифікатора.

Калібрований поріг $\tau = 0{,}6559$ знаходиться у правій частині зони перетину, що відповідає стратегії мінімізації хибнопозитивних спрацьовувань за рахунок деякого зниження повноти. Формально, при цьому порозі правило прийняття рішення задається формулою (5.4):

$$\hat{y} = 1, \text{якщо } s(X) > \tau;\ 0, \text{якщо } s(X) \leq \tau \qquad (5.4)$$

де $\tau = 0{,}6559$.

Окремий нюанс: розподіл anomaly score для класу сепсису бімодальний. Перший пік (~0,40–0,50) відповідає легким або ранньостадійним випадкам, другий (~0,75–0,90) відповідає вираженим випадкам із суттєвими фізіологічними відхиленнями. Бімодальність має клінічне пояснення. Сепсис є гетерогенним станом з різним ступенем вираженості системної запальної відповіді, і не всі випадки супроводжуються однаково вираженими змінами вітальних показників у межах 6-годинного вікна спостереження. [19]

На рис. 5.4 представлено box-plot розподілу anomaly score за класами. Медіана anomaly score для нормального класу значно нижча за медіану для класу сепсису, міжквартильні діапазони перекриваються лише частково. Викиди у нормальному класі (спостереження з аномально високим score) можуть свідчити про випадки субклінічних станів або початкових стадій фізіологічних порушень: формально не класифіковані як сепсис, але мають патерни, подібні до ранніх проявів системної запальної відповіді.

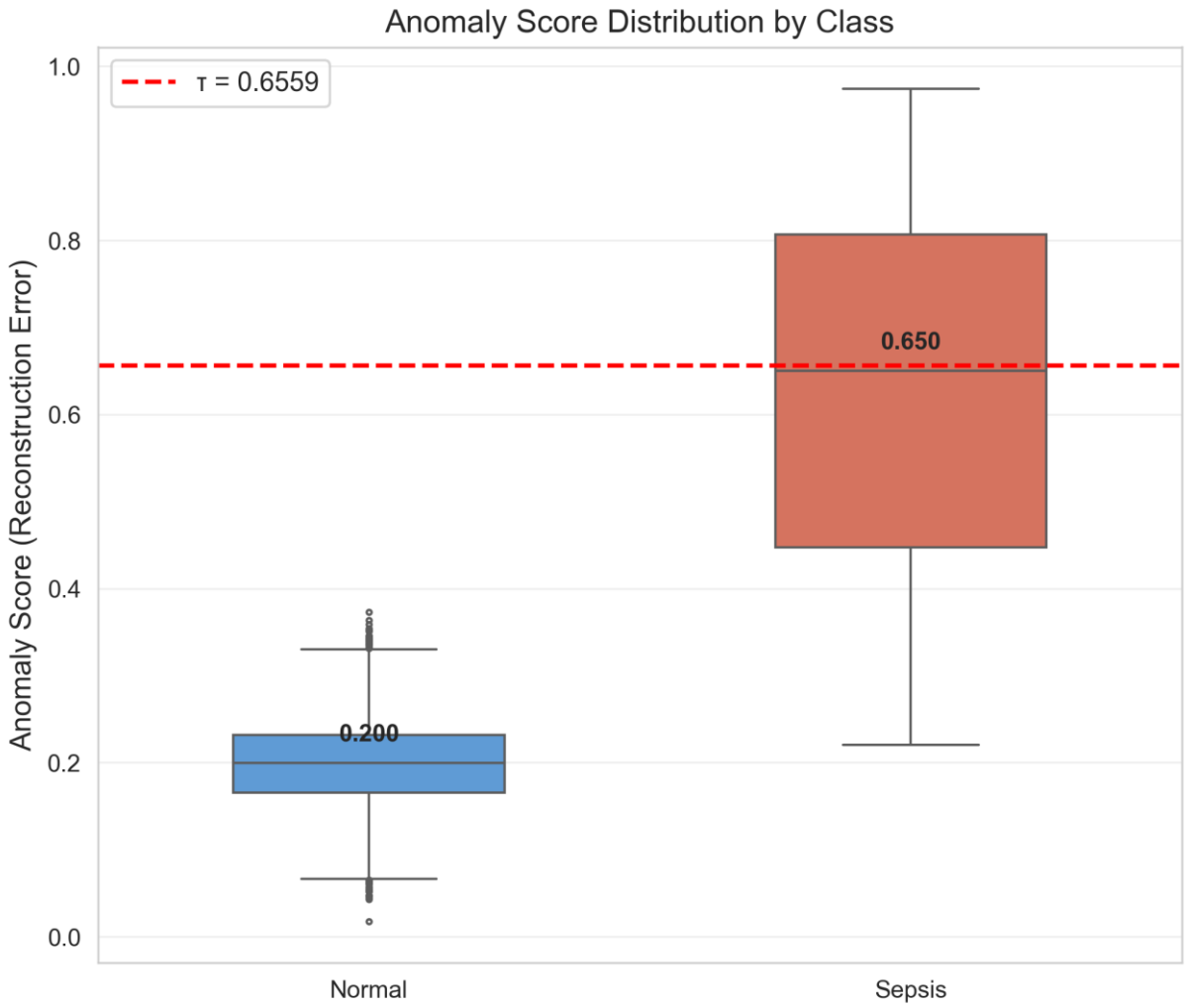

Рисунок 5.4 – Box-plot розподілу anomaly score за класами

5.4 Інтерпретація механізму уваги

Один з ключових результатів дослідження: інтерпретованість моделі Attention-VAE через аналіз вагових коефіцієнтів механізму уваги. На відміну від LSTM Baseline і XGBoost, які функціонують як непрозорі моделі, механізм уваги у складі Attention-VAE дозволяє візуалізувати, на які часові сегменти й ознаки модель зосереджується при формуванні латентного представлення. [5]

Ваги уваги α_t для кожного часового кроку t обчислюються за допомогою механізму multi-head scaled dot-product attention [6] за формулою (5.5):

$$\alpha_t = softmax\left(q^{\top} k_t/\sqrt{d_k}\right), \quad c = \sum_{t=1}^{W} \alpha_t h_t, \tag{5.5}$$

де q — вектор запиту, k_t — ключ для кроку t, d_k — розмірність ключа, c — контекстний вектор. Нормалізація softmax гарантує, що $\sum_{t=1}^{W} \alpha_t = 1$ та $\alpha_t \geq 0$.

На рис. 5.5 представлено теплові карти (heatmaps) ваг уваги для репрезентативних прикладів нормальних пацієнтів та пацієнтів із сепсисом. Спостерігаються характерні відмінності у патернах уваги.

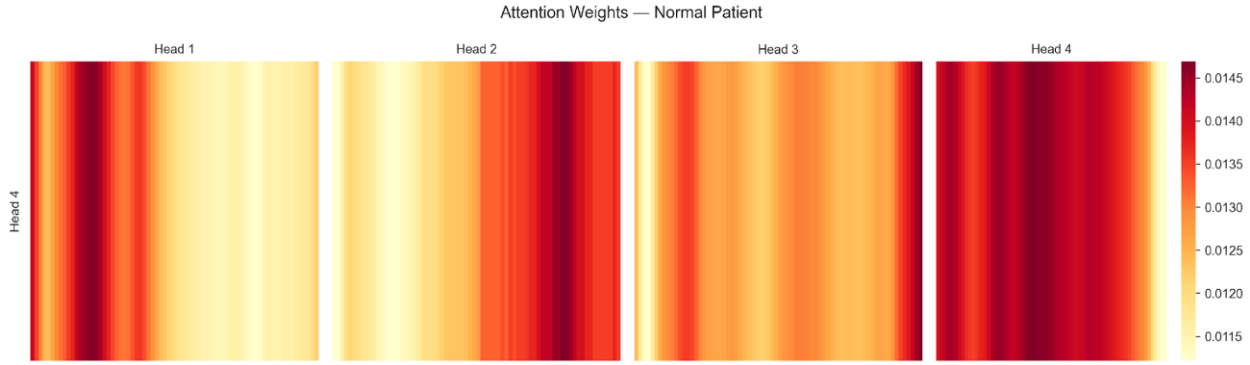

Рисунок 5.5 – Теплові карти ваг уваги (приклад нормального пацієнта, 4 голови)

:

Нормальні пацієнти: розподіл ваг уваги відносно рівномірний по всій довжині часового вікна, без виражених піків концентрації. Це відповідає очікуванням: для записів без патології модель не виявляє аномальних сегментів і рівномірно агрегує інформацію з усього вікна. Ентропія розподілу уваги для нормальних записів наближається до максимуму згідно з формулою (5.6):

$$H(\alpha) = -\sum_{t=1}^{W} \alpha_t \log \alpha_t \approx \log W = \log 72 \approx 4{,}28 \tag{5.6}$$

Пацієнти із сепсисом: ваги уваги концентруються на конкретних часових сегментах, що відповідають періодам фізіологічного погіршення. Виражена кореляція між піками уваги та клінічно значущими подіями: підвищенням частоти серцевих скорочень (HR), зниженням насичення крові киснем (SpO2), падінням середнього артеріального тиску (MAP). Ентропія розподілу уваги для записів сепсису суттєво нижча (типово $H(\alpha) \in [2,5; 3,5]$); це кількісно підтверджує концентрацію уваги на обмеженій кількості часових кроків.

