

Міністерство освіти і науки України
Львівський національний університет імені Івана Франка

М. В. БАРАНОВ

Ю. М. ЩЕРБИНА

НЕЙРОННІ МЕРЕЖІ: фундаментальні принципи та застосування до малих наборів даних

Монографія

Львів
ЛНУ ім. Івана Франка
2026

УДК 004.032.26:303.732.4(06)

Б24

Рецензенти:

д-р техн. наук, проф., чл.-кореспондент НАН України *Н. Д. Панкратова*
(Київський політехнічний інститут імені Ігоря Сікорського);
д-р фіз.-мат. наук, проф. *М. С. Нікітченко*
(Київський національний університет імені Тараса Шевченка);
д-р техн. наук, проф. *М. В. Ткачук*
(Харківський національний університет імені В. Н. Каразіна)

*Рекомендовано до друку Вченою радою
Львівського національного університету імені Івана Франка.
Протокол № 92/12 від 3 грудня 2025 року*

Баранов М. В.

Б24

Нейронні мережі: фундаментальні принципи та застосування до малих наборів даних : монографія / М. Б. Баранов, Ю. М. Щербина. – Львів : ЛНУ ім. Івана Франка, 2026. – 248 с.

ISBN 978-617-10-0978-3.

Викладено фундаментальні принципи роботи нейронних мереж, розглянуто докладно роботу нейронів, багат шарового перцептронну, згорткових нейронних мереж, рекурентних нейронних мереж та архітектур трансформерів. Представлено результати дослідження застосування нейронних мереж до задачі тренування на малих наборах даних.

Для студентів спеціальностей «Інформатика», «Прикладна математика», «Системний аналіз».

УДК 004.032.26:303.732.4(06)

*Відтворення цієї книги або будь-якої її частини
заборонено без письмової згоди видавництва.
Будь-які спроби порушення авторських прав
будуть переслідуватися у судовому порядку*

© Баранов М. В., Щербина Ю. М., 2026

© Львівський національний університет
імені Івана Франка, 2026

ISBN 978-617-10-0978-3

ЗМІСТ

ВСТУП	9
Розділ 1. ВСТУП ДО МАШИННОГО НАВЧАННЯ	11
1.1 Мотивація та концепція машинного навчання	11
1.1.1 Ідея та завдання машинного навчання	11
1.1.2 Парадигми машинного навчання.....	12
1.1.3 Лінійна регресія.....	14
1.1.4 Логістична регресія.....	22
1.2 Нейронні мережі.....	27
1.2.1 Завдання нейронних мереж.....	27
1.2.2 Поняття штучного нейрону.....	27
1.2.3 Шар лінійної моделі.....	28
1.2.4 Багатошаровий перцептрон.....	30
1.2.5 Тренування багатошарового перцептрону.....	33
1.2.6 Лінійна регресія нейронною мережею.....	35
1.2.7 Нелінійна регресія нейронною мережею.....	35
1.2.8 Логістична регресія.....	37
1.2.9 Задача класифікації.....	38
1.2.10 Багатокласова задача класифікації.....	39
1.3 Робота з даними для тренування нейронних мереж	42
1.3.1 Перенавчання нейронних мереж: поняття та виявлення.....	42
1.3.2 Розбиття набору даних на тренувальну та затверджувальну вибірку.....	44
1.3.3 Перехресне затвердження.....	46
1.3.4 Розбиття часових послідовностей	46
1.3.5 Нормалізація даних.....	48
1.3.6 Типові методи нормалізації даних.....	50

