

О. О. Зарицький¹
В. Я. Данилов¹

ПОРІВНЯННЯ ЕФЕКТИВНОСТІ МЕТОДІВ НАПІВКЕРОВАНОГО НАВЧАННЯ НА ОСНОВІ АВТОЕНКОДЕРІВ ДЛЯ ЗАДАЧ КЛАСИФІКАЦІЇ ФОТОГРАФІЙ РАКОВИХ ЗАХВОРЮВАНЬ ШКІРИ

¹Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Розглянуто та порівняно ефективність методів напівкерованого навчання на базі автоенкодерів для задачі класифікації фотографій ракових захворювань шкіри. У статті запропоновано систему оцінки методів напівкерованого навчання і запропоновано модифікації архітектур, описаних в літературних джерелах. Актуальність роботи полягає у дослідженні ефективності застосування різних методів напівкерованого навчання на прикладних задачах класифікації. Ефективне використання методів напівкерованого навчання для роботи з медичними даними може потенційно суттєво підвищити якість використання штучного інтелекту у медицині, оскільки нерозмічені медичні дані легше отримати і вони не потребують залучення дорогих спеціалістів. Як показали експерименти, проведені в рамках цього дослідження, методи напівкерованого навчання на базі автоенкодерів суттєво покращують метричні показники моделі для задачі класифікації фотографій ракових захворювань шкіри. В рамках дослідження розглянуто звичайний згортковий автоенкодер, варіаційний автоенкодер і сходову нейронну мережу, адаптовані для задач напівкерованого навчання. Як дані для експериментів використано датасет HAM10000. Для кожної моделі проведено 3 експерименти з різною кількістю розмічених даних у навчальній вибірці (5, 10, 30 % маркованих даних). Дослідження показали, що варіаційний автоенкодер демонструє виняткові результати на великій кількості розмічених даних, проте погано адаптується до малої кількості. Згортковий автоенкодер є ефективним тільки для великої кількості розмічених даних і дуже поступається складнішим методам НН у всіх експериментах. Дослідження показало, що сходова нейронна мережа є дуже ефективною для малої кількості маркованих даних порівняно з іншими розглянутими архітектурами. Отже, враховуючи специфіку задач класифікації медичних даних, можна зробити висновок, що сходова нейронна мережа є найефективнішою для такого типу задач.

Ключові слова: ракові захворювання шкіри, напівкероване навчання, автоенкодер, варіаційний автоенкодер, сходова нейронна мережа, HAM10000.

Вступ

Використання глибоких нейронних мереж у сфері аналізу медичних даних стає дуже розповсюдженим методом у медицині. Використання автоматизованих систем для аналізу дає можливість зменшити вплив людського фактора на рішення, а також дає можливість проводити певні діагностичні заходи без необхідності відвідування пацієнтом лікарні.

Особливо важливою автоматизація діагностики є у сфері виявлення онкологічних захворювань шкіри. Як відомо, онкологічні захворювання шкіри досить ефективно лікуються за умови ранньої діагностики. Отже, використання систем на основі глибоких нейронних мереж, що дають можливість діагностики без необхідності відвідувати лікарню, мають дуже великий потенціал збільшити ймовірність раннього ефективного лікування ракових захворювань шкіри.

Проте, для навчання глибоких нейронних мереж необхідна велика кількість розмічених даних, добування яких не є простою задачею у сфері медицини. Для розмітки медичних даних у більшості випадків необхідно наймати велику кількість спеціалістів, і це потребує великих витрат часу та грошей, водночас нерозмічені анонімізовані дані можна отримати від будь-якого медичного за-

кладу за умови укладення договору.

Саме тому для завдання аналізу медичних даних за допомогою нейронних мереж дуже важливими є дослідження у сфері напівкерованого навчання. Напівкероване навчання дає змогу використовувати для тренування моделей не тільки розмічені дані, а й нерозмічені.

Напівкероване навчання (далі може використовуватися скорочення НН) — це парадигма машинного навчання, що посідає позицію між керованим і некерованим навчанням. Воно дає змогу використовувати невелику кількість розмічених даних і велику кількість нерозмічених даних для навчання моделей.

