

Терентьев Александр Александрович

Доктор технічних наук, професор кафедри інформаційних технологій проектування та прикладної математики, декан факультету автоматизації і інформаційних технологій,

<https://orcid.org/0000-0001-9499-6635>

Київський національний університет будівництва і архітектури, Київ

Соловей Богдан Анатолійович

Аспірант кафедри інформаційних технологій проектування та прикладної математики,

<https://orcid.org/0009-0008-0328-1123>

Київський національний університет будівництва і архітектури, Київ

РЕКУРЕНТНА НЕЙРОННА МЕРЕЖА З МЕХАНІЗМОМ УВАГИ ТА БАЙЄСІВСЬКИМ НАБЛИЖЕННЯМ ДЛЯ ПРОГНОЗУВАННЯ СТІЙКОСТІ МОНОРЕЙКОВОГО КРАНА НА ПЕРЕКИДАННЯ

***Анотація.** Дослідження присвячено розробці моделі прогнозування стійкості монорейкового крана у першому положенні. Об'єктом дослідження є процес прогнозування стійкості крана в експлуатаційних умовах. Проблема, що вирішується у роботі, полягає у підвищенні точності прогнозування, яка в класичних рекурентних нейронних мережах обмежена через ефект згасання градієнтів та втрату довгострокових залежностей у часових рядах. У роботі розроблено рекурентну нейронну мережу типу GRU з механізмом уваги та байєсівським наближенням, що дозволяє враховувати вплив попередніх станів системи та оцінювати невизначеність прогнозів. Запропонована архітектура забезпечує стабільність процесу навчання, що підтверджується відсутністю згасання градієнтів та ефективним оновленням ваг. Отримані результати показали, що модель досягає середньоквадратичної помилки 0,009, середньої абсолютної помилки 0,066 і відносної точності 0,96. Це свідчить про можливість прогнозування коефіцієнта стійкості крана з похибкою менше 10% у 96% випадків. Висока точність пояснюється ефективністю механізму уваги, який дозволяє моделі зосереджуватись на найінформативніших параметрах, та байєсівським наближенням, що підвищує надійність прогнозів за рахунок врахування невизначеності. Відмінною рисою розробленої моделі є поєднання рекурентної структури GRU, механізму уваги та стохастичної (байєсівської) параметризації, що забезпечує стабільне навчання, узагальнення результатів і підвищену точність у порівнянні з класичними нейронними архітектурами. Практичне значення отриманих результатів полягає у можливості використання запропонованої моделі в інтелектуальних системах моніторингу безпечної роботи монорейкового крана, а також у створенні основи для подальших досліджень, зокрема оцінювання стійкості на зсув, що виникає при переміщенні стріли поперек рейкового шляху.*

***Ключові слова:** рекурентні нейронні мережі; механізм уваги; байєсівське наближення; прогностичне моделювання; машинне навчання; стійкість крана; інтелектуальні системи моніторингу*

Вступ

Стійкість вантажопідіймального крана – одна з ключових характеристик, що визначають його надійність та безпеку. Допустимі значення коефіцієнтів стійкості регламентуються нормативними документами, зокрема правилами охорони праці [1]. Тому точне прогнозування та моніторинг стійкості вантажопідіймального крана в режимі реального часу є основою будь-якої сучасної системи забезпечення безпеки кранів.

У роботі [2] була запропонована модель Байєсівської нейронної мережі для прогнозування

стійкості монорейкового крана для 1-го положення, коли стріла направлена уздовж рейкового шляху, а кран знаходиться під навантаженням. Проте архітектура запропонованої моделі під час навчання не враховує історію станів крана за певний часовий період, що може знизити точність прогнозування динамічних змін стійкості під час реальної експлуатації.

Останні досягнення в галузі штучного інтелекту сприяли застосуванню рекурентних нейронних мереж (RNN) та їх модифікацій – LSTM (рекурентна нейронна мережа з довгою короткочасною пам'яттю) і GRU рекурентна нейронна мережа з блочними керованими елементами пам'яті – для оцінювання

стійкості кранів завдяки їхній здатності моделювати послідовні часові ряди даних [3]. Однак навчання рекурентних моделей штучного інтелекту (RNN, LSTM і GRU) пов'язане з проблемою згасаючого або вибухаючого градієнта, яка призводить до втрати довгострокових залежностей у часових рядах і, як наслідок, до зниження точності прогнозування при тривалих послідовностях даних [4].

Для подолання цієї проблеми застосовують механізм уваги (attention mechanism), який дозволяє моделі зосереджуватися на вагомих елементах вхідної послідовності та покращує здатність мережі відстежувати довготривалі залежності між станами системи [5].

У даному дослідженні розглядається можливість застосування GRU з механізмом уваги і байесівським наближенням для прогнозування стійкості монорейкового крана для 1-го положення.

Проблема дослідження

Нехай стан монорейкового крана в момент часу t описується вектором ознак $x_t \in \mathbb{R}^n$, де n – кількість параметрів. Тоді еволюція стану крана протягом періоду з T часових кроків, може бути представлена як послідовність векторів $X = (x_1, x_2, \dots, x_T)$.

Стандартна архітектура рекурентних нейронних мереж, до яких належить і GRU, обробляє таку послідовність рекурентно. Прихований стан h_t на кожному кроці є функцією поточного входу x_t та попереднього прихованого стану h_{t-1} . Ключовим недоліком такої архітектури при прогнозуванні є те, що для висновку використовується лише останній прихований стан h_T , в який має бути стиснута вся інформація про попередню історію. Це призводить до проблеми згасаючого градієнта (vanishing gradient) під час навчання, що значно погіршує здатність моделі враховувати події, що відбулися на ранніх етапах послідовності.