1.4 Оптимізація цільової функції втрат.....	51
1.4.1 Стохастичний метод градієнтного спуску.....	51
1.4.2 Оптимізація методом градієнтного спуску з моментом.....	53
1.4.3 Оптимізація методом Adagrad.....	55
1.4.4 Оптимізація методом RMSProp.....	55
1.4.5 Оптимізація методом Adam.....	56
1.5 Згорткові нейронні мережі.....	57
1.5.1 Основи комп'ютерного бачення.....	57
1.5.2 Проблеми роботи з зображеннями багатоплановим перцептроном.....	59
1.5.3 Основи згорткових операцій.....	62
1.5.4 Згортковий шар нейронної мережі.....	66
1.5.5 Схема роботи згорткових нейронних мереж.....	69
1.5.6 Багатовимірний згортковий шар нейронної мережі.....	73
1.5.7 Практичні аспекти реалізації та використання згорткових шарів.....	78
1.5.8 Застосування згорткових мереж.....	80
1.6 Пошук об'єктів на зображеннях з використанням згорткових нейронних мереж.....	81
1.6.1 Задача пошуку об'єктів.....	81
1.6.2 Локалізація об'єктів.....	82
1.6.3 Локалізація та класифікація об'єктів.....	83
1.6.4 Нульова гіпотеза при пошуку об'єктів.....	84
1.6.5 Пошук багатьох об'єктів.....	86
1.6.6 Метод пригнічення немаксимумів.....	86
1.7 Автокодувальники.....	88
1.7.1 Ідея архітектури автокодувальників.....	88
1.7.2 Автокодувальник з усуненням шумів.....	90
1.7.3 Згортковий автокодувальник.....	90
1.7.4 Варіаційний автокодувальник.....	94
1.8 Рекурентні нейронні мережі.....	95
1.8.1 Обмеження прямих нейронних мереж.....	95

1.8.2 Базова структура рекурентних нейронних мереж та їх класифікація.....	97
1.8.3 RNN: Класична рекурентна нейронна мережа.....	100
1.8.4 LSTM: Модель з довготривалою та короткочасною пам'яттю.....	102
1.8.5 GRU: Модель з вентиляльованим рекурентним вузлом	103
1.9 Обробка природної мови	104
1.9.1 Проблема векторизації слів.....	104
1.9.2 Класичні методи векторизації слів.....	105
1.9.3 Векторизація Word2Vec.....	106
1.9.4 Архітектура моделі класифікації тексту	109
1.9.5 Seq2Seq: Нейронний машинний переклад.....	109
1.10 Трансформери.....	110
1.10.1 Проблема обробки довгих послідовностей.....	110
1.10.2 Базовий механізм уваги	111
1.10.3 Узагальнений механізм уваги	115
1.10.4 Архітектура трансформерів.....	117
1.10.5 Позиційне кодування	118
1.10.6 Механізм багатонаправленої уваги	119
1.10.7 Декодувальник трансформера.....	120
Розділ 2. ОГЛЯД ЛІТЕРАТУРИ	122
2.1 Підходи до навчання нейронних мереж на малих наборах даних.....	122
2.1.1 Переносне навчання.....	122
2.1.2 Дистиляція знань моделі	123
2.1.3 Сіамські нейронні мережі.....	124
2.1.4 Прототипні мережі.....	125
2.1.5 Мережі з відповідністю	126
2.1.6 Реляційні нейронні мережі	126
2.1.7 Модель-незалежне мета-навчання.....	127
2.2 Стохастичні та ймовірнісні нейронні мережі	128
2.2.1 Басівські нейронні мережі	128

2.2.2	Ансамблі детерміністичних моделей	129
2.2.3	Ймовірнісні нейронні мережі.....	130
2.3	Моделі на основі множини експертів.....	131
2.4	Висновки.....	131
Розділ 3.	МЕТОДОЛОГІЯ	133
3.1	Ймовірнісні нейронні мережі.....	133
3.1.1	Ідея та мотивація ймовірнісного підходу у нейронних мережах	133
3.1.2	Перехід від класичних нейронних мереж до мереж ймовірнісного типу	134
3.1.3	Цільова функція втрат ймовірнісних моделей	136
3.1.4	Розв'язування регресійних задач за допомогою ймовірнісних нейронних мереж.....	137
3.1.5	Розв'язування класифікаційних задач за допомогою ймовірнісних нейронних мереж.....	137
3.1.6	Переваги використання ймовірнісного підходу у нейронних мережах	138
3.2	Навчання моделей ймовірнісним підходом на малих наборах даних	139
3.2.1	Ідея та мотивація.....	139
3.2.2	Побудова реляційної ймовірнісної нейронної мережі.....	140
3.2.3	Обґрунтування ймовірнісного підходу для реляційних нейронних мереж	142
3.2.4	Переваги ймовірнісного підходу у реляційних нейронних мережах	146
3.2.5	Алгоритм адаптивної побудови пар для оптимізації процесу тренування	147
3.2.6	Теоретична оцінка пришвидшення тренування з використанням адаптивного алгоритму побудови пар.....	149
3.2.7	Балансування позитивних та негативних пар у реляційних ймовірнісних моделях	150
3.2.8	Цільова функція втрат реляційної ймовірнісної мережі	151