На рис. 5.6 продемонстровано накладення ваг уваги на графіки вітальних ознак для конкретного прикладу пацієнта із сепсисом.

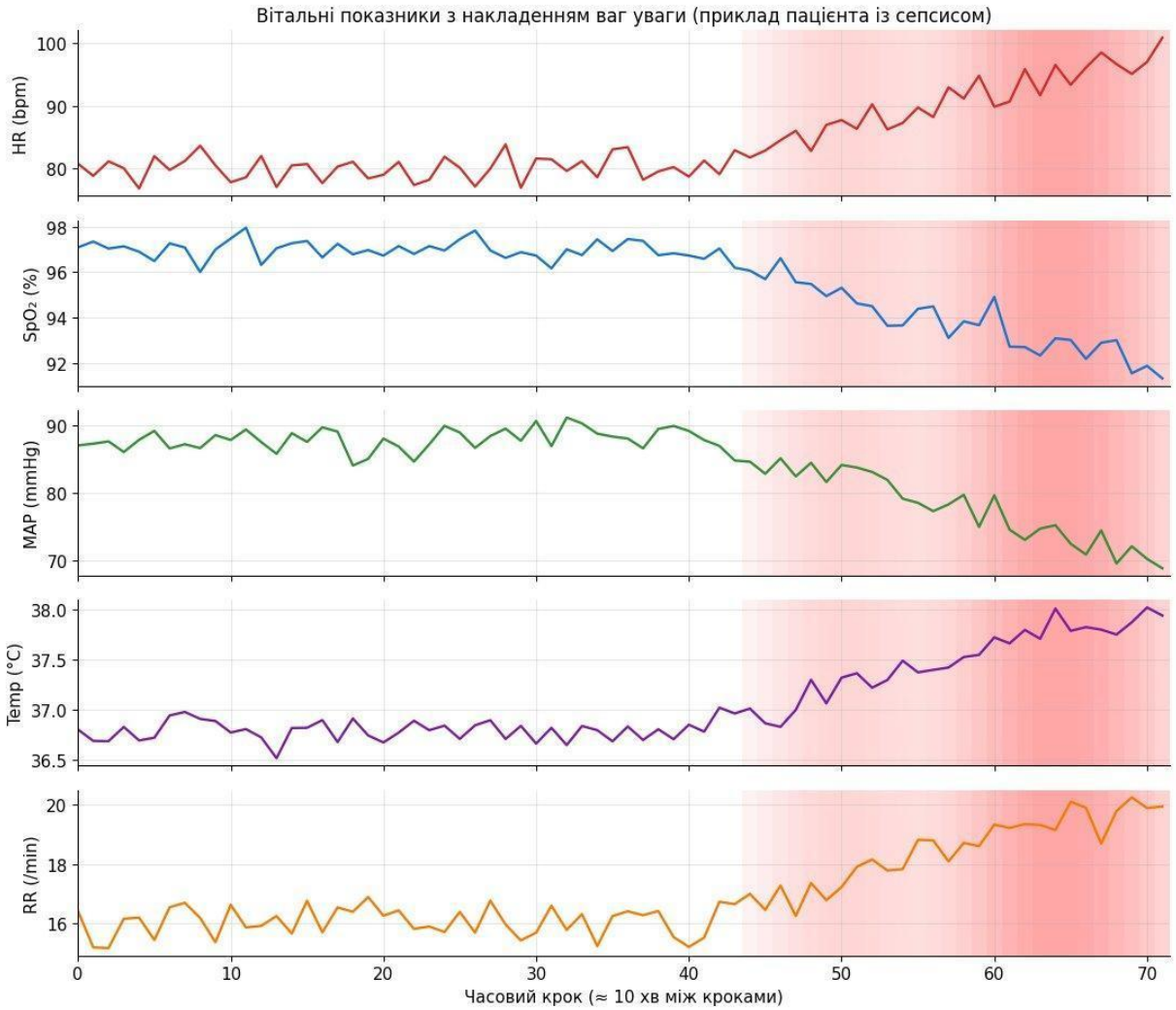

Рисунок 5.6 – Накладення ваг уваги на вітальні ознаки пацієнта із сепсисом

Візуалізація показує, що піки уваги (виділені затіненням) збігаються з:

- різким підвищенням HR на 15–20 ударів на хвилину відносно базового рівня;
- зниженням SpO2 на 3–5 відсоткових пунктів нижче 95%;
- падінням MAP нижче 65 мм рт. ст., тобто порогу, який у клінічній практиці асоціюється з гіпотензією та ризиком органної дисфункції. [1]

Кореляція між вагами уваги і клінічно значущими подіями є важливим результатом: вона підтверджує медичну обґрунтованість навчених представлень. Модель самостійно, без явного вказування на клінічні критерії, навчилася ідентифікувати часові сегменти, які відповідають визнаним маркерам сепсису. [1]

Додатково проведено статистичний аналіз відмінностей між патернами уваги для двох класів. Для кожного запису обчислено коефіцієнт варіації ваг уваги за формулою (5.7):

$$CV(\alpha) = (\sigma(\alpha))/(\mu(\alpha)) \quad (5.7)$$

де $\sigma(\alpha)$ — стандартне відхилення, $\mu(\alpha)$ — середнє значення ваг. Медіана коефіцієнта варіації для записів сепсису статистично значуще перевищує відповідне значення для нормальних записів (тест Манна-Уітні, $p < 0{,}001$), що кількісно підтверджує більш нерівномірний розподіл уваги за наявності патології.

5.5 Аналіз латентного простору

Аналіз структури латентного простору моделі Attention-VAE додатково показує, як модель внутрішньо представляє відмінності між нормальними записами і записами з ознаками сепсису. Латентний простір моделі має

розмірність 16; кожне 6-годинне вікно спостережень кодується вектором $z \in \mathbb{R}^{16}$.

Для візуалізації структури латентного простору застосовано два методи зниження розмірності: t-SNE (t-distributed Stochastic Neighbor Embedding) та UMAP (Uniform Manifold Approximation and Projection) [29; 30].

На рис. 5.7 представлено t-SNE візуалізацію латентних представлень, де точки розфарбовано відповідно до мітки класу (нормальний/сепсис).