Рис. 1 демонструє обмеження стандартної GRU у виявленні довгострокових залежностей.

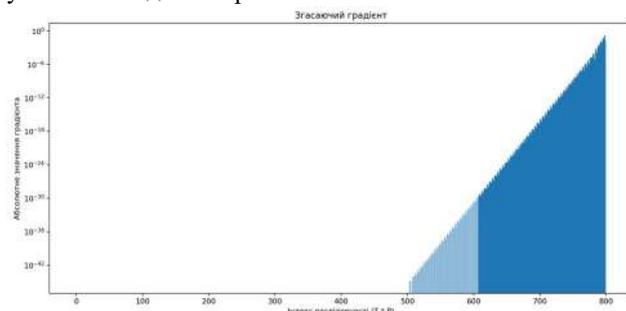


Рисунок 1 – Аналіз значення градієнта похибки при навчанні стандартної GRU

Для штучно створеної послідовності довжиною 800 кроків, де лише початковий вхід x_0 визначає фінальний стан, обчислено градієнт похибки прогнозу відносно кожного вхідного кроку.

Величина градієнта для останніх кроків послідовності є дорівнює 0.164 для кроку 798, однак вона згасає при русі назад у часі. Вже для середніх кроків послідовності її значення становить ($1.0e-45$ для кроку 503), а для початкових кроків (зокрема, для $t=0$) градієнт є практично нульовим.

Аналіз значень градієнта на рис. 1 емпірично підтверджує, що модель глибинного навчання типу стандартної GRU нездатна навчитися зв'язку між початковою критичною подією в експлуатації крана та кінцевим результатом прогнозування його стану. Будь-які оновлення вагових коефіцієнтів моделі, засновані на фінальній похибці, матимуть незначний вплив на те, як модель обробляє початкові вхідні дані. Це унеможливує вивчення довгострокових залежностей і, як наслідок, отримання достовірних прогнозів, щодо стану стійкості монорейкового крана.

Мета та задачі дослідження

Метою дослідження є розробка рекурентної нейронної мережі з блочними (керованими) елементами пам'яті (GRU) з механізмом уваги та байесівським наближенням для прогнозування стійкості монорейкового крана у 1-му положенні.

Практичне значення роботи полягає у використанні розробленої моделі для систем моніторингу безпеки монорейкового крана.

Для досягнення поставленої мети визначено такі задачі:

- розробити архітектуру рекурентної нейронної мережі GRU з механізмом уваги і байесівським наближенням;
- оцінити точність розробленої моделі у прогнозуванні стійкості монорейкового крана у 1-му положенні за критерієм допустимого рівня похибки, рівного 10%, який має бути досягнутий не менше ніж у 95% прогнозів.

Аналіз останніх досліджень і публікацій

Моделі LSTM та GRU суттєво покращили прогнозування часово-залежних параметрів у інженерних системах. У [6] RNN на основі LSTM зменшила коливання кранів; у [7] застосували LSTM для виявлення помилок у контейнерних кранах; у [8] поєднали CNN та LSTM для автоматизації кранів за складних візуальних умов; а у [9] інтегрували CNN та GRU для захоплення просторово-часових залежностей, досягнувши високої точності прогнозування.

Рекурентна нейронна мережа (RNN) обробляє послідовні вхідні сигнали, оновлюючи приховані стани h_t та формуючи вихідні сигнали. Моделі LSTM розширюють цю концепцію за допомогою входових, забуваючих та вихідних воріт (рис. 2), що дозволяє моделювати довгострокові залежності шляхом

керованого оновлення стану клітинки C_t . Моделі GRU спрощують цю архітектуру, об'єднуючи входів та забуваючі ворота в єдине ворото оновлення та усуваючи явний стан клітинки, покладаючись лише на прихований стан h_t [10].

Для прогнозування стійкості монорейкового крана необхідна модель, здатна ефективно захоплювати просторово-часові залежності сигналів при обмеженому обсязі даних та швидко навчатися. GRU в цьому контексті має кілька переваг: спрощена архітектура порівняно з LSTM з меншим числом параметрів забезпечує швидше навчання та зменшує ризик перенавчання, при цьому зберігаючи здатність моделювати довгострокові залежності. Крім того, інтеграція механізму уваги дозволяє виділяти ключові сигнали, які найбільше впливають на стійкість крана [10].

Але стандартна GRU з механізмом уваги не дає змогу кількісно оцінювати невизначеність прогнозів, що обмежує їхню надійність у завданнях моніторингу безпеки кранів.

Таким чином, необхідно розробити архітектуру рекурентної моделі GRU з механізмом уваги і байєсівським наближенням для прогнозування стійкості монорейкового крана.

Виклад основного матеріалу

Архітектура GRU з байєсівським наближенням розглядає всі параметри моделі як випадкові величини та використовує Гауссові апіорні розподіли [11]. Процес прямого поширення та інтеграція механізму уваги представлені на рис. 3.

