

ДОСЛІДЖЕННЯ АЛГОРИТМІВ РОЗПІЗНАВАННЯ РУКОПИСНИХ СИМВОЛІВ РІЗНИХ МОВ ЗА ДОПОМОГОЮ МОДЕЛІ НЕЙРОННИХ МЕРЕЖ KAN

У роботі були проаналізовані найефективніші існуючі методи оптичного розпізнавання символів, які використовують у своїй структурі глибоке навчання нейромереж. У ході аналізу було виявлено, що сучасні архітектури нейромереж, з найкращими показниками точності розпізнавання, мають сталу межу точності. Також було встановлено, що кожна проаналізована архітектура нейромереж містить у своїй структурі багатошаровий перцептрон. Для оптимізації показників розпізнавання нейромереж було запропоновано використовувати мережу Колмогоров-Арнольда як альтернативу для мереж на основі багатошарового перцептрону. Архітектура створеної моделі заснована на двокомпонентному трансформері, перший компонент – візуальний трансформер, що використовується у якості кодера, другий – мовний трансформер, який використовується у якості декодера. Мережа Колмогоров-Арнольда замінює мережу прямого поширення на основі багатошарового перцептрону, у кожному трансформері – кодері та декодері. Поліпшення результатів існуючих нейромереж забезпечується за допомогою трансферного навчання, для чого у створеній моделі використовуються групові раціональні функції, у якості основних елементів мережі Колмогоров-Арнольда, що навчаються. Навчання моделі проводилося на наборах зображень рядків тексту з трьох різних систем письма: алфавітної, абугіда та логографічної; які представлені писемностями: англійської, деванатарі та китайської. В результаті експериментальних досліджень були виявлені високі показники розпізнавання символів для наборів даних китайської та деванатарі, але низькі для англійської писемності, для моделі з мережею Колмогоров-Арнольда. Отримані результати свідчать про нові можливості підвищення надійності і ефективності сучасних систем розпізнавання рукописних текстів.

Ключові слова: оптичне розпізнавання символів, нейромережа, мережа Колмогоров-Арнольда, трансформерна архітектура, трансформер Колмогорова-Арнольда, раціональні функції.

A.V. Serhiienko, E.A. Kolomoichenko. Study of handwritten character recognition algorithms for different languages using the KAN Neural Network Model. The paper analyzed the most effective existing methods of optical character recognition that use deep learning neural networks in their structure. The analysis revealed that modern neural network architectures with the best recognition accuracy indicators have a constant accuracy limit. It was also found that each analyzed neural network architecture contains a multilayer perceptron in its structure. To optimize the recognition performance of neural networks, it was proposed to use the Kolmogorov-Arnold network as an alternative to multilayer perceptron based networks. The architecture of the created model is based on a two-component transformer, the first component is a visual transformer used as an encoder, the second is a language transformer used as a decoder. The Kolmogorov-Arnold network replaces the feedforward network based on a multilayer perceptron, in each transformer – encoder and decoder. Improvement of existing neural network results is ensured through transfer learning, for which group rational functions are used as the main learning elements of the Kolmogorov-Arnold network. The model was trained on sets of images of text

¹ канд. техн. наук, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: 0000-0003-1328-2572, sergienko_a_v@pstu.edu

² магістр, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, kolomoichenko_e_a@pstu.edu

lines from three different writing systems: alphabetic, abugida and logographic; which are represented by the scripts: English, Devanagari and Chinese. As a result of experimental studies, high character recognition rates were found for the Chinese and Devanagari data sets but low for the English script, for the model with the Kolmogorov-Arnold network. The obtained results indicate new possibilities for increasing the reliability and efficiency of modern handwriting recognition systems.

Keywords: optical character recognition, neural network, Kolmogorov-Arnold network, transformer architecture, Kolmogorov-Arnold transformer, rational functions.

Постановка проблеми. Розпізнавання рукописних символів (скорочено HWR) є ключовою галуззю досліджень у сфері комп'ютерного зору та штучного інтелекту. Дослідження у галузі HWR спрямовані на створення алгоритмів та систем, здатних автоматично інтерпретувати рукописні символи та перетворювати їх у цифровий формат.

В загальному випадку розпізнавання рукописного тексту поділяється на дві основні категорії: офлайн та онлайн розпізнавання. Офлайн розпізнавання передбачає обробку статичних зображень рукописного тексту. У цьому випадку система розпізнавання не має доступу до інформації про динаміку написання символів, такої як траєкторія руху пера або швидкість письма [Помилка! Джерело посилання не знайдено.]. Ці динамічні характеристики можуть допомогти розрізнити схожі за формою символи або вирішити неоднозначності, які виникають при аналізі лише статичних зображень і тому процес офлайн розпізнавання є більш складним. У роботі було прийнято рішення досліджувати методи офлайн розпізнавання, через їх ширше застосування та більшу алгоритмічну складність.

Сучасний стан області досліджень розпізнавання рукописного тексту базується на застосуванні машинного навчання [2] та глибокого навчання нейромереж [3], що являють собою моделі багат шарового перцептрона, зокрема згорткових нейронних мереж, які забезпечують автоматичне виділення ознак та демонструють значно кращу точність порівняно з традиційними алгоритмами [4].

Нові дослідження у сфері розпізнавання рукописних символів спрямовані на підвищення точності, ефективності та універсальності існуючих систем, з особливим акцентом на адаптивність до мов, які ще не мають високого рівня розпізнавання у існуючих системах, і до різноманітних стилів письма.

Метою даної роботи є аналіз та вдосконалення сучасних алгоритмів розпізнавання рукописних символів різних мов з використанням інноваційних моделей нейронних мереж Колмогорова-Арнольда.

Аналіз останніх досліджень і публікацій. При аналізі методів розпізнавання рукописних символів доцільно зосередитись на сучасних ефективних методах, що засновані на штучному інтелекті та машинному навчанні. Зокрема, найкраще себе зарекомендували методи на основі навчання нейромереж, що базуються на багат шаровому перцептроні, але з різною архітектурою побудови мережі.

