

Г. С. Іващенко, Д. О. Тимошенко, О. В. Близнюк, О. М. Кононенко

Харківський національний університет радіоелектроніки, Харків, Україна

МОДЕЛІ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Анотація. Актуальність. Прогнозування часових рядів є одним із важливих інструментів для різних сфер людської діяльності, оскільки воно дозволяє аналізувати минулі тенденції, розуміти динаміку подій та приймати обґрунтовані рішення на основі попередньо зібраних історичних даних. За останні роки моделі штучних нейронних мереж глибокого навчання показали значний потенціал у сфері прогнозування часових рядів. **Метою даної роботи** є аналіз використання моделей глибокого навчання для короткострокового прогнозування часових рядів різного походження та з можливою наявністю викривлень. **Об'єктом дослідження** є процес прогнозування часових рядів. **Предметом дослідження** є використання моделей глибокого навчання на основі CNN, RNN, TCNN та LSTM архітектур для прогнозування часових рядів. **Результати.** Експериментальні дослідження показали, що прогнози нестационарних часових рядів за допомогою штучної нейронної мережі на основі архітектури LSTM виявились найближчими до реальних даних, порівняно з іншими моделями глибокого навчання. **Висновок.** Отримані результати у більшості випадків підтверджують перевагу використання моделей на основі LSTM перед іншими розглянутими моделями глибокого навчання для прогнозування часових рядів.

Ключові слова: прогнозування часових рядів, машинне навчання, штучна нейронна мережа, моделі глибокого навчання, згорткові мережі, довга короткострокова пам'ять.

Вступ

Наразі дані відіграють все більш важливу роль у прийнятті рішень та процесі планування. Поширення аналізу даних, представлених у вигляді часових рядів, відбувається у зв'язку з розвитком сучасних технологій, зокрема, засобів машинного навчання.

Часовий ряд являє собою послідовність спостережень за певним показником у різні моменти часу з однаковим (як правило) інтервалом [1]. Як приклад часових рядів можна привести дані про обсяги продажу продукції за вказаний період або дані про середню температуру повітря на певній території в різні роки. При вирішенні завдання прогнозування часових рядів, в якості вхідних даних можна використовувати не тільки цільовий показник (його минулі значення), а й зовнішні параметри, які можуть прямо або опосередковано впливати на результати прогнозу [1].

Існує багато методів прогнозування часових рядів, але найбільш актуальними є методи на основі підходів машинного навчання [2], таких як генетичні алгоритми, метод висновку за прецедентами, штучні імунні системи та штучні нейронні мережі (ШНМ).

Найбільш поширеним інструментом для прогнозування часових рядів з урахуванням зовнішніх факторів є саме штучні нейронні мережі, через їх можливість враховувати численні параметри для отримання точних прогнозів, а процес налаштування та навчання може бути автоматизовано за допомогою інших засобів обчислювального інтелекту, що забезпечує можливість адаптації моделей прогнозування до нових даних [3]. Перспективним є використання моделей ШНМ глибокого навчання, які дозволяють навчатися для створення прогнозу на викривлених даних, характеризуються можливостями донавання та урахування зовнішніх факторів, що впливають на величину, яка прогнозується [4]. Існують різні архітектури глибокого навчання, такі як рекурентні нейронні мережі, згорткові нейронні мережі, нейронні мережі з довгостроковою та короткостроковою пам'яттю. Завдяки своїй здатності вивчати складні нелі-

нійні залежності та адаптуватися до різноманітних вхідних даних, ці моделі дозволяють досягти значних результатів у прогнозуванні часових рядів [4].

Однією із проблем вирішення завдання прогнозування часових рядів за допомогою ШНМ є вибір архітектури, з метою чого доцільно розглянути актуальні наукові роботи, у яких широко представлені моделі і методи прогнозування часових рядів на основі штучних нейронних мереж.

У [5] проведено порівняльний аналіз прогнозування часових рядів на основі моделі BiLSTM та статистичними методами, для вивчення розвитку та поширення пандемії COVID-19 в Саудівській Аравії. Відповідно до результатів дослідження, запропонована модель BiLSTM показала значно вищу ефективність відносно використання статистичних методів прогнозування часових рядів. Проте, порівняно з іншими поширеними моделями ШНМ, такими як LSTM (Long Short-Term Memory) і GRU (Gated Recurrent Unit), при прогнозуванні часових рядів на 14 та 60 днів модель BiLSTM показала лише аналогічну точність.