3.2.9	Регулювання рівня невизначеності в моделі	153
3.3	Нейронні мережі з розподіленими знаннями.....	154
3.3.1	Ідея та мотивація нейронних мереж з розподіленими знаннями	154
3.3.2	Архітектура моделі	156
3.3.3	Переваги й обмеження нейронної мережі з розподіленими знаннями	158
3.4	Реляційні ймовірнісні нейронні мережі для задачі багатокласової класифікації	159
3.4.1	Задача багатокласової класифікації у реляційних нейронних мережах	159
3.4.2	Інтеграція концепції розподілених знань у реляційні моделі	160
3.4.3	Інтеграція ймовірнісного тренування у реляційні нейронні мережі з розподіленими знаннями.....	161
3.4.4	Переваги реляційних нейронних мереж з розподіленими знаннями	162
3.5	Висновки	164
Розділ 4. РЕЗУЛЬТАТИ	165
4.1	Огляд набору даних для апробації моделі	165
4.1.1	Відкритий набір даних Galaxy Zoo DECaLS	165
4.2	Реалізація та апробація реляційної ймовірнісної моделі.....	166
4.2.1	Вимоги до побудови реляційної ймовірнісної моделі	166
4.2.2	Набір даних для апробації реляційної ймовірнісної моделі.....	167
4.2.3	Архітектура реляційної ймовірнісної моделі	168
4.2.4	Тренування реляційної ймовірнісної моделі	172
4.2.5	Постановка експериментів та результати	176
4.2.6	Апробація класифікатора на базі реляційної ймовірнісної моделі.....	182
4.3	Нейронна мережа з розподіленими знаннями	189
4.3.1	Вимоги до побудови нейронної мережі з розподіленими знаннями	189

4.3.2	Реалізація нейронної мережі з розподіленими знаннями.....	190
4.4	Побудова та апробація реляційної ймовірнісної нейронної мережі для задачі багатокласової класифікації	193
4.4.1	Мотивація та задача реляційної ймовірнісної нейронної мережі	193
4.4.2	Архітектура реляційної ймовірнісної мережі для задачі багатокласової класифікації	194
4.4.3	Набір даних для апробації реляційної ймовірнісної нейронної мережі для задачі багатокласової класифікації	197
4.4.4	Особливості реалізації та векторизації архітектури. Процес тренування моделі	197
4.4.5	Вирішення проблеми дисбалансу даних багатокласової класифікації та підхід формування пар	199
4.4.6	Тренування реляційної ймовірнісної моделі з розподіленими знаннями	201
4.4.7	Постановка експериментів та результати	202
4.5	Висновки	208
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ		209
ДОДАТКИ		216
ДОДАТОК А. Розвідувальний аналіз даних Galaxy Zoo Decals		216
ДОДАТОК Б. Детальні результати апробації класифікатора на базі реляційної ймовірнісної моделі		226
ДОДАТОК В. Детальні результати апробації багатокласового класифікатора на базі реляційної ймовірнісної моделі		239

ВСТУП

Протягом останніх років нейронні мережі стають незамінним інструментом в розв'язуванні багатьох задач у різноманітних сферах, таких як розпізнавання людської мови, обробка сигналів, розпізнавання образів тощо. Завдяки властивості узагальнення тренувальних даних, моделі на базі нейронних мереж здатні розв'язувати задачі, рішення яких практично неможливо віднайти без використання алгоритмів машинного навчання.

Важливість використання нейронних мереж в сучасному світі важко переоцінити. За допомогою нейронних мереж вдалось автоматизувати значну кількість рутинної роботи, як от аналіз якості вироблених продуктів на підприємствах чи моніторинг дорожнього руху тощо. Інтеграція моделей машинного навчання також дала змогу значно здешевити ці процеси. Наприклад, використання звичайної оптичної камери є набагато дешевшим рішенням для підрахунку об'єктів на конвеєрі у порівнянні з фізичними пристроями зчитування. Окрім цього, інтелектуальний аналіз даних дає можливість підвищити рівень безпеки в сучасному суспільстві шляхом безперервного моніторингу камер спостереження та автоматичного реагування на надзвичайні ситуації. На додаток нейронні мережі є важливим інструментом у наукових дослідженнях, оскільки моделі дають змогу автоматично встановити неочевидні статистичні залежності між випадковими величинами, які людині було б вкрай проблематично виявити та зрозуміти.