У цій роботі розглянуто порівняння саме методів напівкерованого навчання на базі архітектури автоенкодерів. Автоенкодери широко використовуються для напівкерованого навчання класифікаційних моделей. Згідно з дослідженнями у роботах [1]—[4] методи на основі архітектури енкодер-декодер показують виняткові результати у задачах класифікації зображень. На відміну від класичних методів напівкерованого навчання типу Pseudo-Label або Co-train, методи на основі автоенкодерів не потребують окремого навчання моделі-вчителя, що спрощує процес тренування.

Проте у більшості досліджень автори демонструють ефективність методів на примітивних класифікаційних датасетах типу MNIST або CIFAR, для яких припущення напівкерованого навчання виконуються занадто очевидно. Важливо дослідити, наскільки ефективні ці методи для прикладних датасетів зі складнішою будовою. Ефективність таких методів напівкерованого навчання для медичних даних є недослідженою.

Для порівняння у роботі розглянуто методи на базі згорткового автоенкодера, варіаційного автоенкодера, а також сходової нейронної мережі (Ladder Network).

Метою роботи є емпіричне порівняння ефективності методів напівкерованого навчання на базі автоенкодерів за метричними показниками натренованих за допомогою них класифікаційних моделей.

Результати дослідження

Як вибірку вибрано датасет HAM10000. Цей датасет містить 10000 зображень ракових захворювань шкіри (або захворювань шкіри, що можуть з високою ймовірністю призвести до утворення раку шкіри), а також незлоякісних утворень (родимки) з відповідними до них класами. Великий

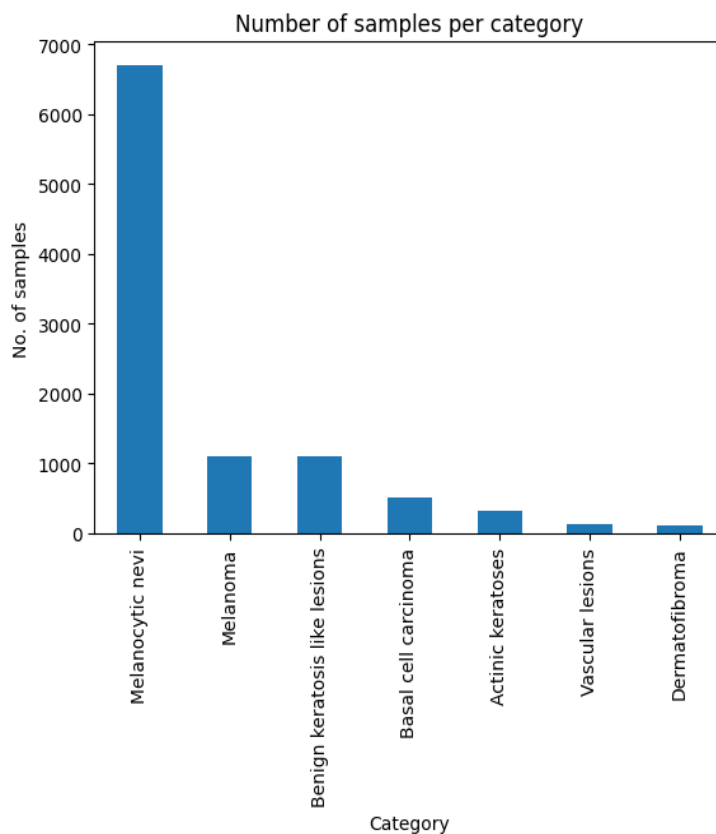


Рис. 1. Розподіл класів HAM10000 до ребалансування

розмір датасету дав авторам можливість віднести великий відсоток даних до нерозмічених і дослідити ефективність методів напівкерованого навчання для екстремально низького співвідношення розмічених до нерозмічених даних. Попередня обробка даних складалася з нормування зображень, зменшення розміру, а також з розподілення всіх класів зображень на два класи: доброякісні (клас 0) та злоякісні утворення (клас 1). Ці дії необхідні для того щоб навчальна і тестова вибірки були збалансованішими. Розподіл класів у датасеті HAM10000 до ребалансування класів зображено на рис. 1. Розподіл класів HAM10000 після ребалансування зображено на рис. 2.

