31 Octobre;
Paris, République française

SECTION 12.
TECHNOLOGIES ET SYSTÈMES D'INFORMATION

DOI 10.36074/logos-31.10.2025.022

ОПТИМІЗАЦІЯ МУЛЬТИМОДАЛЬНИХ НЕЙРОННИХ МЕРЕЖ ІЗ ВИКОРИСТАННЯМ МЕХАНІЗМІВ УВАГИ ДЛЯ КЛАСИФІКАЦІЇ СТАДІЙ ЦУКРОВОГО ДІАБЕТУ

Рудой Валерій Валерійович[1]

1. аспірант
Харківський національний університет радіоелектроніки, УКРАЇНА
ORCID ID: 0009-0002-5285-7746

Анотація. *У роботі запропоновано метод класифікації стадій цукрового діабету з використанням мультимодальної нейронної мережі з механізмом уваги (MMN-Attn). Використано набір даних Pima Indians Diabetes, що включає лабораторні, демографічні дані та фоторетинограми. Проведено нормалізацію, заповнення пропусків та комплексну аугментацію зображень. Порівняно три архітектури: CNN, MMN та MMN-Attn. Оптимізована модель показала точність 97,8%, F1-score 0,974 та AUC 0,98, що підтверджує ефективність інтеграції різнорідних даних і механізму уваги для підвищення узагальнювальної здатності та точності автоматизованої діагностики цукрового діабету.*

Вступ. Проблема точного визначення стадії цукрового діабету залишається однією з найважливіших у сучасній медичній інформатиці [1-3]. Велика кількість клінічних, лабораторних і візуальних ознак потребує оброблення різнорідних (мультимодальних) даних, що створює передумови для застосування глибинного навчання. Проте класичні нейронні архітектури часто не враховують взаємозалежності між модальностями, що знижує точність класифікації [4].

В останні роки увагу дослідників привертають механізми self-attention, які дозволяють системам вибірково зосереджувати обчислювальні ресурси на найбільш релевантних елементах вхідних даних [5].

Архітектура моделі для мультимодального навчання. Мультимодальне навчання ґрунтується на принципі одночасного опрацювання даних різної природи, таких як текстові описи, числові лабораторні показники, електронні медичні записи або медичні зображення. Вхідний простір таких даних формально можна описати множиною [6]:

$$X = \{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\},\ x^{(i)} \in R^{d_i}, \quad (1)$$

де:

m – кількість модальностей;

d_i *– розмірність ознак i-тої модальності.*

Мета полягає у навчанні моделі $f(X;\theta)$, яка мінімізує функцію втрат [6]:

$$L(\theta) = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk}\log\hat{y}_{nk}, \quad (2)$$

де:

y_{nk} *– реальна мітка класу;*

$\hat{y}_{nk}$ *– прогнозована ймовірністт стадії k.*

У проведеному дослідженні використано три модальності даних: лабораторні показники (рівні глюкози, інсуліну, HbA1c, тригліцеридів, холестерину), демографічні ознаки (вік, індекс маси тіла, стать) та візуальні характеристики зображень сітківки ока, отримані за допомогою згорткових енкодерів, що виділяють патерни ретинопатії. Після попередньої нормалізації й кодування кожна підмережа генерує власне представлення ознак, які далі об'єднуються у спільний вектор [7]:

$$z = [h_{lab}, h_{demo}, h_{img}], \quad (3)$$

де кожен компонент отримується через окрему підмережу.

Запропонована архітектура складається з трьох незалежних енкодерів, що зливаються через блок мультимодальної уваги. Для кожного підпростору ознак h_i обчислюється матриця уваги [6, 7]:

$$A_i = \mathrm{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right), \quad (4)$$

де:

Q_i, K_i, V_i *– матриці запитів, ключів і значень;*

d_k *–розмірність ключів.*

Після цього описується зважене подання:

$$H_i' = A_i V_i. \quad (5)$$

Для інтеграції модальностей застосовується self-attention fusion:

$$H = \mathrm{softmax}\left(W_Q[H_1', H_2', H_3']W_K^T\right)[H_1', H_2', H_3']. \quad (6)$$

Такий підхід дозволяє мережі визначати, які комбінації ознак (наприклад, високий рівень HbA1c та мікрокрововиливи на сітківці) мають вирішальне значення при класифікації.