Розглянуте дослідження [6] спрямоване на порівняння ефективності прогнозування за допомогою ШНМ на архітектурі LSTM, MLP (Multilayer Perceptron) та CNN (Convolutional Neural Network). Використаний набір даних щодо спалаху COVID-19 у Єгипті містив записи про підтверджені випадки зараження, смерті та одужання у проміжку від 14 лютого 2020 року до 30 червня 2021 року. 90% даних було використано для навчальної вибірки та 10% для тестової. Виходячи з отриманих результатів, зроблено висновок, що модель на основі архітектури LSTM має найбільшу точність при прогнозуванні часових рядів на тиждень і на місяць уперед. Найгірші результати показала модель CNN, яка під час прогнозування часових рядів з горизонтом, що дорівнює один місяць, почала періодично генерувати аномальні значення.

Виходячи з аналізу сучасних наукових робіт можна зробити висновок, що завдання прогнозування за

допомогою моделей глибокого навчання є актуальним, і потребує детального дослідження.

Метою цієї роботи є аналіз ефективності використання моделей ШНМ глибокого навчання для короткострокового прогнозування часових рядів.

Постановка завдання

Часові ряди являють собою послідовний набір значень, вимірюваних протягом N періодів часу,

$$x_t = (x_{t-N}, x_{t-N+1}, x_{t-1}, \dots, x_{t-2}, x_{t-1}),$$

де t – мітка поточного часу, $0 \leq t \leq N$ [7].

Вимірювання, зроблені під час процесу, що описується за допомогою часового ряду, розташовані в належному відсортовано-хронологічному порядку.

В дослідженні використання моделей глибокого навчання для прогнозування часових рядів у якості вихідних даних для аналізу обрано дані щодо числа захворілих під час пандемії COVID-19 в Україні (зібрані Всесвітньою Організацією Охорони Здоров'я), дані середньорічної температури повітря за період 2009-2023 років та історичні дані курсів акцій NASDAQ.

Таким чином, для експериментальних досліджень були обрані ряди різного походження, розміру, з відсутніми та наявними зовнішніми факторами та викривленнями.

Дослідження передбачає використання таких моделей ШНМ, як згорткова нейронна мережа (CNN), рекурентна нейронна мережа (RNN), часова згорткова нейронна мережа (TCNN) та довга короткострокова пам'ять (LSTM). Для порівняльного аналізу також використовується багатошаровий перцептрон (MLP).

Основна частина

RNN, CNN, TCNN та LSTM – це моделі штучних нейронних мереж, що відносяться до сімейства моделей глибокого навчання. Кожна з моделей може мати три типу шарів: вхідний, прихований і вихідний шар, з'єднані ациклічними зв'язками. Моделі також можуть мати більше одного прихованого шару [8].

Використана архітектура ШНМ на основі RNN для прогнозування часових рядів передбачає декілька фіксованих функціональних блоків активації, по одному на кожний часовий крок [9]. Кожен модуль має внутрішній стан, який визначає попередні знання, які мережа містить на даному етапі. Внутрішній стан оновлюється на кожному кроці, щоб відобразити зміну знань мережі про минуле на поточному етапі [9].

Прихований стан оновлюється з використанням наступного рекурентного відношення:

$$h(t) = f(h_{t-1}, x_t), \quad (1)$$

де h_t – поточний стан, h_{t-1} – попередній стан, x_t – стан входу. У якості функції активації використовується гіперболічний тангенс зваженої суми:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{hx}x_t), \quad (2)$$

де W_{hh} – вага рекурентного нейрона, W_{hx} – вага вхідного нейрона.

Використовувана для прогнозування часових рядів CNN складається з вхідного шару, певної кількості прихованих шарів та вихідного шару [10].

Приховані шари зазвичай складаються зі згорткових шарів, шарів агрегування, нормалізуючих та повнозв'язних шарів. Ці шари пов'язані між собою шарами з визначеними активаційними функціями. Головним елементом згорткової нейронної мережі для прогнозування часових рядів є згорткові шари, де до даних з попереднього шару застосовується операція згортки [10].