Загальний підхід при роботі з методами на основі машинного навчання нейромереж включає кілька ключових етапів. Перш за все, здійснюється збір та підготовка великих та різноманітних наборів даних рукописних символів обраних мов, що забезпечує якісне навчання моделей. Наступним етапом є попередня обробка даних, яка включає нормалізацію зображень, видалення шуму, вирівнювання та інші техніки для покращення якості вхідних даних. Після цього проводиться вибір архітектури моделі, де визначаються найбільш підходящі структури моделі для машинного навчання. Далі слідує етап навчання моделі, під час якого використовуються алгоритми оптимізації для налаштування параметрів моделі на основі підготовлених даних з метою мінімізації помилок розпізнавання. Після навчання модель проходить етап оцінки та валідації, де тестується на окремих вибірках даних для оцінки її точності, здатності до узагальнення та стійкості до різних варіацій рукопису. Заключним етапом є оптимізація та налаштування моделі, що включає внесення корективів до архітектури або процесу навчання для покращення продуктивності та адаптивності до специфічних вимог задачі [5]. Такий системний підхід дозволяє створювати гнучкі та високоефективні системи розпізнавання рукописних символів, здатні обробляти різноманітні мовні та стилістичні варіації, що є ключовим фактором успішного впровадження в реальні застосунки.

Найбільш поширеним методом для розпізнавання рукописних символів є використання нейромереж з архітектурою CNN-LSTM, де згорткові шари (CNN) ефективно виділяють локальні ознаки зображень, а операція пулінгу підвищує стійкість до геометричних перетворень [6]. Використання пакетної нормалізації та додаткових методів регуляризації покращує збіжність і генералізацію моделі [7], а застосування трансформаційно-інваріантних шарів підвищує її стійкість до різних варіацій вхідних даних [8]. Водночас згорткові шари обмежені у захопленні глобального контексту, що компенсується рекурентними шарами LSTM, здатними зберігати інформацію протягом тривалих інтервалів та долати проблему зникання градієнтів [9, 10]. Для навчання на послідовностях змінної довжини, без явного вирівнювання між вхідними та вихідними даними, використовується функція втрат CTC, яка враховує всі можливі вирівнювання та покращує ефективність розпізнавання рукописних символів [11].

Альтернативним методом вирішення задачі розпізнавання рукописного тексту є навчання нейромережі з архітектурою трансформеру. Ця архітектура стала революційною у галузі обробки природної мови та комп'ютерного зору завдяки здатності обробляти великі послідовності у постійній кількості операцій, одночасно забезпечуючи ефективний паралелізм. Центральним елементом трансформеру є механізм уваги, особливо багатоголова увага, що дозволяє моделі одночасно фокусуватися на різних частинах вхідної інформації [12, 13]. Механізм багатоголової уваги паралельно застосовує кілька незалежних голів уваги, кожна з яких виконує власні перетворення вхідних даних через лінійні проєкції запитів, ключів та значень. Це забезпечує вивчення різних аспектів залежностей у даних, що підвищує здатність моделі розпізнавати неоднорідний контекст рукописних символів. Архітектура трансформеру також використовує позиційні кодування, аби зберегти інформацію про порядок елементів послідовності, оскільки сама модель не має вбудованого механізму для цього [14]. Додатково застосовуються шари нормалізації, залишкові зв'язки та feed-forward шари, які покращують стабільність тренування та виразну здатність моделі.

Виклад основного матеріалу. Використання трансформерної архітектури у різних задачах комп'ютерного зору, зокрема і в розпізнаванні рукописних символів, показало високі результати. Однак майже усі останні дослідження, які направлені на покращення ефективності роботи мереж на основі трансформерів, були зосереджені на поліпшенні механізму уваги [15, 16]. У той час, як майже не була розглянута заміна для іншого компонента мережі – багат шарового перцептронну. Лише зовсім нещодавні дослідження використовують інноваційну архітектуру – мережу Колмогоров-Арнольда на заміну MLP [17].

По суті, KAN представляє нову теорію навчання для тренування нейронних мереж, засновану на теоремі Колмогорова-Арнольда про представлення неперервної функції з кількома параметрами на обмеженому інтервалі – формула 1.

$$f(x) = f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{p,q}(x_p) \right), \quad (1)$$

де кожна функція $\phi_{p,q}$ – є дійснозначною функцією, визначеною на інтервалі від 0 до 1 включно, а Φ_q є неперервною дійсною функцією, що приймає будь які дійсні значення.

Основна зміна процесу навчання полягає в навчанні функцій активації замість ваг у звичайному MLP. Також у KAN кожен ваговий параметр замінюється одновимірною функцією, зазвичай параметризованою як сплайн. Як наслідок, KAN взагалі не мають лінійних ваг. Вузли в KAN просто підсумовують вхідні сигнали без застосування будь-яких нелінійностей. На рисунку 1 [17] зображені основні відмінності моделі MLP від KAN.

На основі розширених експериментів із використанням синтетичних математичних і фізичних наборів даних KAN показав [17]:

- покращену точність навчених моделей порівняно з аналогами MLP;
- покращену інтерпретацію, завдяки демонстрації його здатності витягувати символічні зв'язки, на основі яких був синтезований набір даних;
- покращений розмір навчених параметрів, де моделям KAN потрібна менша кількість параметрів, щоб досягти того самого рівня точності, що й моделям MLP;

– вивчення нового завдання попередньо навченої моделі KAN на старому завданні менше впливає на ефективність старого завдання, ніж у випадку MLP, коли мережа може «забути» старе завдання.

