

**УДК 004.85**

**В. Шипка**

(Тернопільський національний університет імені Івана Пулюя, Україна)

## **АПАРАТНО-АДАПТИВНИЙ ПІДХІД ДО ОПТИМІЗАЦІЇ ОНЛАЙН-НАВЧАННЯ НА ПРИБОРАХ TINYML**

**UDC 004.85**

**V. Shypka**

## **HARDWARE-ADAPTIVE APPROACH TO OPTIMIZATION OF ONLINE LEARNING ON TINYML DEVICES**

Інструмент TinyML (Tiny Machine Learning) дозволяє виконувати моделі штучного інтелекту на мікроконтролерах з низьким споживанням потужності. Це відкриває можливості для створення нового покоління «розумних» автономних пристроїв. Проте, більшість сучасних рішень зосереджені на статичному виконанні попередньо навчених моделей. Можливість онлайн-навчання моделей штучного інтелекту, тобто адаптації моделі до нових даних безпосередньо на пристрої, залишається актуальним через жорсткі обмеження пам'яті та обчислювальної потужності.

Стандартні методи оптимізації, такі як прунінг та квантизація, ефективно зменшують розмір моделей для виконання. Однак вони не призначені для процесу навчання, який вимагає обчислення градієнтів та збереження проміжних активацій, що є нетривіальною задачею для мікроконтролерів. Існуючі «апаратно-незалежні» (hardware-agnostic) підходи до онлайн-навчання [1] часто призводять до неефективного використання ресурсів, оскільки ігнорують специфіку архітектури конкретного мікроконтролера.

Процес оптимізації відбувається не в загальному, а з урахуванням унікальних характеристик цільової платформи.

Метод базується на моделі, яка враховує: час, пам'ять, обчислень для конкретної архітектури. Перед початком донавчання алгоритм аналізує апаратні обмеження та динамічно визначає оптимальну комбінацію стиснення, яка максимізує точність при мінімальних витратах ресурсів.

Проведення обчислювального експерименту необхідно здійснити на реальних мікроконтролерних платах, провести порівняльний аналіз запропонованого методу з базовими підходами за такими критеріями:

- Використання пікової SRAM-пам'яті під час циклу донавчання (КБ).
- Час, необхідний для одного циклу донавчання (мс).
- Кінцева точність моделі після N циклів адаптації (%).

В результаті отримані дані дозволять створити автономні та адаптивні апаратні пристрої. Такий підхід дозволить скоротити вимоги до оперативної пам'яті та прискорити процес донавчання, при збереженні високої точності моделі. Розробка такого підходу є актуальною задачею, що дозволить створювати по-справжньому автономні та адаптивні інтелектуальні пристрої.

### **Література**

1. Cai H., Gan C., Han S. TinyOL: TinyML with Online-Learning on Microcontrollers. arXiv preprint arXiv:2104.10260. 2021. URL: <https://arxiv.org/abs/2104.10260> (дата звернення: 06.11.2025).
2. David R., Duke J., Jain A. et al. TensorFlow Lite Micro: Embedded machine learning on tiny-RAM-T-devices. Proceedings of Machine Learning and Systems. 2021. Vol. 3. P. 800–811.