Дані у датасеті розподілено на тестову та тренувальну вибірку у співвідношенні 1:9. Водночас тренувальну вибірку розділено на розмічену і нерозмічену частину. У цьому дослідженні розглянуто варіанти з 30%, 10% та 5% розмічених даних у навчальній вибірці.

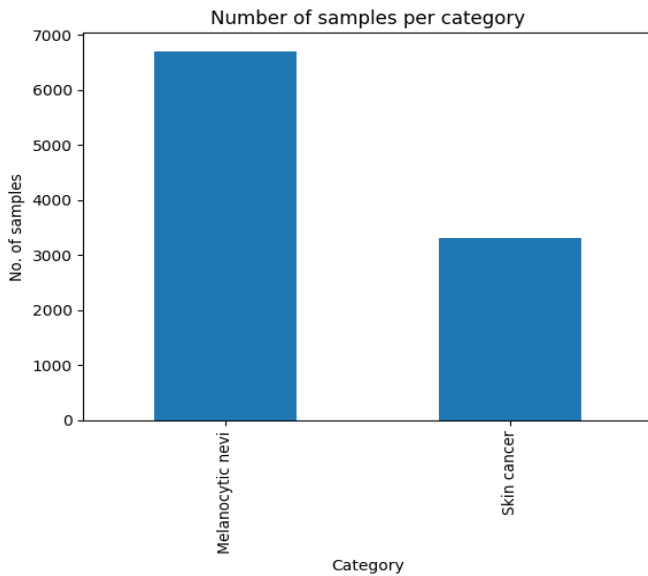


Рис. 2. Розподіл класів HAM10000 після ребалансування

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 32, 96, 96]	128
BatchNorm2d-2	[-1, 32, 96, 96]	64
Conv2d-3	[-1, 64, 96, 96]	18,496
BatchNorm2d-4	[-1, 64, 96, 96]	128
Conv2d-5	[-1, 128, 24, 24]	73,856
BatchNorm2d-6	[-1, 128, 24, 24]	256
Conv2d-7	[-1, 256, 6, 6]	295,168
BatchNorm2d-8	[-1, 256, 6, 6]	512
Linear-9	[-1, 256]	2,359,552
Linear-10	[-1, 64]	16,448
Linear-11	[-1, 2]	130
=====		
Total params: 2,764,738		
Trainable params: 2,764,738		
Non-trainable params: 0		
=====		

Рис. 3. Будова базової моделі

Для емпіричної оцінки ефективності методів напівкерowanego навчання взято просту згорткову нейронну мережу (далі – базова модель), що побудована з 4 шарів послідовно з'єднаних модулів згортки, батч-нормалізації та активаційної функції. Вихід шарів згортки передається в повнозв'язний шар з трьох елементів. Будову базової моделі показано на рис. 3.

На основі цієї нейронної мережі побудовано 3 різні архітектури автоенкодерів: звичайний згортковий автоенкодер, варіаційний автоенкодер, сходова нейронна мережа. Після навчання всіх 3 моделей, отримано метричні показники кожної моделі на тестовій вибірці для різних співвідношень розмічених і нерозмічених даних у навчальній вибірці. Як показники використовуватимуться стандартні класифікаційні показники, такі як accuracy, precision, recall і f1 score. Метричні показники моделі — це основні критерії оцінки ефективності методу напівкерowanego навчання. Зазвичай більшу увагу приділяли правильній класифікації зображень зі злякисними утвореннями, оскільки false negative сценарій для ракових захворювань має набагато серйозніші наслідки.

Також для порівняння методів НН вищезгадану модель чотиришарової згорткової нейронної мережі натреновано тільки на розмічених даних (еталонна «погана» модель), а також на повністю розміченому датасеті (еталонна модель ідеалу). Очевидно, що метричні показники ідеального методу НН мають наблизитись до показників моделі, навченої на повністю розмічених даних. Якщо метричні показники моделі подібні до еталонної «поганої» моделі, то це означає, що метод напівкерowanego навчання є неефективним, оскільки не дав приросту якості моделі за рахунок нерозмічених даних.

Згортковий автоенкодер

Для того, щоб мінімально модифікувати базову модель, за енкодер у схемі енкодер–декодер взято послідовні чотири шари, аналогічні базовій моделі. Для декодера створено послідовні чотири шари оберненої згортки з відступом (padding) для вихідних шарів. Аналогічно до енкодера, після кожної оберненої згортки застосовано батч-нормалізацію, а також активаційну функцію RELU або sigmoid.