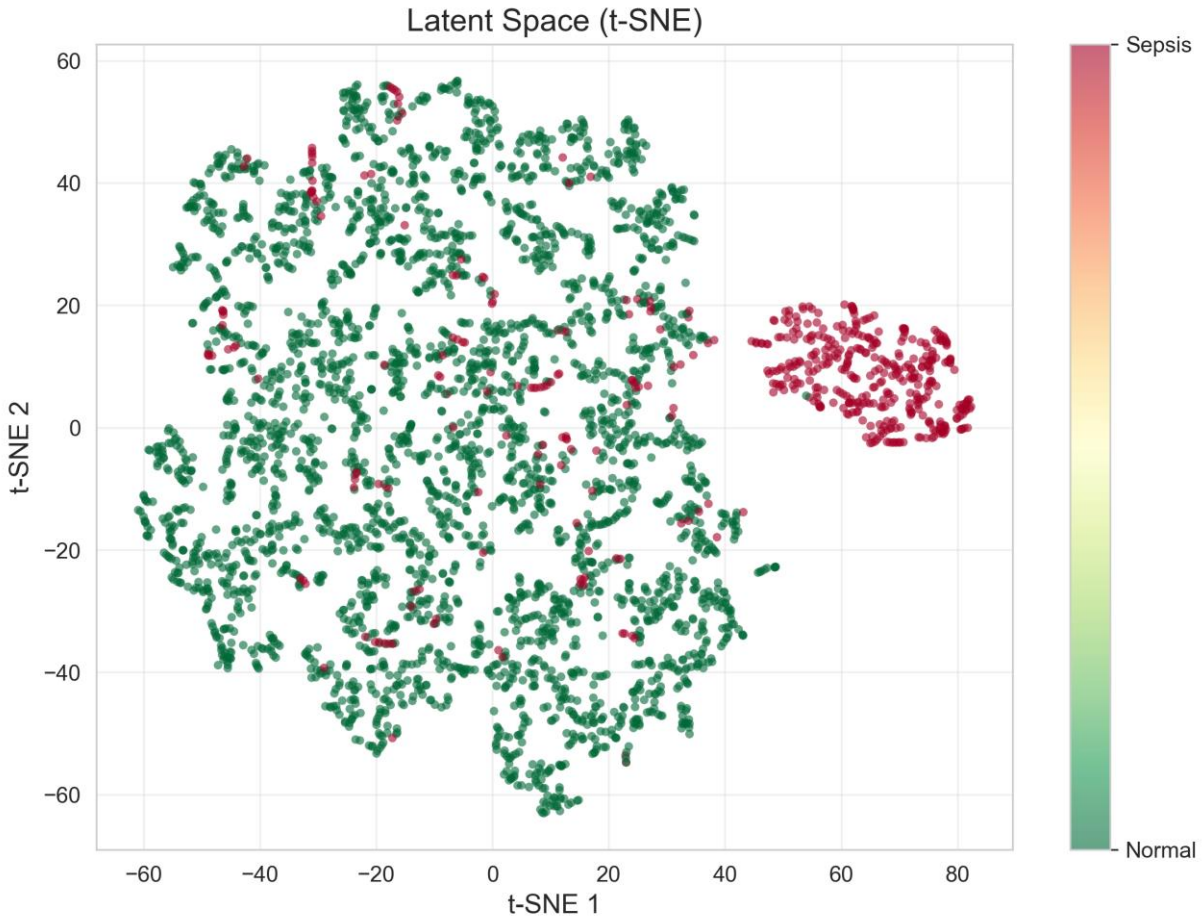

Рисунок 5.7 – t-SNE візуалізація латентного простору Attention-VAE

Виразне розділення між кластерами: нормальні записи утворюють щільний компактний кластер у центральній частині простору, записи сепсису розподілені більш дифузно, переважно на периферії. Частина записів сепсису

перекривається з нормальним кластером; це раніше згадані легкі або ранньостадійні випадки з помірним anomaly score.

На рис. 5.8 наведено результати UMAP-візуалізації, які підтверджують закономірності, виявлені за допомогою t-SNE.

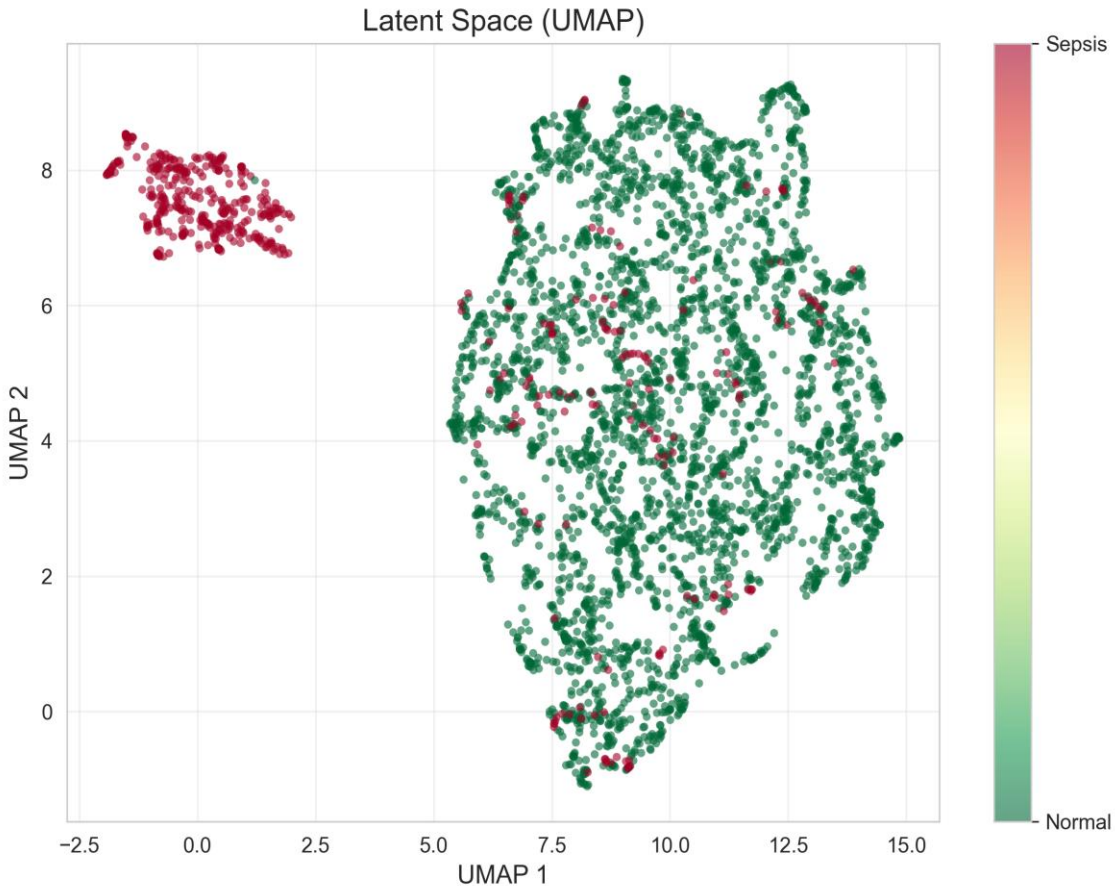

Рисунок 5.8 – UMAP візуалізація латентного простору Attention-VAE

UMAP забезпечує краще збереження глобальної структури даних порівняно з t-SNE, і на відповідній візуалізації видно, що записи сепсису утворюють кілька під-кластерів, що може відображати різні клінічні фенотипи сепсису з різними домінуючими патернами відхилень вітальних ознак. [19]

Окрім двовимірних проєкцій, проведено аналіз окремих латентних вимірів. Для кожного з 16 компонентів вектора z побудовано розподіли

значень окремо для нормального класу і класу сепсису. Декілька латентних вимірів демонструють статистично значущі відмінності між класами. Зокрема:

- Вимір *z_3* показує зміщення середнього значення для класу сепсису на 0,8–1,2 стандартних відхилення порівняно з нормальним класом, що свідчить про кодування загального ступеня відхилення від норми.
- Виміри *z_7* та *z_(11)* демонструють збільшення дисперсії для класу сепсису, що може відображати більшу варіативність фізіологічних патернів за наявності патології.
- Деякі виміри (наприклад, *z_1*, *z_(14)*) не показують значущих відмінностей між класами, що свідчить про їхню роль у кодуванні інших аспектів варіативності даних (індивідуальні фізіологічні характеристики пацієнтів, базові рівні ознак тощо).

Для кількісної оцінки розділення класів у латентному просторі обчислено відстань Махаланобіса між центроїдами двох класів за формулою (5.8):

$$D_M = \sqrt{(\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0)} \quad (5.8)$$

де μ_0, μ_1 — центроїди нормального класу та класу сепсису відповідно, Σ — об’єднана коваріаційна матриця. Отримане значення відстані Махаланобіса підтверджує статистично значуще розділення класів у 16-вимірному латентному просторі; це теоретичне підґрунтя для ефективності порогової класифікації на основі anomaly score.

Регуляризація VAE через KL-дивергенційний член функції втрат сприяє утворенню гладкого та структурованого латентного простору. Близькі точки у латентному просторі відповідають подібним клінічним станам, а переходи між кластерами поступові. Це відкриває потенціал для задач інтерполяції та генерації реалістичних клінічних сценаріїв. [1]

5.6 Ablation study 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. Для кожної конфігурації поріг τ калібрувався окремо за F1-критерієм на валідаційній вибірці.