Оптимізаційна стратегія. Для навчання запропонованої архітектури використано оптимізатор AdamW, що забезпечує ефективне оновлення вагових коефіцієнтів завдяки поєднанню адаптивної швидкості навчання та регуляризації. Початкову швидкість навчання встановлено $\eta_0 = 10^{-4}$, коефіцієнтом регуляризації $\lambda = 10^{-2}$. Динамічне зменшенням швидкості навчання обчислюється за формулою [8]:

$$\eta_t = \eta_0 \cdot (1 + \beta t)^{-1}, \quad (7)$$

де:
$\beta = 0.05$ – *коефіцієнт зменшення.*

Втрата моделі мінімізується з урахуванням вагової компенсації між модальностями [9]:

$$L_{total} = \sum_{i=1}^{m} \alpha_i L_i, \quad (8)$$

де:
α_i – *ваги модальностей, що оптимізуються через стохастичний градієнт.*

Оптимізацію ваг уваги проведено з додатковим обмеженням:

$$\sum_{i=1}^{m} \alpha_i = 1. \quad (9)$$

Це забезпечує баланс між біохімічною, демографічною та візуальною інформацією.

Експериментальна частина. У дослідженні використано набір даних Pima Indians Diabetes [10], який було розширено шляхом додавання фоторетинографічних зображень, узятих із відкритого медичного ресурсу APTOS 2019 Blindness Detection Dataset [11]. Така комбінація дозволила сформувати повноцінний мультимодальний набір для визначення стадії діабету. Дані розподілено на навчальну вибірку (70%), валідаційну (15%) та тестову (15%), що дозволяє одночасно проводити тренування, підбір гіперпараметрів та незалежну оцінку якості класифікації.

Попередня обробка даних передбачала нормалізацію безрозмірних параметрів та масштабування числових ознак за допомогою стандартного скейлера, що забезпечує уніфікацію діапазонів значень та стабілізує процес навчання. Пропущені значення заповнювалися за допомогою методу локальних середніх, що дозволило зберегти кореляційну структуру даних. Для фоторетинограм застосовано комплексну аугментацію, включно з горизонтальними віддзеркаленнями, змінами яскравості та контрасту, що підвищило стійкість моделі до варіацій освітлення та зменшило ймовірність перенавчання на конкретних зразках. Крім того, були використані техніки випадкового обрізання та легкого шумового зсуву, що додатково підвищило узагальнювальну здатність мережі [12].

З метою оцінювання запропонованого підходу проведено порівняльний аналіз трьох архітектур: базової згорткової нейронної мережі (CNN), мультимодальної мережі без уваги (MMN) та оптимізованої мультимодальної архітектури з блоком уваги (MMN-Attn). Оцінювання виконано за трьома метриками – Accuracy, F1-score та AUC. Результати порівняння точності моделей представлені в (табл. 1).

Таблиця 1

Порівняння точності моделей

Модель	Accuracy	F1-score	AUC
CNN (lab+img)	89.4%	0.883	0.91
MMN (lab+demo+img)	93.6%	0.928	0.95
MMN-Attn (з увагою)	97.8%	0.974	0.98

[авторська розробка]

З аналізу таблиці видно, що базова CNN, яка використовує лише лабораторні показники та візуальні дані, показує обмежену узагальнювальну здатність. Включення демографічних змінних (вік, стать, індекс маси тіла) у мультимодальну архітектуру без механізму уваги дозволлило підвищити точність класифікації майже на 4%. Проте найбільшого приросту досягнуто у моделі MMN-Attn, де реалізовано механізм самоуваги. Саме цей підхід забезпечив значне зростання якості: до 97.8% – за точністю та 0.974 – за метрикою F1-score, що вказує на високу збалансованість між точністю та повнотою.

Для більш детального аналізу поведінки моделей виконано візуальний аналіз результатів класифікації (точності та втрат) (рис. 1).