Вибірка вагових матриць для вентилів (gates) та зсувів (biases) GRU визначається як:

$$W_x = \begin{bmatrix} W_{zx} \\ W_{rx} \\ W_{hx} \end{bmatrix} = \mu_x + \sigma_x \cdot T_x = \mu_x + \text{softplus}(\rho_x) \cdot T_x, \quad (1)$$

$$W_h = \begin{bmatrix} W_{zh} \\ W_{rh} \\ W_{hh} \end{bmatrix} = \mu_h + \sigma_h \cdot T_h = \mu_h + \text{softplus}(\rho_h) \cdot T_h, \quad (2)$$

$$b = \mu_b + \text{softplus}(\rho_b) \cdot \epsilon_b, \quad (3)$$

де $\epsilon_x, \epsilon_h, \epsilon_b \sim \mathcal{N}(0, I)$ – це Гаусів шум для вибірки; $\mu_x, \mu_h \in \mathbb{R}^{3H \times D}$ – параметри математичних сподівань

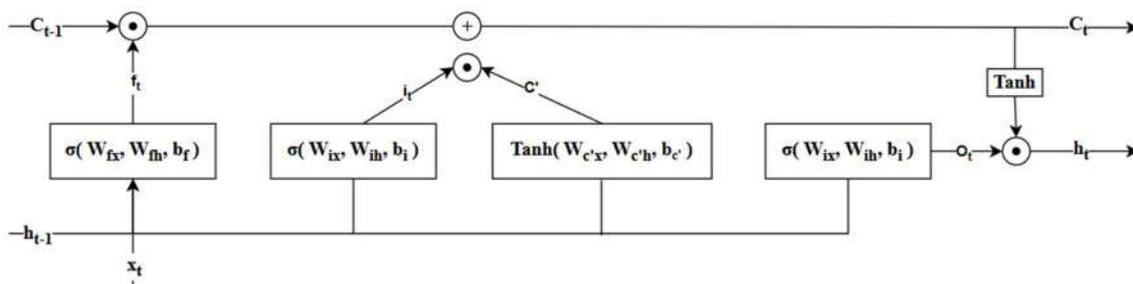


Рисунок 2 – Архітектура однієї клітини моделі LSTM

розподілів ваг для входних та прихованих станів; $\rho_x, \rho_h \in \mathbb{R}^{3H \times D}$ та $\rho_b \in \mathbb{R}^{3H}$ – параметри, що навчаються, і використовуються для параметризації стандартних відхилень σ_x, σ_h розподілів ваг у байєсівській нейронній мережі за допомогою функції softplus; $\mu_b \in \mathbb{R}^{3H}$ – параметер математичного сподівання розподілу зсувів.

Для навчання моделі використовується техніка усіченого зворотного поширення в часі (TBPTT). Після розбиття входної послідовності довжиною T на m частин довжиною k : $[1, k], [k+1, 2k], \dots, [T-k+1, T]$, рівняння прямого поширення для кожної частини (i) визначаються як:

$$h_t = \text{GRU}(x_t, h_{t-1}), t \in [T[i], T[i+k-1]]. \quad (4)$$

Метою навчання є максимізація нижньої межі доказовості (Evidence Lower Bound, ELBO) [12]. Для досягнення мети навчання за допомогою стандартних градієнтних оптимізаторів, це еквівалентно мінімізації негативного ELBO, що еквівалентно мінімізації негативної ELBO, яка визначається як функція втрат L_{loss} . Для кожної частини (i) на інтервалі $[t_i, t_{i+k-1}]$ функція втрат має вигляд:

$$L_{\text{loss}}^{(i)} = -E_{q(W)} \left[\sum_{t=t_i}^{t_i+k-1} \log p(y_t | x_t, h_{t-1}, W) \right] + \frac{1}{m} \text{KL}(q(W) \| p(W)), \quad (5)$$

де $p(y_t | x_t, h_{t-1}, W)$ – правдоподібність спостереження істинного цільового значення y_t ; $q(W)$ – апроксимований варіаційний апостеріорний розподіл ваг; $p(W)$ – фіксований апіорний розподіл ваг; $\text{KL}(q(W) \| p(W))$ – дивергенція Кульбака–Лейблера між апостеріорним та апіорним розподілами; X та Y – послідовності входних даних та відповідних цільових значень.

Перший член рівняння (5) вимірює відповідність прогнозів моделі даним, а член KL діє як регуляризатор, контролюючи складність моделі. Градієнт $\nabla_W \mathcal{L}_{\text{loss}}^{(i)}$ обчислюється за допомогою правила ланцюга:

$$\nabla_{\rho_W} \mathcal{L}_{\text{loss}}^{(i)} = \sum_{j=t}^{t+k-1} \frac{\partial \mathcal{L}_{\text{loss}}^{(i)}}{\partial W_j} \cdot \frac{\partial W_j}{\partial \rho_W}. \quad (1)$$

$$\nabla_{\mu_W} \mathcal{L}_{\text{loss}}^{(i)} = \sum_{j=t}^{t+k-1} \frac{\partial \mathcal{L}_{\text{loss}}^{(i)}}{\partial W_j} \cdot \frac{\partial W_j}{\partial \mu_W}. \quad (2)$$