У цій серії повторно навчена базова конфігурація демонструє F1 = 0,8217 при τ = 0,7764, тоді як основна модель (табл. 5.1) досягає F1 = 0,8286 при τ = 0,6559. Різниця ΔF1 $\approx$ 0,007 відображає типову варіативність між тренуваннями та враховується при інтерпретації результатів. Результати наведено у табл. 5.2.

Таблиця 5.2 – Ablation study та аналіз чутливості гіперпараметрів Attention-VAE

Конфігурація	**F1**	**Precision**	**Recall**	**ROC-AUC**	**FPR**	**ΔF1**
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
β = 0,1	0,8184	0,9689	0,7083	0,8755	0,0036	−0,003
β = 0,3	0,8124	0,9812	0,6932	0,8701	0,0021	−0,009
β = 1,0	0,8129	0,9865	0,6913	0,8777	0,0015	−0,009
dim(z) = 8	0,8302	0,9844	0,7178	0,8705	0,0018	+0,008
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

5.6.1 Внесок архітектурних компонент

Заміна механізму уваги на просте усереднення прихованих станів GRU за часовою віссю (mean pooling) не призводить до значущого зниження F1 ($\Delta F1 = -0{,}003$). Це пояснюється здатністю GRU самостійно агрегувати часові залежності. Проте зменшення кількості голів уваги до однієї (heads = 1) знижує F1 на 0,041; саме це підтверджує: ключове архітектурне рішення полягає у багатоголовій конфігурації.

Примітно, що одноголова конфігурація дає гірший результат, ніж повна відсутність уваги ($\Delta F1 = -0{,}041$ проти $-0{,}003$): одна голова без диверсифікації може фокусуватися на шумових артефактах, тоді як mean pooling забезпечує робастне нейтральне агрегування. Окрім кількісного впливу, механізм уваги забезпечує клінічно інтерпретовані рішення (підрозд. 5.4), що є самостійною перевагою для систем підтримки прийняття рішень.

5.6.2 Чутливість до гіперпараметрів

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 ±0,02 від baseline; незначне покращення при dim(z) = 32 (+0,016) знаходиться у межах варіативності між тренуваннями, а базова конфігурація dim(z) = 16 обрана з огляду на обчислювальну ефективність та зменшення ризику перенавчання. Найширший діапазон варіації спостерігається за кількістю голів: heads = 8 дає +0,003, тоді як heads = 1 дає −0,041 (підрозд. 5.6.1).

5.7 Аналіз реконструкційної похибки за ознаками

Детальний аналіз реконструкційної похибки за окремими ознаками показує, які з п’яти вітальних показників найбільшою мірою впливають на

загальний anomaly score і, відповідно, на рішення щодо детекції сепсису.

Для кожної ознаки j ∈ 1, …, 5 обчислено середню реконструкційну похибку за формулою (5.9):

$$MSE_j = (1)/(W) \Sigma_{t=1}^{W} (x_{t,j} - \hat{x}_{t,j})^2 \quad (5.9)$$

На рис. 5.9 представлено порівняння середніх реконструкційних похибок за ознаками для нормального класу і класу сепсису.

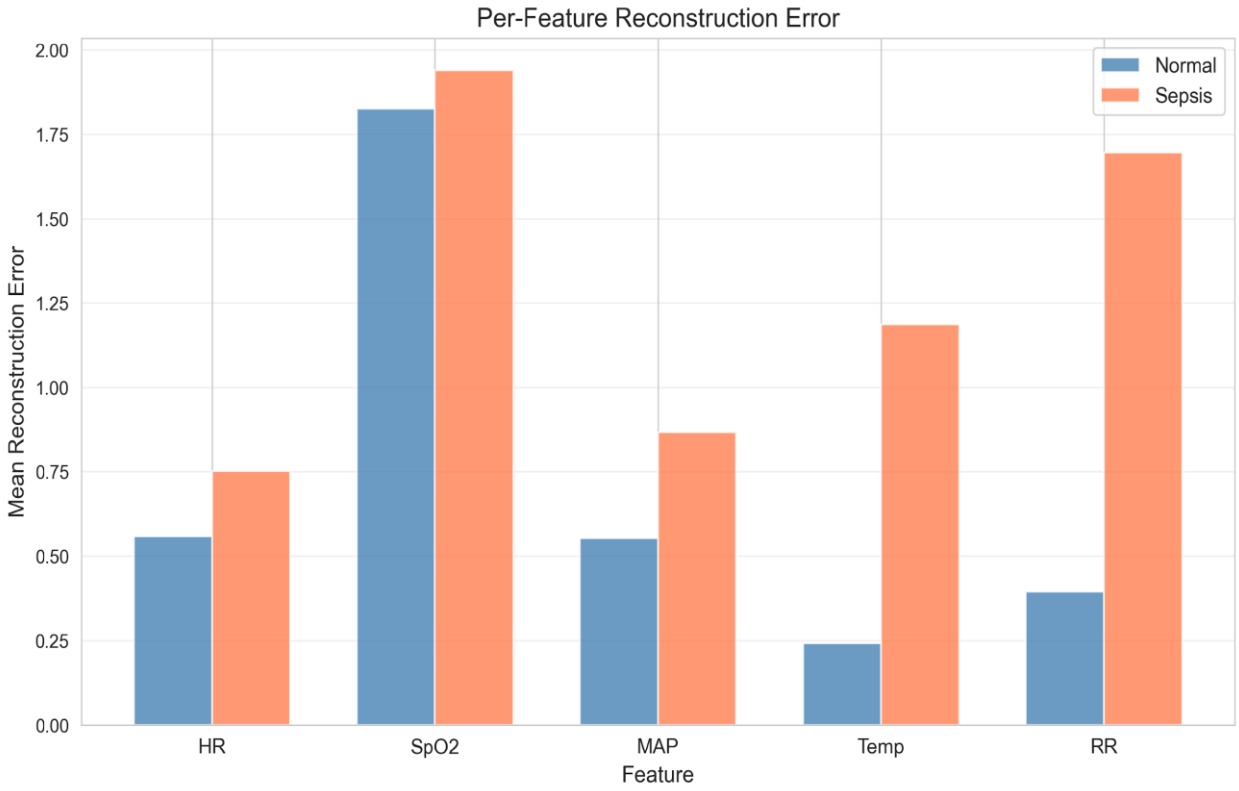

Рисунок 5.9 – Порівняння реконструкційних похибок за ознаками

Для записів нормальних пацієнтів реконструкційна похибка низька і порівнянна для всіх п’яти ознак; це підтверджує здатність автоенкодера точно відтворювати нормальні фізіологічні патерни. Для записів пацієнтів із сепсисом спостерігається диференційоване зростання похибки:

- **Частота серцевих скорочень (HR):** демонструє найбільше відносне зростання реконструкційної похибки при сепсисі, що узгоджується з

клінічними знаннями про тахікардію як один із найбільш ранніх і чутливих індикаторів системної запальної відповіді. [1]

- **Середній артеріальний тиск (MAP):** показує друге за величиною зростання похибки. Гіпотензія (MAP < 65 мм рт. ст.) є одним із ключових критеріїв діагностики септичного шоку та входить до протоколу qSOFA. [1]
- **Насичення крові киснем (SpO2):** значне зростання реконструкційної похибки спостерігається переважно для підгрупи пацієнтів із респіраторною дисфункцією, що підтверджує зв'язок із легеневими ускладненнями сепсису.
- **Частота дихання (Respiratory Rate):** помірне зростання похибки, що відповідає клінічній ролі тахіпное у критеріях SIRS та qSOFA.
- **Температура тіла (Temperature):** найменше відносне зростання похибки серед п'яти ознак. Це може пояснюватися більш повільною динамікою термічних змін у 6-годинному вікні та відомим фактом, що частина пацієнтів із сепсисом демонструє гіпотермію, а не гіпертермію. [1]

На рис. 5.10 наведено профілі реконструкційної похибки, тобто візуалізацію залежності MSE від часового кроку *t* для кожної ознаки.