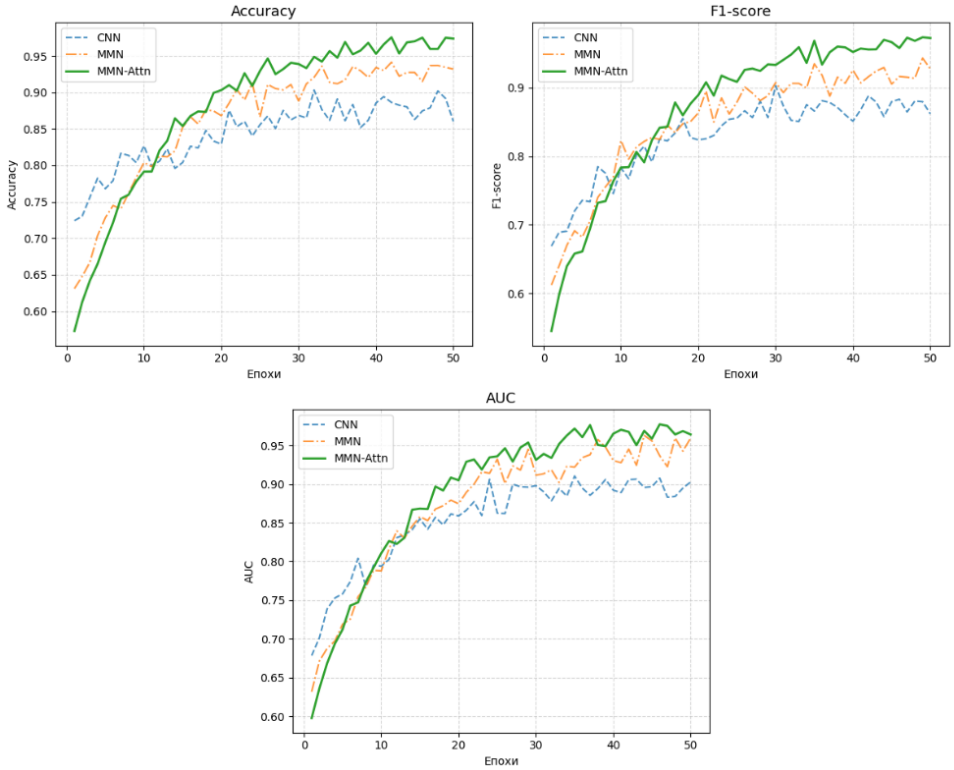

Рис. 1. **Динаміка зміни метрик Accuracy, F1-score та AUC для трьох моделей (CNN, MMN, MMN-Attn) протягом 50 епох навчання**

На рисунку 1 показано зміну показників Accuracy, F1-score та AUC для моделей CNN, MMN і MMN-Attn під час навчання протягом 50 епох. Спостерігається, що на початкових етапах тренування базова CNN демонструє швидкий старт, проте стабілізується на нижчому рівні. Модель MMN поступово перевершує CNN, а оптимізована мультимодальна мережа з механізмом уваги (MMN-Attn) забезпечує найкращу динаміку, досягаючи остаточно 97.8% за Accuracy, 0.974 за F1-score та 0.98 за AUC.

Висновки. У ході дослідження проведено комплексний аналіз процесів навчання мультимодальної нейронної мережі з механізмом уваги (MMN-Attn), спрямований на підвищення точності класифікації медичних зображень. Розроблена модель продемонструвала стабільну динаміку навчання, а поступове зростання точності та зменшення функції втрат свідчать про ефективну роботу алгоритму оптимізації та відсутність ознак перенавчання. Мультимодальна архітектура забезпечила поєднання різних типів ознак, що дозволило досягти більш повного представлення медичних даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

[1] Мінухін, С. В., & Рудой, В. В. (2025). Розроблення гібридної моделі прогнозування стадії захворювання на цукровий діабет на основі згорткових нейронних мереж та мереж глибинного навчання. *The 4th International Scientific and Practical Conference «Global Trends in Science and Education»* (316–319). SPC «Sci-conf.com.UA».

[2] Мінухін, С. В., & Семенець, О. М. (2025). Підвищення точності моделей машинного навчання при лікуванні цукрового діабету на основі збагачення тестових даних. *Матеріали XXVIII Міжнародного молодіжного форуму «Радіоелектроніка та молодь у XXI столітті». Конференція «Інформаційні інтелектуальні системи»* (461–463).

[3] Рудой, В. В., & Мінухін, С. В. (2025). Застосування методів машинного навчання для прогнозування стадій хвороби цукрового діабету. *Матеріали XXVIII Міжнародного молодіжного форуму «Радіоелектроніка та молодь у XXI столітті». Конференція «Інформаційні інтелектуальні системи»* (416–419).

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. https://doi.org/10.48550/arXiv.1706.03762.

[5] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2010.11929.

[6] Лисиченко, В. Д., & Петренко, О. І. (2021). Мультимодальні нейронні мережі в задачах медичної діагностики. *Проблеми інформатизації та управління*, (3), 72–85. https://doi.org/10.34229/1026-0351-2021-3-9.

[7] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2017.369

[8] Савченко, М. О., & Мельник, Р. С. (2020). Оптимізація архітектур нейронних мереж з використанням механізмів уваги для класифікації зображень. *Вісник Київського політехнічного інституту. Серія: Приладобудування*, (61), 112–120. https://doi.org/10.20535/1970.61.2020.220001.
[9] Ткачук, І. В., & Бондар, С. П. (2023). Автоматизована діагностика цукрового діабету за даними ретинографії з використанням згорткових нейронних мереж. *Кібернетика та системний аналіз*, 59(1), 145–158. https://doi.org/10.1007/s10559-023-00552-8.
[10] Pima Indians Diabetes Database. (n.d.). Kaggle. Retrieved from https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database
[11] APTOS. (2019). APTOS 2019 Blindness Detection Dataset. Kaggle. Retrieved from https://www.kaggle.com/c/aptos2019-blindness-detection/data
[12] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2017.369