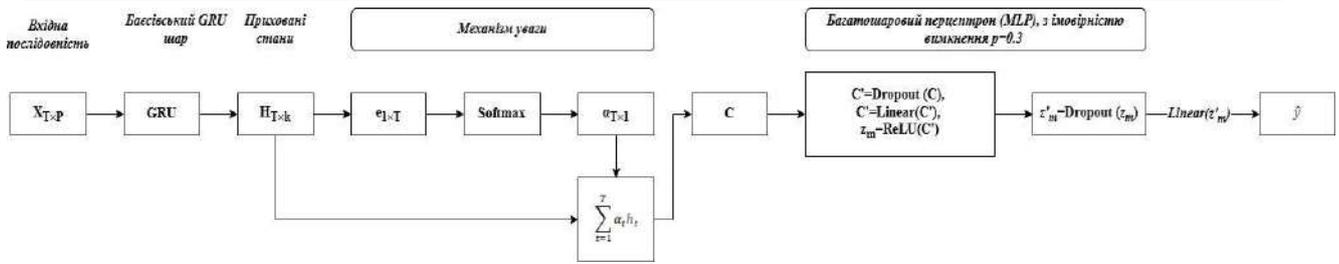


Рисунок 3 – Архітектура GRU з механізмом уваги і байесівським наближенням

Параметри μ_w, ρ_w оновлюються за допомогою методу градієнтного спуску:

$$\mu_w^{(i)} = \mu_w^{(i)} - \alpha \nabla_{\mu_w} \mathcal{L}_{loss}^{(i)}, \quad (8)$$

$$\rho_w^{(i)} = \rho_w^{(i)} - \alpha \nabla_{\rho_w} \mathcal{L}_{loss}^{(i)}, \quad (9)$$

де α – швидкість навчання.

Архітектура GRU з механізмом уваги і байесівським наближенням

Механізм уваги вимагає власного набору вагових коефіцієнтів, які в байесівському контексті також розглядаються як випадкові величини з розподілами. Вагові матриці для механізму уваги визначаються так:

$$W_a = \mu_{W_a} + \text{softplus}(\rho_{W_a}) \cdot \epsilon_{W_a}, \quad (10)$$

$$v_a = \mu_{v_a} + \text{softplus}(\rho_{v_a}) \cdot \epsilon_{v_a}, \quad (11)$$

де W_a, v_a – вагові коефіцієнти шарів уваги; μ та ρ – параметри розподілів, що навчаються; ϵ – Гауссів шум.

Процес прямого поширення складається з трьох етапів. Спочатку GRU-кодер обробляє вхідну послідовність X , генеруючи повну послідовність прихованих станів $H = [h_1, h_2, \dots, h_T]$:

$$h_t = \text{GRU}(x_t, h_{t-1}), \forall t \in [1, T]. \quad (12)$$

Механізм уваги обчислює оцінки значущості e_t для кожного прихованого стану $h_t \in H$:

$$e_t = v_a^T \tanh(W_a h_t). \quad (13)$$

Оцінки e_t нормалізуються за допомогою функції softmax для отримання вагових коефіцієнтів уваги α_t :

$$\alpha_t = \frac{\exp(e_t)}{\sum_{j=1}^T \exp(e_j)}. \quad (14)$$

Вектор \vec{c} механізму уваги, c зваженою сумою всіх прихованих станів H :

$$c = \sum_{t=1}^T \alpha_t h_t. \quad (15)$$

Прогнозоване значення стійкості монорейкових пересувних кранів \hat{y} є вихідним нейроном багатошарового перцептрона (MLP), який приймає на вхід контекстний вектор c :

$$\hat{y} = \text{MLP}(\vec{c}). \quad (16)$$

Використання байесівського наближення – Monte Carlo Dropout – під час активації нейронів у процесі навчання дозволяє отримати не лише середнє значення прогнозу, але й оцінити його дисперсію [13].

Функція витрат (5) для навчання Байесівської GRU з механізмом уваги приймає вигляд:

$$\mathcal{L}_{loss}^{(i)} = -E_{q(\tilde{w})} \left[\log p(y|X, \tilde{w}) \right] + \text{KL}(q(\tilde{w})||p(\tilde{w})), \quad (17)$$

де $\tilde{W} = \{W_a, v_a, W_x, W_h, b\}$ – набір усіх вагових коефіцієнтів моделі; $p(y|X, \tilde{W})$ правдоподібність спостереження істинного цільового значення (y) за умови, що предиктивний розподіл моделі згенеровано з усієї вхідної послідовності X ; дивергенція Кульбака – Лейблера $\text{KL}(q(\tilde{w})||p(\tilde{w}))$ включає апіорні та апостеріорні розподіли для нових вагових коефіцієнтів уваги, таким чином регуляризуючи складність усієї моделі.

Основні байесівські принципи для вагових коефіцієнтів (1–3) та правила оновлення градієнтів (6–9) залишаються концептуально незмінними, але тепер поширюються і на нові шари механізму уваги.

Підготовка експерименту дослідження

В роботі [2] визначена математична модель стійкості S монорейкового крана для 1-го положення, коли стріла направлена уздовж рейкового шляху, а кран знаходиться під навантаженням:

$$S = \begin{cases} 1, & M_{ут}/M_{пер} \geq 1.4 \\ 0, & M_{ут}/M_{пер} < 1.4 \end{cases}, \quad (18)$$

де $M_{ут} = G_{кр}(k-u[;1]) + G_{гр}(a-u[;2])$ – сума моментів від власної ваги крана та від групи вантажу з урахуванням деформацій балки; $M_{пер} = G_{гр}(a-k-u[;3])$ – момент від вантажу, який прагне перекинути кран, скоригований на переміщення; u_1, u_2, u_3 – переміщення у відповідних вузлах, які входять до вектор-стовпця невідомих вузлових переміщень і кутів повороту u_f .

Для моделювання експлуатаційних умов було створено синтетичний набір даних, отриманий шляхом вибірки з рівномірних розподілів. Межі вибірки визначено відповідно до типових технічних характеристик монорейкових кранів, наведених у табл. 1 [14].