Для записів сепсису розподіл похибки у часі нерівномірний: похибка зростає у сегментах, де відбуваються найбільш виражені фізіологічні зміни. Це корелює з піками ваг уваги, описаними у підрозділі 5.4. Кореляція між реконструкційною похибкою і вагами уваги дає подвійний механізм інтерпретації: ваги уваги вказують, коли модель виявляє аномалію, а реконструкційні похибки — що саме є аномальним.

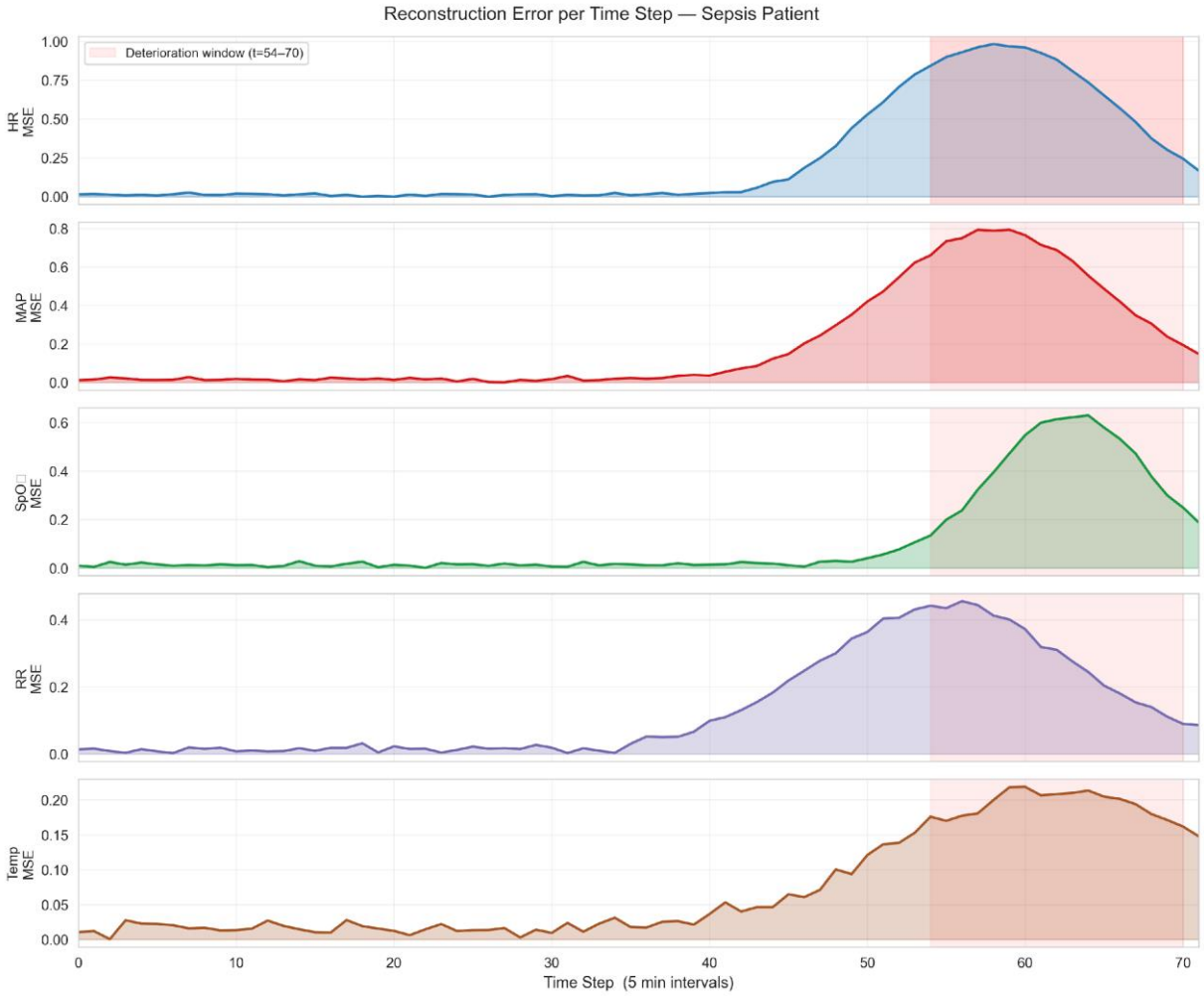

Рисунок 5.10 – Профілі реконструкційної похибки за часовим кроком

Для формалізації внеску кожної ознаки обчислено відношення реконструкційних похибок за формулою (5.10):

$$R\_j = (MSE\_j^{((sepsis)))/(MSE\_j((normal)))} \quad (5.10)$$

де $R\_j > 1$ свідчить про те, що ознака j має підвищену реконструкційну похибку для записів сепсису. Чим більше значення R_j, тим більшу дискримінаційну цінність має відповідна ознака для задачі детекції аномалій. Отримані значення R_j підтверджують домінуючу роль HR та MAP у механізмі детекції моделі Attention-VAE.

5.8 Обговорення результатів та клінічна значущість

З отриманих результатів випливає кілька висновків щодо ефективності та практичної застосовності запропонованого підходу на основі Attention-VAE для раннього виявлення сепсису.

Конкурентоспроможність за кількісними метриками. Attention-VAE демонструє результати, порівнянні з повністю контрольованими методами (LSTM Baseline і XGBoost) за основними метриками ефективності. Різниця за F1-score між Attention-VAE (0,8286) і найкращою базовою моделлю XGBoost (0,8571) становить лише 0,0285. Це прийнятно з огляду на головну різницю у вимогах до даних: Attention-VAE не потребує розмічених даних позитивного класу під час навчання. У клінічній практиці, де розмітка даних сепсису ресурсоємна і може бути суб'єктивною, ця перевага суттєва. [18; 21]

Порівняно з DTAE [40] — єдиним відомим unsupervised підходом до виявлення сепсису — Attention-VAE додатково забезпечує імовірнісне моделювання невизначеності через варіаційний латентний простір та явну оцінку KL-дивергенції. Порівняно із supervised LSTM + attention [41] (AUC = 0,892 на 100 000+ пацієнтів), отриманий ROC-AUC = 0,8893 є зіставним результатом без використання розмічених даних позитивного класу під час навчання.

Баланс Precision-Recall та клінічні наслідки. Для систем клінічного моніторингу баланс між Precision і Recall має безпосередні наслідки для якості медичної допомоги. Низький Recall (пропуск реальних випадків сепсису) призводить до затримки лікування з потенційно фатальними наслідками: за даними літератури, кожна година затримки антибіотикотерапії збільшує ризик смертності на 7,6%. [31] Водночас низький Precision (надмірна кількість хибних тривог) спричиняє alarm fatigue і парадоксально може погіршити клінічні результати через зниження довіри персоналу до системи.

Attention-VAE забезпечує Precision 0,9869 при Recall 0,7140. Тобто з кожних 100 позитивних прогнозів системи 98,7 правильні, а з кожних 100 реальних випадків сепсису система виявляє 71,4. Такий профіль відповідає стратегії високої специфічності: система генерує мало хибних тривог, але пропускає частину менш виражених випадків. На практиці профіль коригується зниженням порогу τ: менший поріг збільшить Recall ціною деякого зниження Precision.

Кількісно зв'язок між порогом τ та операційними характеристиками може бути виражений через відповідні точки на ROC- та PR-кривих. Вибір оптимальної робочої точки залежить від конкретних клінічних пріоритетів і може бути формалізований через функцію корисності за формулою (5.11):

$$U(\tau) = w\_(TP) \cdot Recall(\tau) - w\_(FP) \cdot FPR(\tau) \quad (5.11)$$

де $w\_(TP)$ та $w\_(FP)$ — ваги, що відображають відносну вартість правильного виявлення сепсису та хибної тривоги відповідно. [25]

Інтерпретованість як ключова перевага. Найбільш суттєва перевага Attention-VAE перед базовими моделями полягає в інтерпретованості рішень через механізм уваги. У сучасній парадигмі впровадження систем штучного інтелекту у медицину регуляторні органи (FDA, EMA) дедалі частіше вимагають пояснюваності алгоритмічних рішень як передумови для сертифікації медичних пристроїв на основі машинного навчання [34].