На виході енкодера розташований повнозв'язний шар, аналогічний вказаному в базовій моделі. Вихід з повнозв'язного шару оцінюється функцією втрат Cross-entropy loss, що є стандартною функцією втрат для задач класифікації.

Також сигнал з виходу енкодера передається у декодер, який реконструює зображення. У подальшому оригінальне зображення та реконструйоване порівнюються за допомогою функції втрат MSE.

Схематичне зображення будови моделі згорткового автоенкодера для напівкерowanego навчання показано на рис. 4.

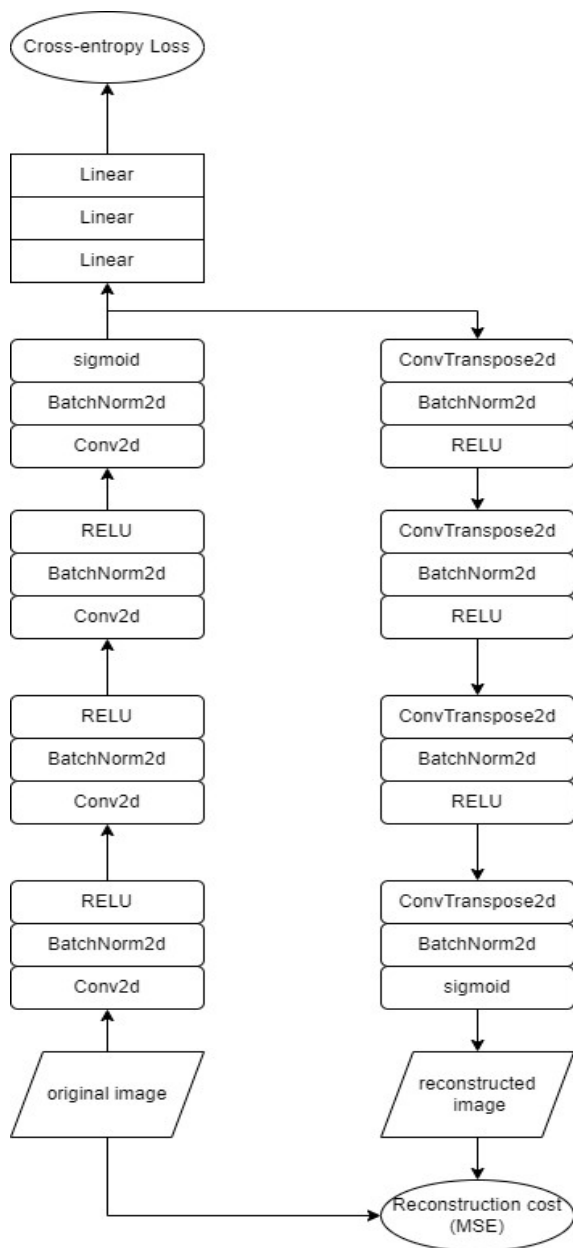


Рис. 4. Будова моделі згорткового автоенкодера

минулого шару декодера, а також зашумлена версія відповідного шару енкодера (у статті [1] автори називають цей зв'язок латеральним). Після цього кожний з виходів енкодера та відповідний йому вихід декодера порівнюються. З них формується reconstruction cost. Таким чином, завдання декодера фактично полягає у видаленні шуму і відновлення значень кожного шару енкодера. Варто зазначити, що реконструйоване зображення також порівнюється з оригінальним.

Для дослідження внесено декілька модифікацій відносно запропонованого в [1] та [2]. Лінійні шари замінені на згорткові, а також додано гіперпараметр α , що контролює вміст reconstruction cost у загальному значенні функції втрат аналогічно до згорткового автоенкодера (див. формулу (1)). Реконструйовані значення порівнювалися за допомогою функції втрат MSE. Всі інші архітектурні рішення залишилися незмінними. Структуру моделі сходової нейронної мережі показано на рис. 6.