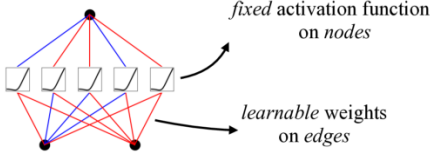
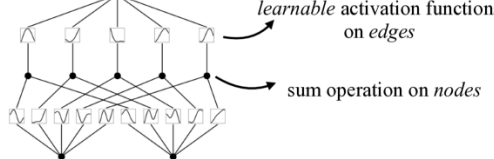
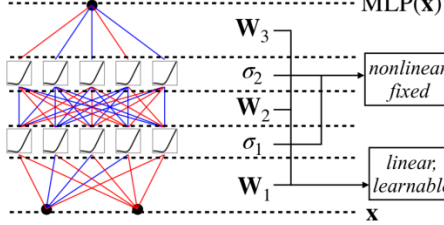
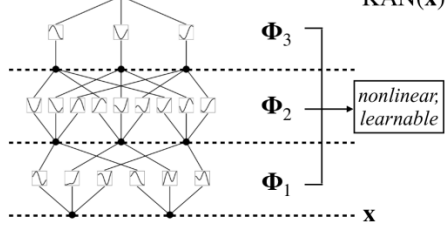
Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(e)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a)  fixed activation functions on nodes learnable weights on edges	(b)  learnable activation functions on edges sum operation on nodes
Formula (Deep)	$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\text{KAN}(\mathbf{x}) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(\mathbf{x})$
Model (Deep)	(c)  $\mathbf{W}_3$ σ_2 $\mathbf{W}_2$ σ_1 $\mathbf{W}_1$ $\mathbf{x}$ nonlinear, fixed linear, learnable	(d)  Φ_3 Φ_2 Φ_1 $\mathbf{x}$ nonlinear, learnable

Рис. 1 – Порівняння моделей MLP та KAN [17]

Основними недоліками KAN є:

1) Базова функція. Стандартні функції B-spline у KAN не ідеальні для архітектур паралельних обчислень, типових для сучасних GPU. B-сплайни вимагають рекурсивних обчислень, що значно сповільнює навіть найбільш оптимізовані реалізації.

2) Неефективність параметрів і обчислень. Через те, що для кожної унікальної пари вводу-виводу в KAN потрібен окремий набір параметрів і базових функцій. Ця необхідність викликає експоненціальне зростання кількості параметрів у міру збільшення прихованого розміру мережі, що призводить до значних обчислювальних витрат і проблем з масштабованістю.

3) Ініціалізація ваги в KAN подібна до ініціалізації в MLP, але вона не відповідає потребам KAN щодо конвергенції. Ця невідповідність може призвести до нестабільності та погіршення продуктивності під час тренувального процесу.

У статті, що описує KAT [18], пропонується вирішення зазначених проблем реалізації KAN. А саме, запропоновано:

- 1) використовувати раціональні функції активації;
- 2) використовувати «груповий KAN», тобто розподіляти функціональні коефіцієнти та базові функції між групами ребер;
- 3) використовувати ініціалізацію із збереженням дисперсії.

Зазначені методи – покращують обчислювальну ефективність і стабільність навчання. Раціональні функції активації забезпечують сумісність з сучасними GPU, що прискорює тренування моделі. Груповий KAN зменшує обчислювальне навантаження без втрати продуктивності, що є критичним при обробці великих обсягів рукописних даних. Ініціалізація із збереженням дисперсії покращує стабільність моделі під час навчання, що може призвести до більш точних результатів. І хоча трансформери на основі мереж Колмогорова-Арнольда ще не використовувалися для розпізнавання рукописних символів, їх інтеграція в цю область може відкрити нові можливості. Трансформер Колмогорова-Арнольда успішно поєднує KAN із трансформерами і показує кращі результати у задачах розпізнавання зображень, виявлення об'єктів і семантичної

сегментації у порівнянні з класичними трансформерами [18]. Потенційне застосування КАТ у розпізнаванні рукописних символів може покращити продуктивність моделей, надаючи їм більш виразні можливості порівняно з традиційними MLP у трансформерах.

Архітектура створеної у роботі моделі являє собою двокомпонентний трансформер, що включає у себе візуальний КАТ у якості кодера та мовний КАТ у якості декодера. Архітектура кодера зображена на рис. 2, та декодера – на рис. 3.

Згідно з рис. 2, основними елементами моделі візуального КАТ, є: шар вбудовування вхідного зображення та L блоків, які, у свою чергу, складаються з: багатоголової уваги, шарової нормалізації та мережі прямого поширення на основі групових раціональних функцій. Отримані дані з кодера являють собою послідовність ознак $T1$, яка передається у декодер (рис. 3).

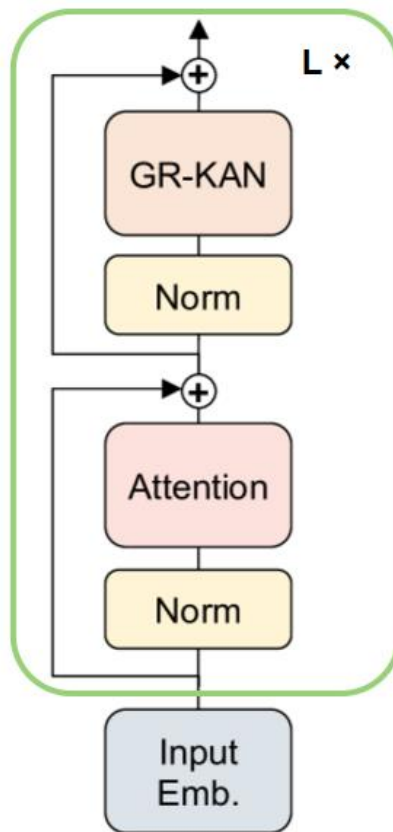


Рис. 2 – Архітектура візуального КАТ як моделі для вилучення ознак

Згідно з рис. 3, основними елементами моделі мовного КАТ є ті самі елементи, що використовуються у візуальному КАТ, але додатково використовується перехресна багатоголова увага між даними ($T1$), отриманими з декодера, та даними, отриманими з попереднього шару, – багатоголової самоуваги.