Ієрархічна архітектура для прогнозування часових рядів TCNN складається з кількох згорткових прихованих шарів. TCNN використовує три основні техніки: причинні згортки, розширені згортки та обробку залишкових зв'язків.

При використанні техніки причинної згортки вихідні дані в момент часу t згортаються лише з елементами з часу t або більш ранніх часових кроків з попереднього рівня [11]. Нульове доповнення використовується в прихованих шарах, щоб забезпечити, що вони мають ту саму розмірність, що й вхідний шар, щоб спростити згортку.

Методика розширеної згортки дозволяє використати довгу пам'ять, що неможливо лише за допомогою причинних згортки. Розширений згортковий оператор F на елементі послідовності s визначається як:

$$F(s) = \sum_{i=0}^{k-1} f(i) * x_{s-d+1}, \quad (3)$$

де x – це послідовний вхід, k – розмір фільтра, d – коефіцієнт розширення.

Третя техніка TCNN реалізована за допомогою залишкових блоків [12], що допомагають подолати проблему зникнення градієнта в мережах з багатьма шарами.

Використовувана для прогнозування часових рядів ШНМ на основі архітектури LSTM складається з блоку входу, блоку виходу, вхідного шлюзу та вихідного шлюзу. Комірки пам'яті або комірки LSTM можна розглядати як шар нейронів у традиційній нейронній мережі прямого зв'язку, де кожен нейрон має прихований шар і поточний стан. LSTM має стан комірки, представлений $C(t-1)$ і $C(t)$ для попередньої та поточної міток часу відповідно [13].

Прихований стан реалізує короткочасну пам'ять, а стан комірки використовується у якості довгострокової пам'яті. При використанні комірки штучної нейронної мережі на основі LSTM спочатку визначається збереження інформації з попереднього часового кроку, на підставі:

$$f_t = \sigma(x_t * U_f + H_{t-1} * W_f), \quad (4)$$

де x_t – введення для поточної позначки часу, U_f – вага, пов'язана з введенням, H_{t-1} – прихований стан попередньої позначки часу, W_f – це вагова матриця, пов'язана з прихованим станом [14].

Вхідний шлюз використовується для кількісної оцінки важливості наданої інформації, за допомогою наступного рівняння вхідного вентиля:

$$i_t = \sigma(x_t * U_i + H_{t-1} * W_i), \quad (5)$$

де x_t – введення з поточною міткою часу t , U_i – вагова

матриця введення, H_{t-1} – прихований стан у попередній мітці часу, W_i – вагова матриця введення, пов'язаного з прихованим станом.

Вихідний шлюз ШНМ поєднує поточний вхід, вихід блоку та значення клітинки пам'яті на останній ітерації:

$$O_t = \sigma(w_o x_t + r_o y_{t-1} + p_o * c_{t-1} + b_o), \quad (6)$$

де w_o , r_o та p_o – вагові коефіцієнти, b_o – вектор зміщення. Тепер, щоб обчислити поточний прихований стан, використовується O_t і гіперболічних тангенс оновленого стану комірки. Таким чином, прихований стан є функцією довгострокової пам'яті та поточного виведення.

Одним із важливих аспектів удосконалення моделей глибокого навчання є оптимізація гіперпараметрів, яка стосується процесу вибору найкращого набору регульованих параметрів, які контролюють процес навчання моделі та можуть суттєво впливати на її продуктивність [15]. Одними з найпоширеніших методів оптимізації за гіперпараметрами є методи випадкового пошуку, байєсовської оптимізації та генетичні алгоритми.

Результати порівняльного аналізу

Для оцінки прогнозу можуть бути використані численні метрики, такі як середньоквадратична помилка (MSE), середня абсолютна помилка (MAE), середньоквадратичний квадрат помилки (RMSE), середня квадратична помилка прогнозування (MSPE) та інші. Ці метрики мають свої особливості, які важливо враховувати у контексті задачі, оскільки у дослідженні використовується часові ряди різного походження.