Метричні показники вихідної моделі для 30 %, 10 % і 5 % розмічених даних у навчальній вибірці подано у формі таблиць. Для порівняння у таблицях також присутні вищезгадані еталонні варіанти моделей. У таблиці згортковий автоенкодер позначається як SA, варіаційний автоенкодер позначається як VA, а сходова нейронна мережа позначається як LN. Коефіцієнт α — гіперпараметр функції втрат у формулі (1). У таблицях також виділені найкращі показники у кожному стовпці.

Загальна функція втрат визначається за формулою

$$Loss = CE + \alpha RC, \quad (1)$$

де $Loss$ — загальна функція втрат; CE — значення Cross-entropy loss; RC — Reconstruction cost, що відповідає MSE між звичайним та реконструйованим зображенням; α — гіперпараметр, значення якого, зазвичай, лежить у діапазоні (0,1).

Варіаційний автоенкодер

Варіаційний автоенкодер — це архітектура штучної мережі, що належить до імовірнісних графових моделей. Основна відмінність цієї архітектури полягає у тому, що зображення не просто стискаються до певного латентного простору. Ознаки, отримані за допомогою шарів енкодера, використовуються для моделювання латентного шару за ймовірнісною моделлю. Вихідне значення у цій моделі представлене вектором середніх значень (μ) та вектором значень стандартних відхилень (σ).

За основу варіаційного автоенкодера для напівкерованого навчання використано модель M2 з дослідження [5]. Авторами змінено лише будову енкодера та декодера відповідно до базової моделі, проте, функції втрат та інші особливості архітектури запозичено з вищевказаної роботи. Загальна архітектура моделі варіаційного автоенкодера показана на рис. 5.

На відміну від моделі звичайного згорткового автоенкодера, функція втрат цієї моделі обчислюються як значення функції втрат ELBO (Evidence Lower Bound) з протилежним знаком. Функція втрат для розмічених даних як і раніше залишається cross-entropy loss.

Сходова нейронна мережа

Сходова нейронна мережа, також відома як ladder network, є модифікацією знешумлюючого автоенкодера для задач НН [1]. До кожного проміжного шару декодера подається об'єднання входу з