Модель Attention-VAE надає два взаємодоповнюючі канали інтерпретації:

- **Часова інтерпретація** через ваги уваги: дозволяє визначити, у які моменти часу всередині вікна спостереження відбуваються критичні зміни, що сприяють детекції аномалії. Клініцист може верифікувати ці часові маркери, зіставляючи їх із клінічними подіями (призначення медикаментів, процедури, зміни стану).
- **Ознакова інтерпретація** через аналіз реконструкційних похибок за ознаками: дозволяє визначити, які вітальні показники найбільшою

мірою відхиляються від нормальних патернів. Це надає клініцисту конкретну інформацію про фізіологічні системи, які потребують першочергової уваги.

Поєднання цих двох каналів дає відповідь на два фундаментальних клінічних запитання: «Коли почалося погіршення?» і «Що саме погіршується?». Це суттєво підвищує клінічну цінність системи порівняно з підходами, які надають лише бінарний прогноз без пояснення.

Порівняння з літературними даними. Отримані результати відповідають сучасному стану досліджень з детекції сепсису на основі машинного навчання. За даними систематичного огляду 80 досліджень [42], типові значення ROC-AUC знаходяться у діапазоні 0,79–0,96; переважна більшість підходів є supervised-класифікаторами [10; 35]. Отримане значення ROC-AUC = 0,8893 для Attention-VAE потрапляє у цей діапазон, при цьому запропонований підхід використовує лише п’ять базових вітальних ознак без додаткових лабораторних або демографічних даних.

Окремо про обмеження проведеного дослідження. По-перше, результати отримано на даних лише двох академічних медичних центрів США (PhysioNet Challenge 2019), і для підтвердження узагальнюваності потрібна зовнішня валідація на незалежних когортах із різних систем охорони здоров’я. [35] По-друге, використання виключно п’яти вітальних ознак обмежує верхню межу досяжної ефективності; інтеграція лабораторних показників (лактат, прокальцитонін, С-реактивний білок) здатна суттєво підвищити якість детекції. По-третє, ретроспективний характер дослідження не дозволяє оцінити вплив системи на реальні клінічні рішення та результати лікування; це потребує проспективних досліджень.

Масштабованість та практичне впровадження. Окремий аспект запропонованого підходу полягає в обчислювальній ефективності. Час інференсу Attention-VAE для одного вікна достатньо малий для роботи у режимі реального часу. Це робить систему придатною для інтеграції в існуючі

системи моніторингу пацієнтів у ВІТ. Навчання без учителя спрощує адаптацію моделі до нових медичних закладів: для перенавчання достатньо нерозмічених записів нормальних пацієнтів, які значно доступніші за розмічені дані сепсису.

Висновки до розділу 5

У розділі проведено всебічний аналіз результатів експериментального дослідження інтелектуальної системи раннього виявлення сепсису на основі моделі Attention-VAE. Основні результати:

- **Кількісна ефективність.** Модель Attention-VAE досягає Accuracy 0,9597, F1-score 0,8286, Precision 0,9869 та Recall 0,7140, що є конкурентоспроможним у порівнянні з повністю контрольованими методами (XGBoost: F1 = 0,8571; LSTM Baseline: F1 = 0,7968). Різниця за F1-score між Attention-VAE та найкращою базовою моделлю становить лише 0,0285.
- **Низький рівень хибних тривог.** FPR = 0,0015 забезпечує прийнятний для клінічного застосування рівень хибнопозитивних спрацьовувань (15 на 10 000 нормальних спостережень), що мінімізує ризик alarm fatigue.
- **ROC- та PR-аналіз.** ROC-AUC = 0,8893 та PR-AUC = 0,8246 підтверджують дискримінаційну здатність моделі у широкому діапазоні порогів. Усі три моделі демонструють характерний перегин PR-кривої, що свідчить про наявність об’єктивно складних для детекції випадків.
- **Інтерпретованість.** Механізм уваги забезпечує клінічно обґрунтовану інтерпретацію рішень: ваги уваги концентруються на часових сегментах із фізіологічним погіршенням та корелюють із підвищенням HR, зниженням SpO2 та падінням MAP.
- **Структура латентного простору.** 16-вимірний латентний простір демонструє виразне розділення між нормальними записами та записами

сепсису, що підтверджується візуалізаціями t-SNE та UMAP, а також аналізом окремих латентних вимірів.

- **Ознакова декомпозиція.** Аналіз реконструкційних похибок за ознаками виявляє домінуючу роль HR та MAP у механізмі детекції, що узгоджується з клінічними знаннями про ранні маркери сепсису.
- **Перевага парадигми навчання.** Навчання Attention-VAE без учителя (лише на нормальних записах) є суттєвою практичною перевагою. Вона спрощує адаптацію системи до нових клінічних умов без потреби у розмічених даних позитивного класу.

Отримані результати підтверджують, що запропонований підхід на основі Attention-VAE забезпечує ефективну та інтерпретовану детекцію ранніх ознак сепсису на основі аналізу багатовимірних часових рядів вітальних показників пацієнтів відділень інтенсивної терапії.

ЗАГАЛЬНІ ВИСНОВКИ

У дисертаційній роботі розглянуто проблему раннього виявлення сепсису у відділеннях інтенсивної терапії на основі аналізу багатовимірних часових рядів життєвих показників пацієнтів. Розроблено інтелектуальну систему на основі архітектури Attention-VAE, яка реалізує підхід виявлення аномалій без потреби у розмічених даних з підтвердженими діагнозами сепсису. Нижче наведено узагальнення основних результатів, формулювання наукової новизни, оцінку практичної цінності та визначення напрямів подальших досліджень.

Узагальнення результатів роботи

У виконаному дослідженні розв'язано комплекс взаємопов'язаних задач.

Центральний теоретичний результат: формалізація задачі раннього виявлення сепсису як задачі детектування аномалій у багатовимірних часових рядах. Обґрунтовано, що такий підхід відповідає клінічній природі проблеми: ранні прояви сепсису часто мають характер поступових відхилень від нормального фізіологічного стану, а не дискретних подій з чітко визначеним моментом початку [3; 7]. Вхідні дані подано як шестигодинні часові вікна п'яти показників: ЧСС, SpO2, MAP, температура та RR.

На цій основі спроєктовано і реалізовано архітектуру Attention-VAE: GRU-енкодер формує послідовність прихованих станів, механізм уваги виділяє найбільш інформативні часові кроки у зважене контекстне представлення, варіаційний декодер реконструює вхідний сигнал з латентного простору. Аномальний стан проявляється через підвищену реконструкційну похибку: нормальні патерни відновлюються точно, аномальні відхиляються від «вивченої норми» [1; 3].

Для навчання та оцінювання реалізовано повний конвеєр передобробки даних PhysioNet/Computing in Cardiology Challenge 2019 [35] — від

завантаження записів до формування ковзних вікон із розбиттям на рівні пацієнтів, що виключає витік інформації між вибірками.

Експериментальне оцінювання підтвердило ефективність підходу: Attention-VAE досягла F1 = 0,8286, ROC-AUC = 0,8893, Precision = 98,69% при FPR = 0,15%. Ці результати конкурентоспроможні з XGBoost (F1 = 0,8571) і LSTM (F1 = 0,7968), навченими з учителем, попри те що Attention-VAE не використовує розмічених міток сепсису під час навчання [15; 18]. Аналіз ваг уваги і латентного простору (t-SNE, UMAP) підтвердив клінічну осмисленість внутрішніх представлень моделі.

Формулювання наукової новизни

Наукова новизна дисертаційного дослідження визначається кількома взаємопов'язаними аспектами, що стосуються як розробленого підходу, так і конкретних архітектурних рішень.

Запропоновано застосування архітектури Attention-VAE для задачі раннього виявлення сепсису на основі багатовимірних часових рядів життєвих показників. Це дозволяє поєднати переваги навчання без учителя (відсутність потреби у розмічених даних) з можливістю інтерпретації рішень моделі через механізм уваги. На відміну від більшості існуючих підходів, що потребують точних міток початку сепсису для навчання класифікатора, запропонований метод моделює розподіл нормальних фізіологічних патернів і детектує відхилення від них як потенційні аномалії [3; 18]. Такий підхід стійкіший до невизначеності у встановленні моменту onset сепсису, яка є типовою проблемою для ретроспективних клінічних баз даних.