Таблиця 1 – Вибірki з рівномірних розподілів для створення набору даних для тренування GRU

Параметр	Опис параметра	Діапазон
L (X1)	Повна довжина основної балки крана, м	[5.0, 15.0]
E (X2)	Модуль Юнга матеріалу балки, Па	[2.0×10 ¹¹ , 2.5×10 ¹¹]
I (X3)	Момент інерції поперечного перерізу балки, м ⁴	[0.5×10 ⁻⁴ , 1.5×10 ⁻⁴]

Закінчення табл. 1

$G_{kr}(X4)$	Вага (сила тяжіння) конструкції крана, кН	[40,100]
$G_{gr}(X5)$	Маса вантажу, кН	[60,120]
$a(X6)$	Відстань від опорної точки балки до вантажу, м	[4.0, 7.0]
$K(X7)$	Відстань від осі обертання крана до ходового колеса, м	[1.0, 2.0]
$S(X8)$	Відстань між рейковими шляхами (база опори), м	[1.5, 3.5]

На основі зазначених параметрів та їх діапазонів буде сформовано синтетичний набір даних з 2000-ми записів.

Метод оцінювання розробленої моделі

Здатність GRU з механізмом уваги і байесівським наближенням прогнозувати в межах допустимого рівня похибки, рівного 10%, визначимо на основі відносної точності (RelAcc), що обчислюється як:

$$RelAcc = \frac{1}{N} \sum_{i=1}^N I\left(\frac{S_i - \hat{S}_i}{|S_i|} < 10\%\right), \quad (19)$$

де $I(\cdot)$ – функція, яка набуває значення 1, якщо умова виконується, і 0 – інакше; S_i – фактичне значення індексу стійкості у наборі даних; \hat{S}_i – прогнозоване значення індексу стійкості.

Точність прогнозування моделі оцінюється за допомогою двох основних метрик:

Середньоквадратична похибка (Mean Square Error, RMSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (S_i - \hat{S}_i)^2. \quad (20)$$

Середня абсолютна похибка (Mean Absolute Error, MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |S_i - \hat{S}_i|. \quad (21)$$

Налаштування експерименту

Модель розроблено з використанням мови програмування Python у середовищі програмного забезпечення: PyTorch (версія 2.7.0+cpu), NumPy (версія 2.2.4), Pandas (версія 2.2.3), Scikit-learn (версія 1.6.1), Matplotlib (версія 3.10.1). Набір даних поділений наступним чином: 70% записів використано для навчальної вибірки, а решта 30% – для тимчасової вибірки. Тимчасову вибірку надалі розділено, у результаті чого отримали 15% контрольної вибірки та 15% тестової вибірки. Для забезпечення відтворюваності результатів усіх експериментів було використано фіксоване «випадкове зерно» зі значенням 42 для всіх процедур поділу даних і ініціалізації ваг моделей.

Результати

На рис. 4 представлені абсолютні значення градієнтів у логарифмічній шкалі, отримані в процесі зворотного поширення помилки на різних етапах обробки даних для розробленої моделі GRU з

механізмом уваги і байесівським наближенням. Значення градієнтів не згасають до нуля навіть на ранніх етапах послідовності. Хоча спостерігається певна варіативність, градієнти залишаються в стабільному діапазоні (переважно між 10^{-6} та 10^{-4}), що є достатнім для ефективного оновлення ваг моделі.

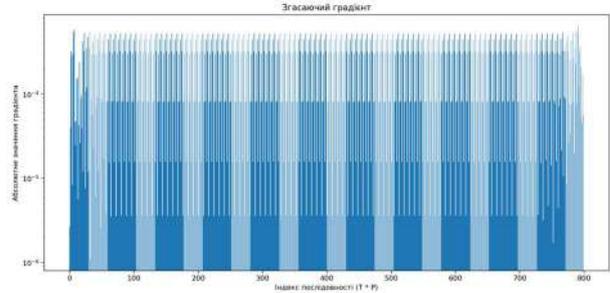


Рисунок 3 – Аналіз значень градієнта похибки при навчанні GRU з механізмом уваги

Стабільність градієнта досягається завдяки розробленому механізму уваги, який визначає найважливіші частини вхідної послідовності (14) і створює поширення градієнтів до цих часових кроків, минаючи кроки з низькими ваговими коефіцієнтами.

На рис. 5, 6 представлено роботу механізму уваги для двох різних спостережень з тестової вибірки. Висота стовпця відповідає рівню "важливості", який модель призначила ознаці. Помаранчева пунктирна лінія показує масштабоване вхідне значення для кожної ознаки.

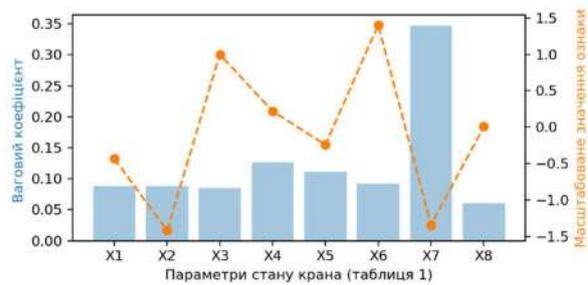


Рисунок 4 – Розподіл ваги уваги для спостереження з індексом 6 з набору для навчання

На рис. 5 для спостереження з індексом 6 вага уваги для ознаки X7 є домінуючою і перевищує 0.3, тоді як ваги для інших ознак (X1-X6, X8) є приблизно однакові та близькі до 0.1. Це означає, що для прогнозу спостереження 6 модель вважала значення ознаки X7 найважливішим. Помаранчева лінія показує, що ознака X7 має низьке (негативне) значення, близьке до -1.5. Хоча ознака X6 має високе позитивне значення (близько 1.4), модель проігнорувала X6 і рішення моделі базувалося на одному значенні ознаки X7.