Перевагою використання дво-трансформерної архітектури створеної системи є спільні елементи в кодері та декодері. У обох трансформерах (рис. 2, та рис. 3) отриманні дані з багатошарової уваги ($x_0^{(l)}$) передаються у мережу прямого поширення на основі KAN. Мережа прямого поширення, у свою чергу, складається з декількох шарів групової раціональної мережі Колмогорова-Арнольда. KAN на основі раціональної функції з груповим обчисленням параметрів визначається за формулою 2.

$$\text{GR-KAN}(x) = W_{\text{GR-KAN}} F(x) = \begin{bmatrix} w_{1,1} & \dots & w_{1,d_{\text{in}}} \\ \vdots & \ddots & \vdots \\ w_{d_{\text{out}},1} & \dots & w_{d_{\text{out}},d_{\text{in}}} \end{bmatrix} \times \left[F_{[1/d_g]}(x_1) \dots F_{[d_{\text{in}}/d_g]}(x_{d_{\text{in}}}) \right]^T, \quad (2)$$

де:

- W_{GR-KAN} є вагами, що навчаються та визначається за формулою 3;
- $F(x)$ це раціональна функція, що визначається за формулою 4 [19].

$$W_{GR-KAN} \in \mathbb{R}^{d_{in} \times d_{out}}, \quad (3)$$

$$\phi(x) = wF(x) = w \frac{P(x)}{Q(x)} = w \frac{\sum_{j=0}^m a_j x^j}{1 + |\sum_{i=1}^n b_i x^i|}, \quad (4)$$

де:

- w – коефіцієнт масштабування;
- a_j, b_i, w – це параметри, що навчаються;
- m, n – параметри функції, у роботі прийняті значення $m = 5, n = 4$.

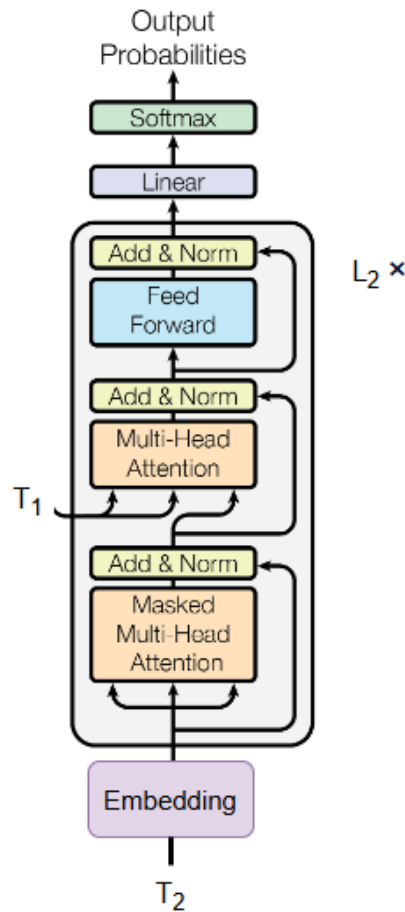


Рис. 3 – Архітектура мовного КАТ у якості декодера загальної моделі

У початковій моделі ViT мережа прямого поширення складається з двох лінійних шарів з двома функціями активації. Тому, у формулі обчислення виходу мережі прямого поширення на основі KAN (формула 5), використовуються 2 однакові шари GR-KAN. Після чого отримані данні передаються на початок наступного блоку, що знову обчислює данні за формулою 2 та 5 доти, доки данні не пройдуть через кожен блок.

$$x_l = GR - KAN \left(GR - KAN \left(LayerNorm \left(x_0^{(l)} \right) \right) \right) + x_0^{(l)}, [KAT]. \quad (5)$$

Для навчання створеної моделі використовувалися три датасети з трьох різних письмовностей: китайської, англійської, деванагарі. Кожен датасет містить зображення рядків тексту та

відповідну текстову транскрипцію зображених символів.

Модель KAN-Transformer спершу навчалась на китайському датасеті CASIA-HWDB [20], де за 34 епохи втрати знизилися з 3,5 до 0,6, графік втрат зображено на рис. 4.

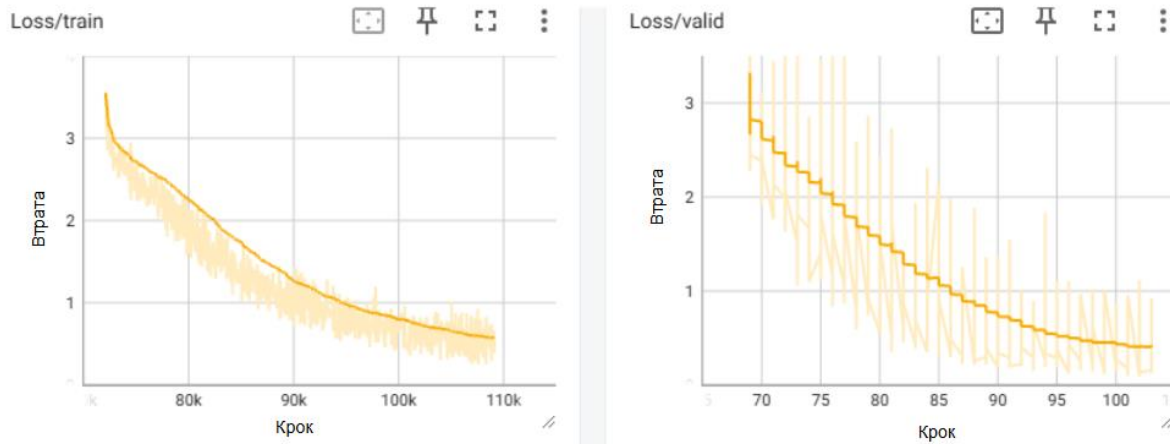


Рис. 4 – Графік втрат на навчальному (графік зліва) і перевірчному (справа) наборі китайських символів, з першого етапу навчання

Далі, з метою перевірки стійкості моделі до забування, її продовжили тренувати на деванагарі датасеті Devanagari Handwritten Dataset [21] до досягнення схожого рівня втрат (0,6), а потім на обох наборах одночасно. Результати показали, що модель не забуває попередні дані, оскільки середні втрати для обох датасетів залишилися на рівні 0,6. Навчання на двох наборах даних одночасно тривало до досягнення точності приблизно 0,25. Перевірка на валідаційних даних засвідчила подальшу можливість покращення моделі, не зважаючи на те, що темп зниження втрат сповільнився з часом. Графік втрат на цьому етапі навчання наведено на рис. 5.