Додатковий елемент новизни: використання механізму уваги не лише як компоненти, що покращує якість реконструкції, а й як засобу інтерпретації рішень моделі. Ваги уваги визначають, які саме часові сегменти і взаємозв'язки між показниками найбільш релевантні для підвищення показника аномальності. Це створює передумови для клінічно осмисленого

пояснення сигналів тривоги, на відміну від моделей типу “чорної скриньки”, де рішення не піддаються структурованій інтерпретації [4; 5].

Також запропоновано методику комплексного аналізу латентного простору Attention-VAE з використанням методів зниження розмірності (t-SNE, UMAP) для верифікації здатності моделі формувати розрізнювані внутрішні представлення нормальних і аномальних станів. Результати візуалізації підтверджують, що латентний простір набуває структурованого характеру: кластери нормальних і потенційно септичних патернів просторово розділені, що додатково підтверджує клінічну валідність моделі.

Зовнішня багатоцентрова валідація є запланованим наступним етапом дослідження і предметом окремої роботи.

Оцінка практичної цінності

Практична цінність розробленої інтелектуальної системи визначається кількома ключовими характеристиками, що безпосередньо впливають на можливість її впровадження у клінічну практику.

Навчання без учителя становить практичну перевагу. Система не потребує розмічених наборів даних з підтвердженими діагнозами сепсису для навчання: достатньо мати записи життєвих показників пацієнтів без ознак сепсису. Це значно спрощує процес впровадження у нових клінічних закладах, де створення великих розмічених вибірок пов’язане зі значними витратами часу і ресурсів кваліфікованих клініцистів [21]. Модель навчається на даних конкретного відділення інтенсивної терапії, що дозволяє адаптувати її до локальних особливостей популяції пацієнтів, протоколів моніторингу та апаратного забезпечення.

Низька частка хибнопозитивних спрацювань (FPR = 0,15%) особливо важлива для реального клінічного застосування. Проблема *alarm fatigue* (зниження реактивності медичного персоналу на сигнали тривоги внаслідок їх надмірної кількості та високої частки хибних спрацювань) визнана одним з критичних факторів, що знижують ефективність систем моніторингу у

відділеннях інтенсивної терапії [8]. Отриманий показник FPR свідчить, що система здатна мінімізувати кількість неінформативних сигналів тривоги при високій чутливості до реальних аномалій.

Забезпечення інтерпретованості рішень через механізм уваги становить ще один аспект практичної цінності. У клінічних умовах сигнал тривоги з поясненням (які саме показники і в які часові проміжки продемонстрували аномальну динаміку) значно корисніший для лікаря, ніж числове значення ризику без контексту. Теплові карти уваги (attention heatmaps) дають візуальне представлення найбільш інформативних ділянок часового вікна; це допомагає клініцисту швидко оцінити ситуацію і прийняти обґрунтоване рішення щодо подальшого обстеження або терапевтичного втручання [5].

Архітектура системи передбачає інтеграцію з існуючими системами моніторингу пацієнтів у реальному часі. Вхідними даними системи є п’ять базових життєвих показників (ЧСС, SpO2, MAP, температура, RR), які безперервно реєструються приліжковими моніторами у відділеннях інтенсивної терапії. Тобто система не потребує додаткового обладнання або лабораторних досліджень для функціонування, і це спрощує її впровадження. Обчислювальна складність моделі дозволяє виконувати інференс із затримкою, прийнятною для клінічних систем підтримки прийняття рішень.

Додаткова практична перевага: модульна програмна архітектура системи. Вона забезпечує поетапне розширення (додавання нових каналів даних: лабораторних показників, медикаментозних втручань; модифікація механізму уваги; заміна стратегій передобробки або калібрування порогів) без перебудови всього конвеєра. Така архітектура також сприяє відтворюваності результатів і спрощує подальшу підтримку та розвиток системи.

Напрями подальших досліджень

Попри отримані результати, існує низка напрямів, розвиток яких суттєво підвищить якість і клінічну застосовність запропонованого підходу.

Першочерговий напрям: оцінювання моделі на більших і різноманітніших наборах даних. Хоча набір PhysioNet/Computing in Cardiology Challenge 2019 — загальновизнаний еталонний ресурс для задач раннього виявлення сепсису, він відображає практику двох академічних медичних закладів США і не повною мірою репрезентує варіабельність клінічних протоколів, демографічних характеристик та апаратного забезпечення інших закладів [35]. Багатоцентрова валідація на даних з різних медичних закладів, географічних регіонів і систем охорони здоров'я є необхідною умовою для підтвердження узагальнюваності отриманих результатів. Така валідація також оцінить робастність моделі до відмінностей у частотах вимірювань, апаратних артефактах і клінічних протоколах.

Перспективний напрям: дослідження альтернативних конфігурацій механізму уваги. Зокрема, варіанти з багатоголовою увагою (multi-head attention) дозволяють моделі паралельно фокусуватися на різних аспектах вхідних даних: одна «голова» виділяє короткострокові коливання, інша — довгострокові тренди, третя — міжканальні кореляції [4; 23]. Також варто дослідити архітектури на основі Transformer-енкодера, які в останні роки продемонстрували високу ефективність у задачах аналізу часових рядів різної природи [23], на відміну від існуючих Transformer-based підходів для виявлення сепсису [40].

Окремий важливий напрям: інтеграція розробленої системи з клінічними системами підтримки прийняття рішень (CDSS). Це передбачає не лише технічну інтеграцію з системами електронних медичних записів (EHR) і моніторингового обладнання, а й розробку адекватних протоколів реагування на сигнали тривоги, адаптивного калібрування порогів аномальності з урахуванням специфіки конкретного відділення і механізмів зворотного зв'язку від клінічного персоналу [25; 8]. Перспективне також онлайн-калібрування порогів на основі накопиченої статистики спрацювань конкретного відділення.

Ще один напрям подальшої роботи: розширення набору вхідних ознак. Включення лабораторних показників (лактат, прокальцитонін, кількість лейкоцитів), даних про медикаментозні втручання (вазопресори, антибіотики) і контекстуальної інформації (вік, супутні захворювання) потенційно підвищує чутливість і специфічність детектування. Проте таке розширення потребує адаптації архітектури для роботи з гетерогенними і асинхронними даними різної природи та частоти вимірювань [6].

Нарешті, ключовий етап, що передує реальному клінічному впровадженню: проспективне оцінювання системи у режимі реального часу. На відміну від ретроспективного аналізу, проспективна оцінка враховує фактори, які не відображаються в історичних даних: затримки у надходженні вимірювань, тимчасову недоступність окремих каналів даних, вплив рішень персоналу на подальший перебіг і реальну корисність сигналів тривоги для клінічного процесу [8]. Цей етап є логічним продовженням дослідження і необхідною передумовою для переходу від дослідницького прототипу до клінічно валідованого інструменту.

У дисертаційній роботі розроблено і експериментально оцінено інтелектуальну систему раннього виявлення сепсису на основі архітектури Attention-VAE, яка реалізує підхід виявлення аномалій у багатовимірних часових рядах життєвих показників. Запропонований підхід поєднує три головні характеристики: здатність до навчання без розмічених даних; конкурентоспроможну з методами навчання з учителем якість детектування; забезпечення інтерпретованості рішень через механізм уваги. Результати експериментального оцінювання на наборі PhysioNet/Computing in Cardiology Challenge 2019 [35] підтверджують ефективність підходу і формують підґрунтя для подальшого розвитку у напрямі багатоцентрової валідації, розширення архітектурних можливостей і клінічного впровадження. Внесок дисертації полягає у розробці та верифікації запропонованого підходу, який зменшує залежність від дорогої та неоднозначної розмітки клінічних даних,

водночас зберігаючи прозорість моделі для кінцевого користувача (лікаря-клініциста відділення інтенсивної терапії).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Singer M., Deutschman C. S., Seymour C. W. et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016. Vol. 315, No. 8. P. 801–810. DOI: 10.1001/jama.2016.0287.
2. Kingma D. P., Welling M. Auto-Encoding Variational Bayes. *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*. 2014. arXiv: 1312.6114.
3. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. 2015. arXiv: 1409.0473.
4. Bowman S. R., Vilnis L., Vinyals O. et al. Generating Sentences from a Continuous Space. *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL)*. 2016. P. 10–21. arXiv: 1511.06349.
5. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. P. 785–794. DOI: 10.1145/2939672.2939785.
6. Vaswani A., Shazeer N., Parmar N. et al. Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*. 2017. Vol. 30. P. 5998–6008. arXiv: 1706.03762.
7. Zhou C., Paffenroth R. C. Anomaly Detection with Robust Deep Autoencoders. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017. P. 665–674. DOI: 10.1145/3097983.3098052.
8. Che Z., Purushotham S., Cho K. et al. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports*. 2018. Vol. 8, No. 1. Article 6085. DOI: 10.1038/s41598-018-24271-9.