На рис. 6 для спостереження з індексом 169 розподіл уваги є розподілений між ознаками X7 (0.23), X6(0.18), X4(0.14), X8 (0.15). Таким чином, рішення моделі базувалося на комбінації ознак, а не на одній.

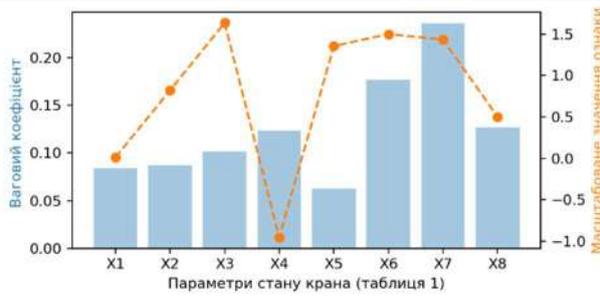


Рисунок 5 – Розподіл ваги уваги для спостереження з індексом 169 з набору для навчання

Рис. 7 показує динаміку функції витрат (17) для наборів тренування та контролю. Крива демонструє падіння з ~2.7 до ~0.2 на ітераціях (1-20), після чого крива поступово вирівнюється, досягаючи значення (близько 0.05), що свідчить про успішне навчання. Крива витрат контрольного набору протягом усього процесу навчання залишається нижчою за навчальну, що свідчить про здатність моделі узагальнювати знання на нових даних. Крива витрат на контрольному наборі не зростає з часом, що доводить відсутність перенавчання.

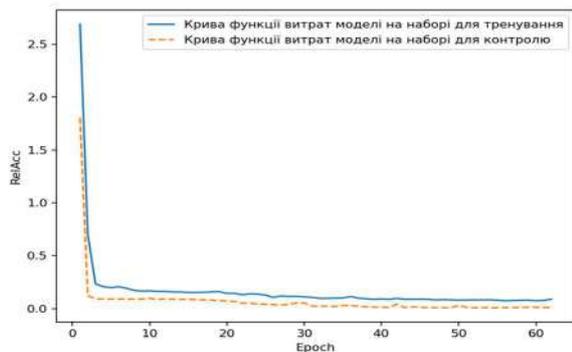


Рисунок 6 – Динаміка функції витрат моделі для наборів тренування та контролю

Рис. 8 демонструє зміну відносної точності RelAcc для наборів тренування та контролю. Відносна точність моделі на наборі для тренування зростає до значень 0.45–0.5 за 60 ітерацій навчання. Відносна точність на контрольному наборі зростає до 0.96, але спостерігаються різкі коливання.

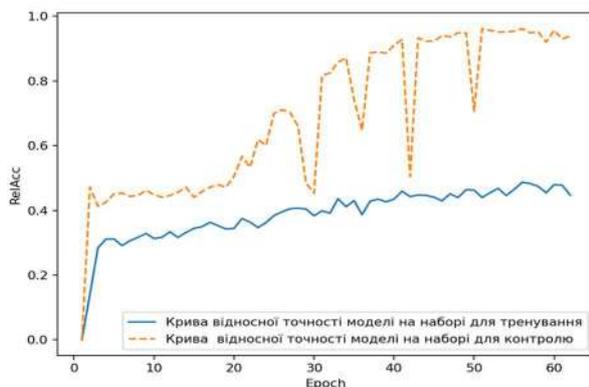


Рисунок 7 – Крива відносної точності RelAcc моделі для наборів тренування та контролю

Відносна точність на контрольному наборі вдвічі вища, ніж для навчального, через використання регуляризації Dropout. Під час навчання частина нейронів "вимикається", що штучно погіршує продуктивність на тренувальних даних. На етапі контролю Dropout вимикається, і модель використовує всю архітектуру, що забезпечує кращі результати.

У табл. 2 наведено оцінки розробленої GRU з механізмом уваги і байєсівським наближенням на тестових даних. Результати показують, що похибка прогнозу менше 10% була досягнута в 96% випадків, а середньоквадратична (MSE) та середня абсолютна (MAE) помилки склали 0.009 та 0.066 відповідно.

Таблиця 2 – Оцінки GRU з механізмом уваги і байєсівським наближенням

Test MSE	Test MAE	RelAcc (<10%)
0.009	0.066	0.96

Таким чином, дані з табл. 2 підтверджують точність розробленої моделі у прогнозуванні стійкості монорейкового крана у 1-му положенні за критерієм допустимого рівня похибки, рівного 10%, який має бути досягнутий не менше ніж у 95% прогнозів.

Висновки

У роботі розроблено та досліджено рекурентну нейронну мережу типу GRU з механізмом уваги та байєсівським наближенням для прогнозування стійкості монорейкового крана у першому положенні.

Аналіз динаміки градієнтів показав відсутність ефекту їх згасання під час навчання, що свідчить про стабільність оновлення ваг моделі. Запропонований механізм уваги забезпечує адаптивний розподіл вагових коефіцієнтів між ознаками, дозволяючи моделі ідентифікувати найбільш значущі параметри, які впливають на стійкість системи. Результати експериментів підтвердили ефективність навчання – функція витрат знизилася до ~0.05 без ознак перенавчання, а відносна точність моделі на контрольному наборі досягла 0.96. Середньоквадратична (MSE) і середня абсолютна (MAE) помилки становили 0.009 та 0.066 відповідно.

Таким чином, запропонована модель GRU з механізмом уваги та байєсівським наближенням забезпечує високу надійність та узагальнювальну здатність при моделюванні експлуатаційної стійкості монорейкових кранів.