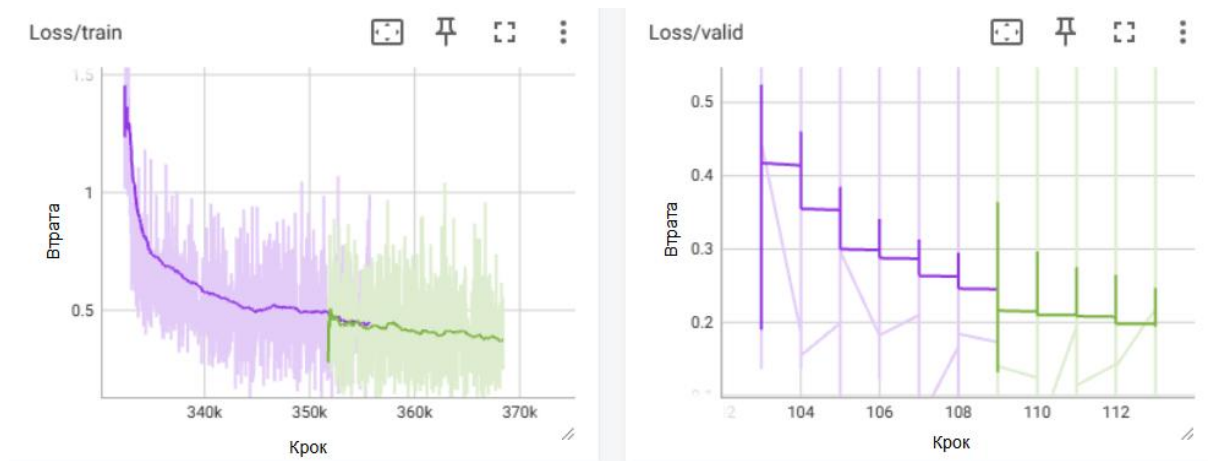


Рис. 5 – Графік втрат на навчальному (графік зліва) і перевірчному (справа) наборах символів деванагарі окремо (фіолетовим) й деванагарі з китайськими разом (зеленим), з другого етапу навчання

Наступним етапом було навчання моделі на англійському датасеті IAM HANDWRITING DATABASE [22]. У перших частинах навчання, графік втрат якої зображено на рис. 6, втрати знижувались, але згодом різниця стала незначною (графік втрат зображено на рис. 7), а на перевірчному наборі процес покращення зупинився, що свідчило про перенавчання. Для покращення було додано тестові приклади до тренувального набору, проте після досягнення мінімальних втрат (0,52) подальшого поліпшення не відбулось.

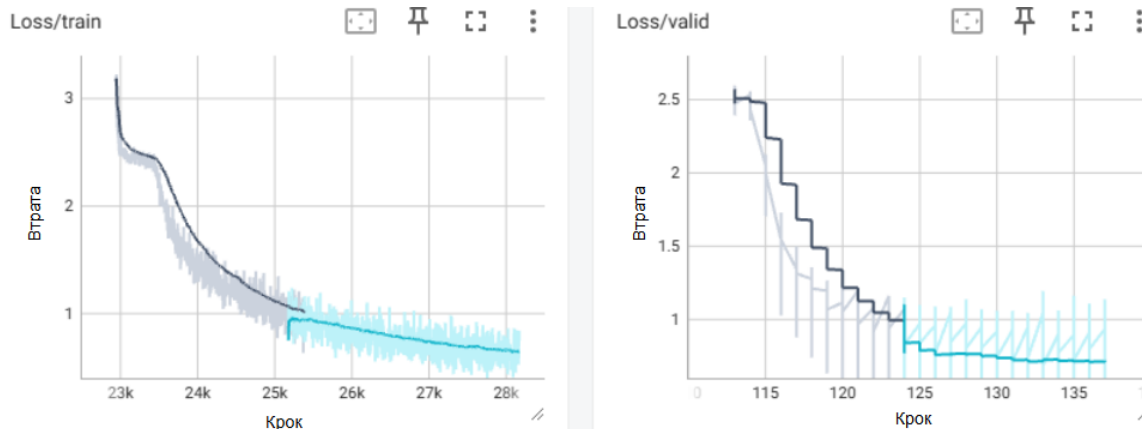


Рис. 6 – Перша частина графіків втрат на навчальному (графік зліва) і перевірчому (справа) наборі символів англійської писемності

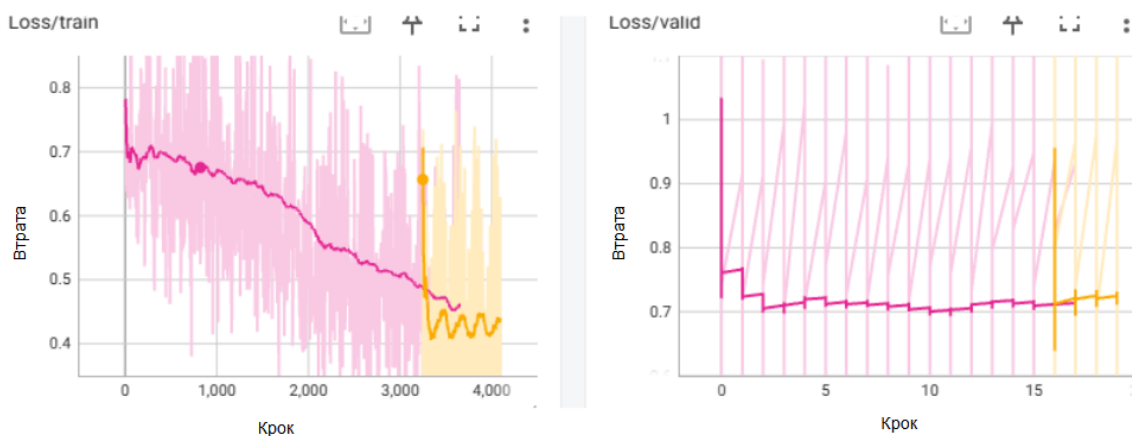


Рис. 7 – Друга частина графіків втрат на навчальному (графік зліва) і перевірчому (справа) наборі символів англійської писемності

Графік втрат останнього етапу навчання, після додавання нових даних у набір IAM, зображено на рис. 8.