9. Xu H., Chen W., Zhao N. et al. Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications. *Proceedings of the 2018 World Wide Web Conference (WWW)*. 2018. P. 187–196. arXiv: 1802.03903.
10. Harutyunyan H., Khachatrian H., Kale D. C. et al. Multitask Learning and Benchmarking with Clinical Time Series Data. *Scientific Data*. 2019. Vol. 6, No. 1. Article 96. DOI: 10.1038/s41597-019-0103-9.
11. Fawaz H. I., Forestier G., Weber J. et al. Deep Learning for Time Series Classification: A Review. *Data Mining and Knowledge Discovery*. 2019. Vol. 33, No. 4. P. 917–963. DOI: 10.1007/s10618-019-00619-1.
12. Vig J. A Multiscale Visualization of Attention in the Transformer Model. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 2019. P. 37–42. arXiv: 1906.05714.
13. Sendelbach S., Funk M. Alarm Fatigue: A Patient Safety Concern. *AACN Advanced Critical Care*. 2013. Vol. 24, No. 4. P. 378–386. DOI: 10.1097/NCI.0b013e3182a903f9.
14. Little R. J. A., Rubin D. B. *Statistical Analysis with Missing Data*. 3rd ed. Hoboken, NJ : John Wiley & Sons, 2019. 464 p. ISBN: 978-0-470-52679-8.
15. Prechelt L. Early Stopping — But When? *Neural Networks: Tricks of the Trade*. Berlin : Springer, 1998. P. 55–69. DOI: 10.1007/3-540-49430-8_3.
16. Ruff L., Kauffmann J. R., Vandermeulen R. A. et al. A Unifying Review of Deep and Shallow Anomaly Detection. *Proceedings of the IEEE*. 2021. Vol. 109, No. 5. P. 756–795. DOI: 10.1109/JPROC.2021.3052449.
17. Pang G., Shen C., Cao L., van den Hengel A. Deep Learning for Anomaly Detection: A Review. *ACM Computing Surveys*. 2021. Vol. 54, No. 2. Article 38. P. 1–38. DOI: 10.1145/3439950.
18. Pollard T. J., Johnson A. E. W., Raffa J. D. et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care

research. *Scientific Data*. 2018. Vol. 5. Article 180178. DOI: 10.1038/sdata.2018.178.

19. Johnson A. E. W., Bulgarelli L., Shen L. et al. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*. 2023. Vol. 10. Article 1. DOI: 10.1038/s41597-022-01899-x.
20. Zhang Y., Zhou Z., He J. et al. Weakly Supervised Anomaly Detection: A Survey. *Advances in Neural Information Processing Systems (NeurIPS)*. 2022. arXiv: 2302.04549.
21. Ruopp M. D., Perkins N. J., Whitcomb B. W., Schisterman E. F. Youden Index and Optimal Cut-Point Estimated from Observations Affected by a Lower Limit of Detection. *Biometrical Journal*. 2008. Vol. 50, No. 3. P. 419–430. DOI: 10.1002/bimj.200710415.
22. Zaharia M., Chen A., Davidson A. et al. Accelerating the Machine Learning Lifecycle with MLflow. *IEEE Data Engineering Bulletin*. 2018. Vol. 41, No. 4. P. 39–45.
23. Yadan O. Hydra — A Framework for Elegantly Configuring Complex Applications. *GitHub*. 2019. URL: https://github.com/facebookresearch/hydra (дата звернення: 01.03.2026).
24. Lécuyer M., Atlidakis V., Geambasu R. et al. Certified Robustness to Adversarial Examples with Differential Privacy. *Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP)*. 2019. P. 656–672. DOI: 10.1109/SP.2019.00044.
25. Niparko K. J., Gong J. J. Heteroscedastic Temporal Variational Autoencoder for Irregularly Sampled Time Series. *Proceedings of the Machine Learning for Healthcare Conference (MLHC)*. 2020. P. 1–16. arXiv: 2007.09579.
26. Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. 2015. arXiv: 1412.6980.
27. Bergstra J., Bengio Y. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*. 2012. Vol. 13, No. 1. P. 281–305.

28. Virtanen P., Gommers R., Oliphant T. E. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. 2020. Vol. 17, No. 3. P. 261–272. DOI: 10.1038/s41592-019-0686-2.
29. van der Maaten L., Hinton G. Visualizing Data using t-SNE. *Journal of Machine Learning Research*. 2008. Vol. 9, No. 86. P. 2579–2605.
30. McInnes L., Healy J., Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint*. 2018. arXiv: 1802.03426.
31. Kumar A., Roberts D., Wood K. E. et al. Duration of Hypotension before Initiation of Effective Antimicrobial Therapy Is the Critical Determinant of Survival in Human Septic Shock. *Critical Care Medicine*. 2006. Vol. 34, No. 6. P. 1589–1596. DOI: 10.1097/01.CCM.0000217961.75225.E9.
32. Rudd K. E., Johnson S. C., Agesa K. M. et al. Global, Regional, and National Sepsis Incidence and Mortality, 1990–2017: Analysis for the Global Burden of Disease Study. *The Lancet*. 2020. Vol. 395, No. 10219. P. 200–211. DOI: 10.1016/S0140-6736(19)32989-7.
33. Jain S., Wallace B. C. Attention is not Explanation. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. 2019. P. 3543–3556. arXiv: 1902.10186.
34. Muehlematter U. J., Daniore P., Vokinger K. N. Approval of Artificial Intelligence and Machine Learning-Based Medical Devices in the USA and Europe (2017–2020): A Comparative Analysis. *The Lancet Digital Health*. 2021. Vol. 3, No. 3. P. e195–e203. DOI: 10.1016/S2589-7500(20)30292-2.
35. Reyna M. A., Josef C., Jeter R. et al. Early Prediction of Sepsis from Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*. 2020. Vol. 48, No. 2. P. 210–217. DOI: 10.1097/CCM.0000000000004145.

36. Paszke A., Gross S., Massa F. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems (NeurIPS)*. 2019. Vol. 32. arXiv: 1912.01703.
37. Hunter J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. 2007. Vol. 9, No. 3. P. 90–95. DOI: 10.1109/MCSE.2007.55.
38. Su Y., Zhao Y., Niu C. et al. Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019. P. 2828–2837. DOI: 10.1145/3292500.3330672.
39. Correia L., Goos J.-C., Klein P. et al. MA-VAE: Multi-head Attention-Based Variational Autoencoder Approach for Anomaly Detection in Multivariate Time-Series. *Proceedings of the 15th International Joint Conference on Computational Intelligence (NCTA)*. 2023. arXiv: 2309.02253.
40. Dou Y., Li W., Zomaya A. Y. Transformer-Based Unsupervised Learning for Early Detection of Sepsis (Student Abstract). *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. 2022. DOI: 10.1609/aaai.v36i11.21669.
41. Zhang D. et al. An Interpretable Deep-Learning Model for Early Prediction of Sepsis in the Emergency Department. *Patterns*. 2021. Vol. 2, No. 2. Article 100196. DOI: 10.1016/j.patter.2021.100196.
42. Zubair M., Din I., Sarwar N. et al. Revolutionizing Sepsis Diagnosis Using Machine Learning and Deep Learning Models: A Systematic Literature Review. *BMC Infectious Diseases*. 2025. Vol. 25, No. 1. Article 1396. DOI: 10.1186/s12879-025-10766-4.
43. Cho K., van Merriënboer B., Gulcehre C. et al. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014. DOI: 10.3115/v1/D14-1179.
44. Mouyart M., Medeiros Machado G., Jun J.-Y. A Multi-Agent Intrusion Detection System Optimized by a Deep Reinforcement Learning Approach with a Dataset Enlarged Using a Generative Model to Reduce the Bias Effect.

Journal of Sensor and Actuator Networks. 2023. Vol. 12, No. 5. Article 68. DOI: 10.3390/jsan12050068.