Розроблена модель може бути використана у системах моніторингу та прогнозування стійкості монорейкового крана у першому положенні, а також слугувати основою для подальших досліджень, зокрема розробки моделі для оцінювання стійкості монорейкового крана на зсув, який може виникати при переміщенні стріли поперек рейкового шляху.

Список літератури

1. Міністерство соціальної політики України. Про затвердження Правил охорони праці під час експлуатації вантажопідіймальних кранів, підіймальних пристроїв і відповідного обладнання : Наказ від 15.01.2014 № 9. URL: <https://zakon.rada.gov.ua/laws/show/z0083-14> (дата звернення: 28.01.2026).
2. Terentyev O., Solovei B. Machine learning of a Bayesian neural network with gamma distribution for stability assessment of a monorail crane. *Management of Development of Complex Systems*. 2025. 62. P. 134–140. DOI: <https://doi.org/2412-9933.2025.62.134-140>.
3. Li W., Law K. E. Deep learning models for time series forecasting: A review. *IEEE Access*. 2024. 12. P. 92306–92327. DOI: <https://doi.org/10.1109/ACCESS.2024.3422528>.
4. Mienye I. D., Swart T. G., Obaido G. Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*. 2024. 15(9). 517. DOI: <https://doi.org/10.3390/info15090517>.
5. Hua Q., Fan Z., Mu W., Cui J., Xing R., Liu H., Gao J. A short-term power load forecasting method using CNN-GRU with an attention mechanism. *Energies*. 2024. 18(1). 106. DOI: <https://doi.org/10.3390/en18010106>.
6. Cui X., Chipusu K., Ashraf M. A., Riaz M., Xiahou J., Huang J. Symmetry-Enhanced LSTM-Based Recurrent Neural Network for Oscillation Minimization of Overhead Crane Systems during Material Transportation. *Symmetry*. 2024. 16(7). 920. DOI: <https://doi.org/10.3390/sym16070920>.
7. Awasthi A., Krpalkova L., Walsh J. Deep Learning-Based Boolean, Time Series, Error Detection, and Predictive Analysis in Container Crane Operations. *Algorithms*. 2024. 17(8). 333. DOI: <https://doi.org/10.3390/a17080333>.
8. Safaei M., Hejazian M., Pedrammehr S., Pakzad S., Eftefagh M., Fotouhi M. Damage Detection of Gantry Crane with a Moving Mass Using Artificial Neural Network. *Buildings*. 2024. 14(2). 458. DOI: <https://doi.org/10.3390/buildings14020458>.
9. Arshad H., Zayed T., Bakhtawar B., Chen A., Li H. Damage assessment of modular integrated construction during transport and assembly using a hybrid CNN-Gated recurrent unit model. *Automation in Construction*. 2025. 174. 106136. DOI: <https://doi.org/10.1016/j.autcon.2025.106136>.
10. Shao X., Ma Y., Chang D., Zhang Q. Predictive Maintenance for Gantry Crane Shafts: Enhancing Sensor Features With Dynamic Dual-Domain Competitive Attention GRU Autoencoder. *IEEE Sensors Journal*. 2024. 24(11). P. 18154–18171. DOI: <https://doi.org/10.1109/JSEN.2024.3391373>.
11. Obi N. E., Nwosu H. U., Shadrack M. U. A new machine learning model for predicting reliability of cranes. *Journal of African Sustainable Development*. 2024.
12. Noor T., Lone S. N., Ramana G. V., Nayek R. A recursive Bayesian neural network for constitutive modeling of sands under monotonic loading. *arXiv preprint arXiv:2501.10088*. 2025.
13. Shorinwa O., Mei Z., Lidard J., Ren A. Z., Majumdar A. A survey on uncertainty quantification of large language models: Taxonomy, open research challenges, and future directions. *ACM Computing Surveys*. 2025.
14. Gorbatyuk E., Bulavka O., Voliyanuk V. Analysis of studies of stationary tower cranes under wind loads. *Гірничі, будівельні, дорожні та меліоративні машини*. 2023. 102. С. 17–23. DOI: <https://doi.org/10.32347/gbdmm.2023.102.01>.

Стаття надійшла до редакції 18.11.2025

Terentyev Oleksandr

DSc (Eng), Professor of the Department of Information Technologies in Design and Applied Mathematics, Dean of the Faculty of Automation and Information Technologies,
<https://orcid.org/0000-0001-9499-6635>

Kyiv National University of Construction and Architecture, Kyiv

Solovei Bohdan

Postgraduate Student of the Department of Information Technologies in Design and Applied Mathematics,
<https://orcid.org/0009-0008-0328-1123>

Kyiv National University of Construction and Architecture, Kyiv

RECURRENT NEURAL NETWORK WITH ATTENTION MECHANISM AND BAYESIAN APPROXIMATION FOR PREDICTING THE OVERTURNING STABILITY OF A MONORAIL CRANE