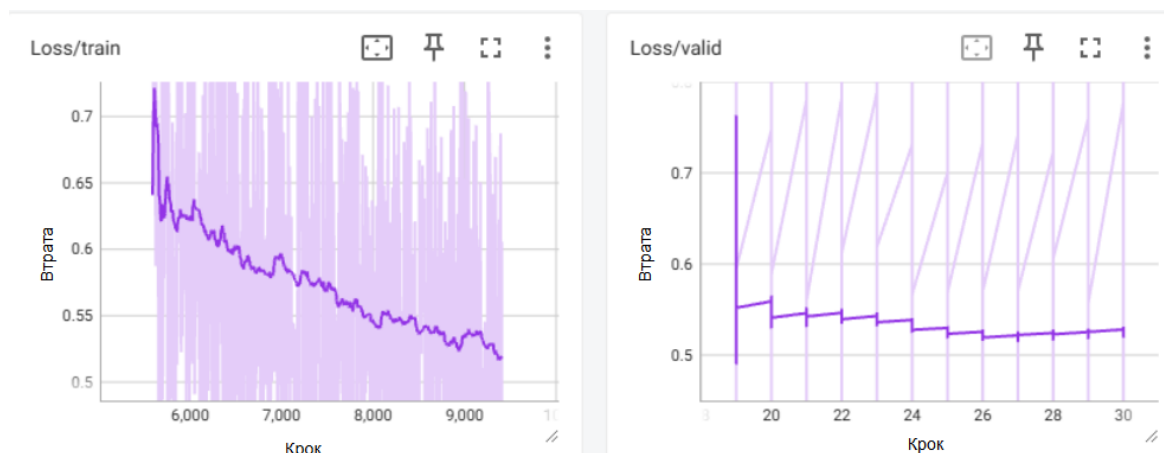


Рис. 8 – Графік втрат на навчальному (графік зліва) і перевірчому (справа) наборі символів англійської писемності, після додавання нових даних у навчальний набір

Слід зазначити, що для набору символів англійської писемності, кількість унікальних зображень рядків тексту є найменшою серед використаних наборів і дорівнює близько 10 тисяч, у той час як для інших використаних датасетів ця кількість становить приблизно 60 тис. унікальних зображень.

Створена модель була оцінена за ключовими метриками: CER (Character Error Rate), WER (Word Error Rate) та F1-score на рівні слів. Результати оцінювання моделі наведено у таблиці 1.

Таблиця 1

Порівняння оцінки створеної моделі за різними метриками для символів різних мов

Метрика	Оцінка точності для символів		
	китайська	деванагарі	англійська
CER	0,14	0,093	0,22
WER	0,14	0,3	0,49
F1	0,93	0,7	0,63

Модель найкраще розпізнавала китайський текст, порівняння оцінки f1 наведено у табл. 2 [23].

Таблиця 2

Порівняння різних моделей на рукописному тексті з набору китайських символів

Модель	F1 оцінка у відсотках
CRNN	48,0
ASTER	45,9
MORAN	30,2
SAR	31,0
SEED	21,1
MASTER	18,5
ABINET	13,8
TrOCR	51,7
Mask _{BASE}	63,7
DTrOCR	81,4
KAN-Transformer (модель, що була створена у результаті роботи)	93,0

Порівняння з іншими архітектурами (таблиця 2), показало, що створена модель KAN-Transformer демонструє найвищі результати для китайської мови (F1 = 93,0%). Також з таблиці 2 видно, що більшість моделей має точність розпізнавання близьку до 51 відсотка. Найкращим результатом, у попередніх дослідженнях, було досягнуто 81,4% на моделі DTrOCR.

Для датасету символів англійської писемності отримані результати розпізнавання гірші за точність розпізнавання існуючих моделей, у яких CER вимірюється від 0,1 до 0,028 [23].

У випадку датасету деванагарі писемності результати розпізнавання створеної моделі є достатньо ефективними. Однак порівняння точності між існуючими архітектурами неможливо провести через відсутність достатньої кількості навчених моделей HWR на деванагарі символах.

Висновки

На основі проведеного дослідження алгоритмів розпізнавання рукописних символів різних мов за допомогою моделі нейронних мереж KAN сформульовано основні висновки. Було детально проаналізовано сучасні методи оптичного розпізнавання символів на основі навчання нейронних мереж, встановлено, що навіть найсучасніші архітектури досягають певної межі точності розпізнавання, що обмежує їх подальшу ефективність. З'ясовано, що більшість існуючих архітектур базуються на багатошаровому перцептроні, що може призводити до обмежень у точності через конструктивні особливості. Для подолання цих обмежень запропоновано використовувати мережу Колмогорова-Арнольда як альтернативу MLP. Реалізована модель нейронної мережі з

архітектурою візуального КАТ та мовного КАТ продемонструвала зростання точності розпізнавання символів китайської письменності на 11% порівняно з найкращою існуючою моделлю, високі показники CER та F1 (CER = 0,093, F1 = 0,93), а також показала, що для навчання моделі необхідно велике число унікальних зображень з різним рукописним текстом. Існує кілька напрямків майбутніх досліджень, що включають дослідження застосування KAN в інших задачах комп'ютерного зору, дослідження оптимізації параметрів моделі для наборів даних, що не були представлені у роботі, а також дослідження використання хмарних обчислень у контексті підвищення швидкості роботи створеної моделі.