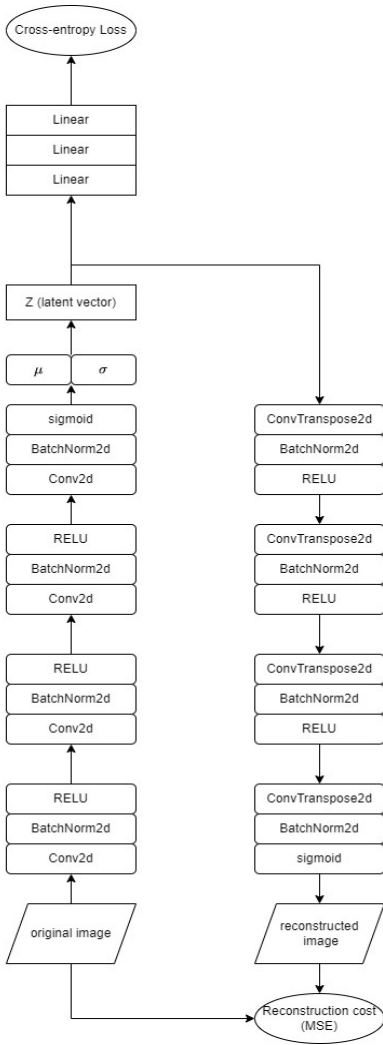


Рис. 5. Будова моделі на основі варіаційного автоенкодера

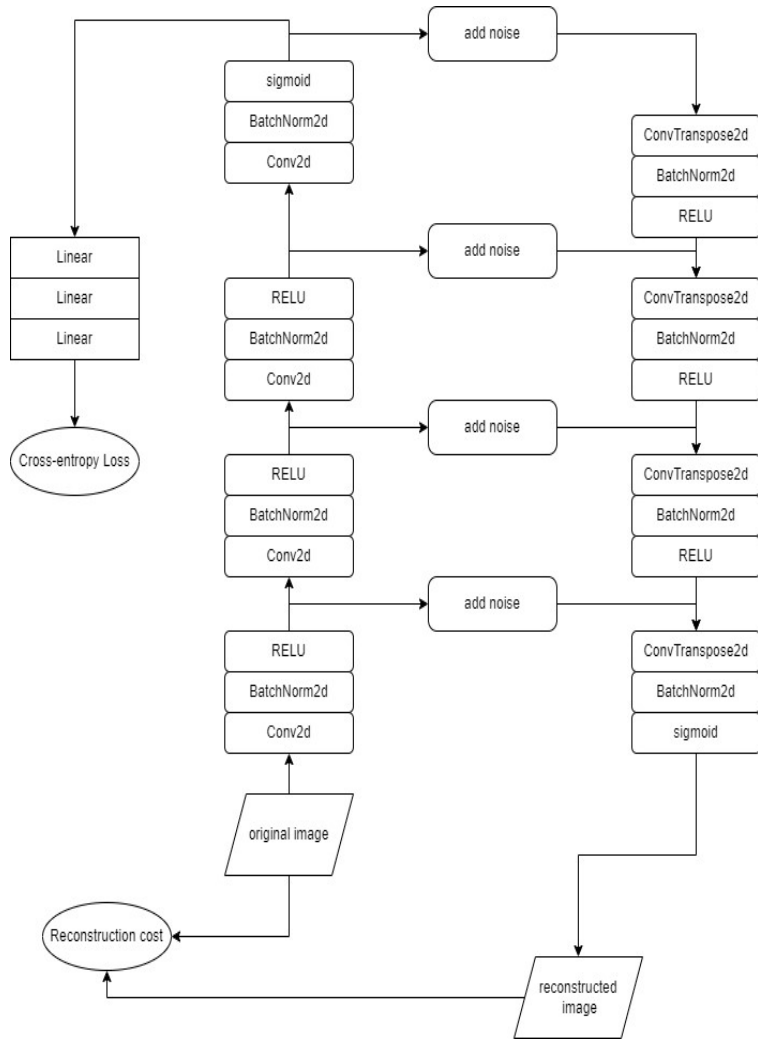


Рис. 6. Модель сходової нейронної мережі

Метричні показники для 30 % розмічених даних у навчальній вибірці подані у табл. 1.

Отримані результати

Таблиця 1

Метричні показники моделей для 30 % розмічених даних у навчальній вибірці

Назва методу напівкерованого навчання	Accuracy	Precision (melanocytic nevi)	Precision (skin cancer)	Recall (melanocytic nevi)	Recall (skin cancer)	F1 (melanocytic nevi)	F1 (skin cancer)
CA ($\alpha = 0,25$)	0,801	0,828	0,739	0,876	0,660	0,851	0,698
CA ($\alpha = 0,5$)	0,815	0,817	0,809	0,922	0,615	0,867	0,698
CA ($\alpha = 1$)	0,799	0,855	0,702	0,833	0,737	0,844	0,719
VA	0,822	0,910	0,656	0,832	0,796	0,869	0,720
LN ($\alpha = 0,25$)	0,807	0,824	0,767	0,896	0,641	0,858	0,699
LN ($\alpha = 0,5$)	0,789	0,814	0,729	0,876	0,626	0,844	0,674
LN ($\alpha = 1$)	0,799	0,791	0,828	0,941	0,534	0,859	0,650
Etalon bad model	0,773	0,779	0,751	0,908	0,519	0,839	0,614
Etalon good model	0,847	0,852	0,796	0,903	0,712	0,883	0,756

Як видно з табл. 1, для 30 % розмічених даних, найкращі метричні показники продемонстрував метод НН на основі варіаційного автоенкодера. Проте, також видно, що для великої кількості розмічених даних підходить і звичайний згортковий автоенкодер F1, score якого для класу skin cancer є майже таким самим (для $\alpha = 1$). Якщо значення Ladder Network $\alpha = 0,25$, то показано хороши

результати. Всі методи напівкерованого навчання суттєво покращили якість моделі порівняно з натренованою тільки на розмічених даних.

Метричні показники для 10 % розмічених даних у навчальній вибірці подані в табл. 2.