Для тренування моделей, що базуються на нейронних мережах необхідна наявність великої кількості проанотованих даних для тренування. Збір масштабних наборів даних не завжди є можливим з огляду на складність отримання даних для деяких сфер (наприклад, рентгенівські знімки, які недоступні з огляду на захист персональних даних). Також практично неможливо створити масштабний набір даних для рідкісних ситуацій (наприклад, спалахи сонячної активності, знімки унікальних морфологій галактик тощо).

Наявність зібраних даних не є достатньою умовою успішної побудови моделі, оскільки алгоритми навчання з вчителем вимагають анотацій до прикладів. Цей процес концептуально не може бути автоматизованим повністю. Таким чином, анотація даних потребує вкладення великої кількості людської праці. Також для певних галузей анотація повинна проводитись виключно спеціалістами цієї галузі. Звідси безпосередньо випливає важливість технологій тренування нейронних мереж в умовах обмеженої кількості даних.

З іншого боку – навіть наявність великомасштабних наборів даних не гарантує побудови універсальної моделі. Отримані моделі на базі нейронних мереж досягають високих показників якості виключно тоді, коли вхідні дані належать тому ж розподілу, що й тренувальний набір даних. Якщо ж природа вхідних даних буде змінена – якість роботи такої моделі зазвичай змінюється в гіршу сторону.

Існує багато задач, де адаптація моделей до нової природи даних є обов'язковим елементом системи. На відміну від розпізнавання номерних знаків (де набір символів алфавіту є сталим і в загальному випадку незмінним), розпізнавання продуктів у супермаркетах потребує постійної неперервної адаптації моделей до нових товарів. Перетренування таких моделей вимагає значної кількості часу (від декількох днів до декількох місяців) при умові наявності проанотованого набору даних. Варто зазначити, що тренування моделей на базі нейронних мереж також вимагає потужних обчислювальних ресурсів.

Техніки навчання на малому наборі даних дають змогу побудувати такі моделі на основі нейронних мереж, які можна натренувати з використанням декількох тренувальних прикладів. Побудова таких моделей є складним процесом з огляду на те, що ризик перенавчання суттєво зростає при тренуванні нейронних мереж з використанням малої кількості тренувальних даних. З іншої сторони модель повинна бути здатною до розв'язування складних задач, що в свою чергу не дає можливості скоротити розмір такої моделі.

Для навчання моделей нейронних мереж на малих наборах даних нещодавно було розроблено множини таких підходів: переносне навчання [1] та дистиляція знань моделі [2], сіамські нейронні мережі [3], прототипні мережі [4], мережі з відповідністю [5], реляційні мережі [6], модель-незалежне мета-навчання [7] тощо.

Незважаючи на значний прогрес розвитку галузі в цьому напрямку, проблема ефективного навчання та уникнення перенавчання залишається актуальною та потребує подальшого дослідження. Існуючі методи умовно можна розділити на дві категорії:

- Методи, що фокусуються на початковій ініціалізації моделі.
- Методи навчання на основі метрик.

Хоча ці два напрямки розглядають незалежні підходи до тренування моделі – проблема залишається спільною: перенавчання на малій тренувальній вибірці.

Розробка нового підходу до подолання ефекту перенавчання моделей на таких вибірках дасть змогу розширити можливості використання нейронних мереж та відкриє можливість використання таких моделей для аналізу унікальних даних. Це є важливим аспектом у науці та має багато потенційних практичних застосувань, як от використання моделей у медицині.

Наукове видання

Микола Вікторович БАРАНОВ
Юрій Миколайович ЩЕРБИНА

НЕЙРОННІ МЕРЕЖІ: фундаментальні принципи та застосування до малих наборів даних

Монографія

Редактор *Наталія Ільків*
Комп'ютерне верстання *Вікторії Лоїк*
Обкладинка *Маргарити Білобринь*

Формат 70x100/16.
Умовн. друк. арк. 19,99. Тираж 50 прим.

Видавець та виготовлювач:
Львівський національний університет імені Івана Франка,
вул. Університетська, 1, м. Львів, 79007.
Свідоцтво про внесення суб'єкта видавничої справи
до Державного реєстру видавців, виготівників
і розповсюджувачів видавничої продукції.
Серія ДК № 3059 від 13.12.2007 р.