Перелік використаних джерел:

1. Plamondon R., Srihari S. N. Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000. Vol. 22. Iss. 1. Pp. 63-84. DOI: <https://doi.org/10.1109/34.824821>.
2. Handwritten Recognition Techniques: A Comprehensive Review / H.A. Alhamad et al. *Symmetry*. 2024. Vol. 16. Article 681. DOI: <https://doi.org/10.3390/sym16060681>.
3. A Deep Learning Approach for Robust, Multi-oriented, and Curved Text Detection / R. Ranjbarzadeh et al. *Cognitive Computation*. 2024. Vol. 16. Pp. 1979-1991. DOI: <https://doi.org/10.1007/s12559-022-10072-w>.
4. Convolutional-Neural-Network-Based Handwritten Character Recognition: An Approach with Massive Multisource Data / Saqib N., Haque K.F., Yanambaka V.P., Abdelgawad A. *Algorithms*. 2022. Vol. 15. Article 129. DOI: <https://doi.org/10.3390/a15040129>.
5. Alzubi J., Anand N., Akshi K. Machine learning from theory to algorithms: an overview. *Journal of physics: conference series*. 2018. Vol. 1142. Pp. 1-15. DOI: <https://doi.org/10.1088/1742-6596/1142/1/012012>.
6. Zeiler M.D., Fergus R. Visualizing and understanding convolutional networks. *Computer Vision – ECCV 2014: Proceedings of the 13th European Conference, Zurich, Switzerland, 6-12 September 2014*. Pp. 818-833. DOI: https://doi.org/10.1007/978-3-319-10590-1_53.
7. Schilling F. The effect of batch normalization on deep convolutional neural networks : thesis of Master of Science – Computer Science. Stockholm, 2016. 102 p.
8. Laptev D., Buhmann J. M. Transformation-invariant convolutional jungles. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 07-12 June 2015. 2015. Pp. 3043-3051. DOI: <https://doi.org/10.1109/CVPR.2015.7298923>.
9. Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness / Ribeiro A. H., Tiels K., Aguirre L. A., Schön T. *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, Palermo, Italy, 26-28 August 2020. Vol. 108. Pp. 2370-2380. DOI: <https://doi.org/10.48550/arXiv.1906.08482>.
10. Hochreiter S., Schmidhuber J. Long Short-Term Memory. *Neural Computation*. 1997. Vol. 9(8). Pp. 1735-1780. DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>.
11. End-to-end sequence labeling via convolutional recurrent neural network with a connectionist temporal classification layer / X. Huang et al. *International Journal of Computational Intelligence Systems*. 2020. Vol. 13. Iss. 1. Pp. 341-351. DOI: <https://doi.org/10.2991/ijcis.d.200316.001>.
12. HTR-VT: Handwritten text recognition with vision transformer / Li Y., Chen D., Tang T., Shen X. *Pattern Recognition*. 2024. Vol. 158. Article 110967. DOI: <https://doi.org/10.1016/j.patcog.2024.110967>.
13. A light transformer-based architecture for handwritten text recognition / Barrere K., Soullard Y., Lemaitre A., Coüasnon B. *International Workshop on Document Analysis Systems*. 2022. Pp. 275-290. DOI: https://doi.org/10.1007/978-3-031-06555-2_19.
14. Attention is all you need / A. Vaswani et al. *NIPS 2017 : 31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 5-10 December 2017. Pp. 1-11. DOI: <https://doi.org/10.48550/arXiv.1706.03762>.
15. Swin transformer: Hierarchical vision transformer using shifted windows / L. Ze et al. *Proceedings of the IEEE/CVF International conference on computer vision*, Las Vegas, Nevada, USA, 10-17 Oct. 2021. Pp. 9992-10002. DOI: <https://doi.org/10.1109/ICCV48922.2021.00986>.

16. Mlp-mixer: An all-mlp architecture for vision / I.O. Tolstikhin et al. *NIPS'21 : Proceedings of the 35th International Conference on Neural Information Processing Systems*, 6-14 December 2021. Pp. 24261-24272. DOI: <https://doi.org/10.48550/arXiv.2105.01601>.
17. Kan: Kolmogorov-Arnold networks / L. Ziming et al. 2024. Pp. 1-50. DOI: <https://doi.org/10.48550/arXiv.2404.19756>.
18. Yang X., Wang X. Kolmogorov-Arnold Transformer. 2024. Pp. 1-19. DOI: <https://doi.org/10.48550/arXiv.2409.10594>.
19. Molina A., Schramowski P., Kersting K. Padé Activation Units: End-to-end Learning of Flexible Activation Functions in Deep Networks. *International Conference on Learning Representations*, 26 April - 26 May 2020. Pp. 1-17. DOI: <https://doi.org/10.48550/arXiv.1907.06732>.
20. Online and offline handwritten Chinese character recognition: benchmarking on new databases / Liu C.-L., Yin F., Wang D.-H., Wang Q.-F. *Pattern Recognition*. 2013. Vol. 46. Iss. 1. Pp. 155-162. DOI: <https://doi.org/10.1016/j.patcog.2012.06.021>.
21. Offline Handwriting Recognition on Devanagari using a new Benchmark Dataset / Dutta K., Krishnan P., Mathew M., Jawahar C. V. *13th IAPR International Workshop on Document Analysis Systems (DAS)*, Vancouver, Canada, 24-27 April 2018. Pp. 25-30. DOI: <https://doi.org/10.1109/DAS.2018.69>.
22. Marti U-V., Bunke H. The IAM-database: an English sentence database for offline handwriting recognition. *International journal on document analysis and recognition*. 2002. Vol. 5. Pp. 39-46. DOI: <https://doi.org/10.1007/s100320200071>.
23. Fujitake M. Dtrocr: Decoder-only transformer for optical character recognition. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, 03-08 January 2024. Pp. 8010-8020. DOI: <https://doi.org/10.1109/WACV57701.2024.00784>.