Таблиця 2

Метричні показники моделей для 10% розмічених даних у навчальній вибірці

Назва методу напівкерованого навчання	Accuracy	Precision (melanocytic nevi)	Precision (skin cancer)	Recall (melanocytic nevi)	Recall (skin cancer)	F1 (melanocytic nevi)	F1 (skin cancer)
CA ($\alpha = 0,25$)	0,797	0,853	0,698	0,831	0,733	0,842	0,715
CA ($\alpha = 0,5$)	0,787	0,830	0,702	0,847	0,676	0,838	0,689
CA ($\alpha = 1$)	0,797	0,845	0,707	0,843	0,710	0,844	0,709
VA	0,775	0,822	0,687	0,831	0,674	0,827	0,681
LN ($\alpha = 0,25$)	0,794	0,844	0,702	0,839	0,710	0,841	0,706
LN ($\alpha = 0,5$)	0,814	0,868	0,721	0,843	0,760	0,855	0,740
LN ($\alpha = 1$)	0,787	0,833	0,699	0,843	0,683	0,838	0,691
Etalon bad model	0,743	0,786	0,654	0,844	0,552	0,819	0,591
Etalon good model	0,847	0,852	0,796	0,903	0,712	0,883	0,756

Для 10 % розмічених даних видно, що найкращі показники має модель Ladder Network з коефіцієнтом $\alpha = 0,5$. Водночас варіаційний автоенкодер показав суттєво гірші результати, порівняно з іншими моделями, не зважаючи на те, що у разі 30 % розмічених даних, він показував безумовно найкращі результати. Це демонструє погану адаптованість цього методу до високого співвідношення розмічених і нерозмічених даних.

У табл. 3 подано метричні показники моделей, навчених на 5 % даних.

Таблиця 3

Метричні показники моделей для 5% розмічених даних у навчальній вибірці

Назва методу напівкерованого навчання	Accuracy	Precision (melanocytic nevi)	Precision (skin cancer)	Recall (melanocytic nevi)	Recall (skin cancer)	F1 (melanocytic nevi)	F1 (skin cancer)
CA ($\alpha = 0,25$)	0,778	0,797	0,727	0,884	0,580	0,838	0,645
CA ($\alpha = 0,5$)	0,777	0,819	0,690	0,843	0,653	0,831	0,671
CA ($\alpha = 1$)	0,778	0,808	0,709	0,865	0,615	0,835	0,658
VA	0,799	0,902	0,607	0,811	0,768	0,854	0,678
LN ($\alpha = 0,25$)	0,771	0,846	0,654	0,794	0,729	0,819	0,690
LN ($\alpha = 0,5$)	0,802	0,803	0,799	0,922	0,576	0,858	0,670
LN ($\alpha = 1$)	0,783	0,806	0,728	0,880	0,603	0,841	0,660
Etalon bad model	0,691	0,913	0,534	0,582	0,897	0,711	0,670
Etalon good model	0,847	0,852	0,796	0,903	0,712	0,883	0,756

З результатів, поданих у табл. 3, видно, що найкращі результати знову показали моделі Ladder Network. Варіаційний автоенкодер показав прийнятні результати. Проте, навіть для такої низької кількості розмічених даних всі методи напівкерованого навчання суттєво покращують якість моделей порівняно з навченою моделлю тільки на розмічених даних.

Висновки

Як видно з проведеного дослідження, усі методи напівкерованого навчання на базі автоенкодерів дають суттєвий приріст більшості метричних показників моделі у порівнянні з навчанням тільки на розмічених даних. Отже, автоенкодер можна ефективно застосовувати для напівкерованого навчання у задачах класифікації зображень ракових захворювань шкіри.

Також експерименти показали, що варіаційний автоенкодер дуже ефективний за відносно великої кількості розмічених даних (на 30 % розмічених даних приріст F1 score становив 3 % для класу 0 та 10,6 % — для класу 1), проте погано адаптується до великої кількості нерозмічених даних. Згортковий автоенкодер має простішу будову і показує хороші результати у випадку 30 % розмічених даних. Так, згортковий автоенкодер з коефіцієнтом $\alpha = 1$ на 30 % розмічених даних показав приріст F1 score на 0,5 % для класу 1 та 10,5 % для класу 2. Відповідно згортковий автоенкодер можна використовувати у ситуаціях, коли співвідношення маркованих до немаркованих даних невелике. Сходова нейронна мережа показала хороші результати на всіх трьох експериментах.