Abstract. This study is devoted to the development of a model for predicting the stability of a monorail crane in its first position. The object of the study is the process of predicting the stability of a monorail crane under operating conditions. The problem addressed in the paper is the improvement of prediction accuracy, which is limited in classical recurrent neural networks due to the gradient vanishing effect and the loss of long-term dependencies in time series. In this work, a Gated Recurrent Unit (GRU) type recurrent neural network with an attention mechanism and Bayesian approximation has been developed, allowing for the consideration of previous system states and the estimation of forecast uncertainty. The proposed architecture ensures the stability of the training process, as confirmed by the absence of vanishing gradients and efficient weight updates. The results obtained showed that the model achieves a Root Mean Square Error of 0.009, a Mean Absolute Error of 0.066, and a relative accuracy of 0.96. This indicates the possibility of predicting the crane's stability coefficient with an error of less than 10% in 96% of cases. The high accuracy is explained by the effectiveness of the attention mechanism, which allows the model to focus on the

most informative parameters, and the Bayesian approximation, which increases the reliability of forecasts by accounting for uncertainty. A distinctive feature of the developed model is the combination of the GRU recurrent structure, the attention mechanism, and stochastic (Bayesian) parameterization, which ensures stable learning, generalization of results, and increased accuracy compared to classical neural architectures. The practical significance of the results lies in the possibility of using the proposed model in intelligent monitoring systems for the safe operation of a monorail crane, as well as serving as a basis for further research, specifically for assessing shear stability that occurs when moving the boom across the rail track.

Keywords: recurrent neural networks; attention mechanism; Bayesian approximation; predictive modeling; machine learning; crane stability; intelligent monitoring systems

References

1. Ministry of Social Policy of Ukraine. (2014). *On approval of the Occupational Health and Safety Rules for the operation of load-lifting cranes, lifting devices and related equipment* (Order No. 9). URL: <https://zakon.rada.gov.ua/laws/show/z0083-14> [In Ukrainian].
2. Terentyev, O., & Solovei, B. (2025). Machine learning of a Bayesian neural network with gamma distribution for stability assessment of a monorail crane. *Management of Development of Complex Systems*, 62, 134–140. <https://doi.org/2412-9933.2025.62.134-140>.
3. Li, W., & Law, K. E. (2024). Deep learning models for time series forecasting: A review. *IEEE Access*, 12, 92306–92327. <https://doi.org/10.1109/ACCESS.2024.3422528>.
4. Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, 15(9), 517. <https://doi.org/10.3390/info15090517>.
5. Hua, Q., Fan, Z., Mu, W., Cui, J., Xing, R., Liu, H., & Gao, J. (2024). A short-term power load forecasting method using CNN-GRU with an attention mechanism. *Energies*, 18 (1), 106. <https://doi.org/10.3390/en18010106>.
6. Cui, X., Chipusu, K., Ashraf, M. A., Riaz, M., Xiahou, J., & Huang, J. (2024). Symmetry-Enhanced LSTM-Based Recurrent Neural Network for Oscillation Minimization of Overhead Crane Systems during Material Transportation. *Symmetry*, 16(7), 920. <https://doi.org/10.3390/sym16070920>.
7. Awasthi, A., Krpalkova, L., & Walsh, J. (2024). Deep Learning-Based Boolean, Time Series, Error Detection, and Predictive Analysis in Container Crane Operations. *Algorithms*, 17 (8), 333. <https://doi.org/10.3390/a17080333>.
8. Safaei, M., Hejazian, M., Pedrammehr, S., Pakzad, S., Etefagh, M., & Fotouhi, M. (2024). Damage Detection of Gantry Crane with a Moving Mass Using Artificial Neural Network. *Buildings*, 14(2), 458. <https://doi.org/10.3390/buildings14020458>.
9. Arshad, H., Zayed, T., Bakhtawar, B., Chen, A., & Li, H. (2025). Damage assessment of modular integrated construction during transport and assembly using a hybrid CNN-Gated recurrent unit model. *Automation in Construction*, 174, 106136. <https://doi.org/10.1016/j.autcon.2025.106136>.
10. Shao, X., Ma, Y., Chang, D., & Zhang, Q. (2024). Predictive Maintenance for Gantry Crane Shafts: Enhancing Sensor Features With Dynamic Dual-Domain Competitive Attention GRU Autoencoder. *IEEE Sensors Journal*, 24 (11), 18154–18171. <https://doi.org/10.1109/JSEN.2024.3391373>.
11. Obi, N. E., Nwosu, H. U., & Shadrack, M. U. (2024). A new machine learning model for predicting reliability of cranes. *Journal of African Sustainable Development*.
12. Noor, T., Lone, S. N., Ramana, G. V., & Nayek, R. (2025). A recursive Bayesian neural network for constitutive modeling of sands under monotonic loading. *arXiv preprint arXiv:2501.10088*.
13. Shorinwa, O., Mei, Z., Lidard, J., Ren, A. Z., & Majumdar, A. (2025). A survey on uncertainty quantification of large language models: Taxonomy, open research challenges, and future directions. *ACM Computing Surveys*.
14. Gorbatiyuk, E., Bulavka, O., & Voliyanuk, V. (2023). Analysis of studies of stationary tower cranes under wind loads. *Mining, Construction, Road and Reclamation Machines*, 102, 17–23. <https://doi.org/10.32347/gbdmm.2023.102.01> [In Ukrainian].

Посилання на публікацію

- | | |
|------|---|
| APA | Terentyev, O., & Solovei, B. (2025). Recurrent Neural Network with Attention Mechanism and Bayesian Approximation for Predicting the Overturning Stability of a Monorail Crane. <i>Management of Development of Complex Systems</i> , 64, 217–224. dx.doi.org/10.32347/2412-9933.2025.64.217-224 . |
| ДСТУ | Терент'єв О. О., Соловей Б. А. Рекурентна нейронна мережа з механізмом уваги та байєсівським наближенням для прогнозування стійкості монорейкового крана на перекидання. <i>Управління розвитком складних систем</i> . Київ, 2025. № 64. С. 217 – 224. dx.doi.org/10.32347/2412-9933.2025.64.217-224 . |