References:

1. R. Plamondon, and S.N. Srihari, «Online and off-line handwriting recognition: a comprehensive survey», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, iss. 1, pp. 63-84, 2000. doi: **10.1109/34.824821**.
2. H.A. Alhamad et al., «Handwritten Recognition Techniques: A Comprehensive Review», *Symmetry*, vol. 16, article 681, 2024. doi: **10.3390/sym16060681**.
3. R. Ranjbarzadeh et al., «A Deep Learning Approach for Robust, Multi-oriented, and Curved Text Detection», *Cognitive Computation*, vol. 16, pp. 1979-1991, 2024. doi: **10.1007/s12559-022-10072-w**.
4. N. Saqib, K.F. Haque, V.P. Yanambaka, and A. Abdelgawad, «Convolutional-Neural-Network-Based Handwritten Character Recognition: An Approach with Massive Multisource Data», *Algorithms*, vol. 15, article 129, 2022. doi: **10.3390/a15040129**.
5. J. Alzubi, N. Anand, and K. Akshi, «Machine learning from theory to algorithms: an overview», *Journal of physics: conference series*, vol. 1142, pp. 1-15, 2018. doi: **10.1088/1742-6596/1142/1/012012**.
6. M.D. Zeiler, and R. Fergus, «Visualizing and understanding convolutional networks», in Proc. 13th European Conf. «Computer Vision – ECCV 2014», Zurich, Switzerland, 2014, pp. 818-833. doi: **10.1007/978-3-319-10590-1_53**.
7. F. Schilling, «The effect of batch normalization on deep convolutional neural networks», M.S. thesis, Stockholm, Sweden, 2016.
8. D. Laptev, and J.M. Buhmann, «Transformation-invariant convolutional jungles», in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015, pp. 3043-3051. doi: **10.1109/CVPR.2015.7298923**.
9. A.H. Ribeiro, K. Tiels, L.A. Aguirre, and T. Schön, «Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness», in Proc. 23d Int. Conf. on Artificial Intelligence and Statistics, Palermo, Italy, 2020, vol. 108, pp. 2370-2380. doi: **10.48550/arXiv.1906.08482**.
10. S. Hochreiter, and J. Schmidhuber, «Long Short-Term Memory», *Neural Computation*, vol. 9(8), pp. 1735-1780, 1997. doi: **10.1162/neco.1997.9.8.1735**.
11. X. Huang, L. Qiao, W. Yu, J. Li, and Y. Ma, «End-to-end sequence labeling via convolutional recurrent neural network with a connectionist temporal classification layer», *International Journal of*

- Computational Intelligence Systems*, vol. 13, iss. 1, pp. 341-351, 2020. doi: **10.2991/ijcis.d.200316.001**.
12. Y. Li, D. Chen, T. Tang, and X. Shen, «HTR-VT: Handwritten text recognition with vision transformer», *Pattern Recognition*, vol. 158, article 110967, 2024. doi: **10.1016/j.patcog.2024.110967**.
 13. K. Barrere, Y. Soullard, A. Lemaitre, and B. Coüasnon, «A light transformer-based architecture for handwritten text recognition», *International Workshop on Document Analysis Systems*, pp. 275-290, 2022. doi: **10.1007/978-3-031-06555-2_19**.
 14. A. Vaswani et al., «Attention is all you need», in Proc. 31st Conf. on Neural Information Processing Systems NIPS 2017, Long Beach, CA, USA, 2017, pp. 1-11. doi: **10.48550/arXiv.1706.03762**.
 15. L. Ze et al., «Swin transformer: Hierarchical vision transformer using shifted windows», in Proc. IEEE/CVF International conference on computer vision, Las Vegas, Nevada, USA, 2021, pp. 9992-10002. doi: **10.1109/ICCV48922.2021.00986**.
 16. I.O. Tolstikhin et al., «Mlp-mixer: An all-mlp architecture for vision», in Proc. 35th Int. Conf. on Neural Information Processing Systems NIPS'21, 2021, pp. 24261-24272. doi: **10.48550/arXiv.2105.01601**.
 17. L. Ziming et al., «Kan: Kolmogorov-Arnold networks», 2024, pp. 1-50. doi: **10.48550/arXiv.2404.19756**.
 18. X. Yang, and X. Wang, «Kolmogorov-Arnold Transformer», 2024, pp. 1-19. doi: **10.48550/arXiv.2409.10594**.
 19. A. Molina, P. Schramowski, and K. Kersting, «Padé Activation Units: End-to-end Learning of Flexible Activation Functions in Deep Networks», in Proc. Int. Conf. on Learning Representations, 2020, pp. 1-17. doi: **10.48550/arXiv.1907.06732**.
 20. C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, «Online and offline handwritten Chinese character recognition: benchmarking on new databases», *Pattern Recognition*, vol. 46, iss. 1, pp. 155-162, 2013. doi: **10.1016/j.patcog.2012.06.021**.
 21. K. Dutta, P. Krishnan, M. Mathew, and C.V. Jawahar, «Offline Handwriting Recognition on Devanagari using a new Benchmark Dataset», in Proc. 13th IAPR International Workshop on Document Analysis Systems (DAS), Vancouver, Canada, 2018, pp. 25-30. doi: **10.1109/DAS.2018.69**.
 22. U-V. Marti, and H. Bunke, «The IAM-database: an English sentence database for offline handwriting recognition», *International journal on document analysis and recognition*, vol. 5, pp. 39-46, 2002. doi: **10.1007/s100320200071**.
 23. M. Fujitake, «Dtrocr: Decoder-only transformer for optical character recognition», in Proc. IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2024, pp. 8010-8020. doi: **10.1109/WACV57701.2024.00784**.

Стаття надійшла 13.08.2024

Стаття прийнята 10.09.2024

УДК 004.8

doi: 10.31498/2225-6733.49.1.2024.321205

© Волошин В.С.*

РОЛЬ НОВОГО ПОКОЛІННЯ СИСТЕМ ЗІ ШТУЧНИМ ІНТЕЛЕКТОМ У РОЗВИТКУ ЛЮДИНИ (НА ПРИКЛАДІ МЕРЕЖІ CHATGPT)

В роботі розглядається «pro-contra» у відносинах між людиною та машиною, маючи на увазі під останню системи штучного інтелекту (AI). Вперше увага приділяється такому аспекту питання, як суб'єктність і об'єктність частин системи «людина-машина». Такі критерії системи, як відношення до термодинамічної нерівноважності, вектори змінення ентропії, що відносяться як до людини, так і до систем

* д-р техн. наук, професор, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: 0000-0002-8423-2663, yvslshn52@gmail.com