Також Ladder Network показала виняткові результати за низької кількості розмічених даних (на 5 % розмічених даних приріст показника F1 score становив максимально 14,7 % для класу 0 та 2 % — для класу 1), що потенційно свідчить про те, що цю архітектуру найдоцільніше використовувати для навчання моделей розглянутої нами задачі. Експерименти показали, що найефективнішим методом НН для малої кількості маркованих даних є Ladder Network з коефіцієнтом $\alpha = 0,5$ (приріст точності 11,1 %, приріст F1 score 14,7 % для класу 0 та 0 % — для класу 1). Найбільший приріст F1 score становив 2 % для класу 1 для 5 % маркованих даних, його показала модель LN, коли $\alpha = 0,25$.

СПИСОК ВИКОРИСТАНОЇ ЛІТУРЕТУРИ

- [1] M. Pazeshki, L. Fan, F. Brakel, A. Courville, and Y. Bengio, “Deconstructing the Ladder Network Architecture,” in *Proceedings of the 33rd International Conference on Machine Learning*, New York, NY, USA, 2016.
- [2] A. Rasmus, R. Valpola, and M. Honkela “Semi-Supervised Learning with Ladder Networks,” in *Advances in Neural Information Processing Systems*, 28, 2015,
- [3] Gogna Anupriya, and Angshul Majumdar, “Semi supervised autoencoder,” *Neural Information Processing*, 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16-21, 2016, Proceedings, Part II 23, Springer International Publishing, 2016.
- [4] Kingma Durk P, et al, “Semi-supervised learning with deep generative models,” *Advances in neural information processing systems*, 27, 2014.
- [5] Xu. Weidi, et al, “Variational autoencoder for semi-supervised text classification,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

Рекомендована кафедрою системного аналізу та інформаційних технологій ВНТУ

Стаття надійшла до редакції 14.03.2024

Зарицький Олексій Олексійович — студент Інституту прикладного системного аналізу, e-mail: zaritskiy.alexey@gmail.com ;

Данилов Валерій Якович — д-р техн. наук, професор, професор Навчально-наукового інституту прикладного системного аналізу, email: danilov1950@ukr.net .

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ

O. O. Zarytskyi¹
V. Ya. Danylov¹

Comparison of the Effectiveness of Semi-Supervised Learning Methods Based on Autoencoders for Skin Cancer Photo Classification Tasks

¹National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”

In this paper, we reviewed and compared the effectiveness of semi-supervised learning methods based on autoencoders for the task of classifying photos of skin cancers. The paper proposes a system for evaluating semi-supervised learning methods and suggests modifications to the architectures described in the literature. The relevance of this work is to study the effectiveness of using various semi-supervised learning methods on applied classification tasks. The effective use of semi-supervised learning methods for working with medical data can potentially significantly increase the quality of artificial intelligence in medicine, since unlabeled medical data is easier to obtain and does not require the involvement of expensive specialists. As the experiments conducted in this study have shown, semi-supervised learning methods based on autoencoders significantly improve the performance metrics of the model for the task of classifying photos of skin cancers. In our study, we considered a conventional convolutional autoencoder, a variational autoencoder, and a ladder neural network, all adapted for semi-supervised learning tasks. The HAM10000 dataset was used as the data for the experiments. For each model, 3 experiments were conducted with different amounts of labeled data in the training set (5, 10, 30 % of labeled data). The studies showed that the variational autoencoder demonstrates exceptional results on a large amount of labeled data, but does not adapt well to a small amount. The convolutional autoencoder is effective only for a large amount of labeled data and is significantly inferior to more complex architectures in all experiments. The study has shown that the ladder neural network is very effective for a small amount of labeled data compared to the other architectures considered. Thus, taking into account the specifics of medical data classification tasks, we can conclude that the ladder neural network is the most effective for this type of task.

Keywords: skin cancer, semi-supervised learning, autoencoders, variational autoencoder, Ladder Network, HAM10000.

Zarytskyi Oleksii O. — Student of the Institute of Applied Systems Analysis, e-mail: zaritskiy.alexey@gmail.com ;

Danylov Valeriy Ya. — Dr. Sc. (Eng.), Professor, Professor of the Institute of Educational and Research of Applied System Analysis, e-mail: danilov1950@ukr.net