MAE вимірює середнє значення абсолютних відхилень, що призводить до меншої чутливості до великих значень помилок у окремих прогнозах.

MAPE визначає відсоткову різницю між фактичними та прогнозованими значеннями, що робить цю метрику корисною для наочної оцінки точності.

MSE вимірює середнє значення квадратів відхилень між фактичними та передбаченими значеннями, і ця метрика доцільна при виявленні аномальних викидів у даних.

Вибір метрики забезпечує розуміння особливостей помилок і недоліків роботи досліджуваної моделі прогнозування часових рядів та її ефективності для конкретного випадку використання.

Для порівняльного аналізу моделей прогнозування на основі архітектур RNN, CNN, LSTM та TCN на першому етапі експериментальних досліджень обраний набір даних, що містить інформацію про розвиток пандемії COVID-19 в Україні. Дані представлені у вигляді більш ніж 135 тисяч записів, кожна з яких складається з 64 полів.

Часові ряди для досліджень можуть містити неповні, відсутні або повторювані дані. Для використання викривлених даних для навчання моделей прогнозування застосовуються підходи щодо попередньої обробки, які полягають у процесі перетворення необроблених даних у значущі за допомогою різних методів.

Першим етапом підготовки вихідних даних часових рядів є усунення пропусків (відсутніх елементів часового ряду).

При навчанні моделей ШНМ наявність пропусків може призводити до неможливості навчання моделі або зниження її точності. При побудові навчальної вибірки за даними стовпця `total_tests`, який був обраний у якості цільового, можна помітити, що дані не мають значних викидів, а кількість та розташування пропусків дозволяє застосувати поліноміальну інтерполяцію для їх заповнення. Цей метод полягає у знаходженні та використанні полінома найменшого можливого ступеня, який повинен проходити через точки набору даних.

Після першого етапу необхідно застосувати згладжування, для усунення помітних позитивних і негативних викидів. Існує багато методів згладжування, але найбільш поширеними є метод ковзних середніх та експонціальне згладжування. На відміну від методу ковзних середніх, в якому всі дані мають однакову вагу, в методі експоненціального згладжування найбільший коефіцієнт застосовується до останнього спостереження.

Для згладжування часового ряду, в якому є викиди даних, доцільним є метод ковзного середнього, який і був використаний при подальших експериментальних дослідженнях.

Для визначення впливу попередньої обробки даних на якість прогнозування часових рядів, використовувалися нейронні мережі в базовій конфігурації при горизонті прогнозування 20 днів і розмірі історії 60 днів.

Аналіз виконується за допомогою метрики MSE, що розраховується згідно:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (7)$$

де n – кількість прикладів в наборі даних, y_i – фактичне значення для i -го прикладу, \hat{y}_i – прогнозоване значення для i -го прикладу.

В таблиці 1 представлені результати тестування впливу попередньої обробки даних на результати прогнозування часових рядів, у якості метрики у цьому дослідженні обрана MSE.

Таблиця 1 – Вплив використання попередньої обробки даних на результати прогнозування

Варіант обробки	Значення Mean Squared Error для моделей прогнозування на основі таких архітектур ШНМ:			
	LSTM	RNN	TCN	CNN
Без попередньої обробки	0,6205	0,5052	0,2805	0,5623
Згладжування <code>new_cases</code>	0,4623	0,4052	0,2001	0,4281
Згладжування <code>new_tests</code>	0,5648	0,4896	0,2249	0,4913
Згладжування <code>new_cases</code> та <code>new_tests</code>	0,4227	0,4006	0,1922	0,4089

Найбільш чутливою до відсутності згладжування виявилася штучна нейронна мережа на основі LSTM. Кращі результати, ніж LSTM, показала ШНМ на основі CNN, але її результати гірші за ШНМ на основі RNN та TCN. Найбільш стійкою до відсутності згладжування є ШНМ на основі TCNN.

У ході дослідження ефективності прогнозування даних щодо поширення COVID-19 було створено 60 варіантів нейронних мереж на основі RNN, CNN, LSTM та TCN архітектур із застосуванням наступних наборів гіперпараметрів:

- для ШНМ на основі LSTM: Units – 16; Activation – tanh; Dropout – 0.4; PredictionHorizon – 20;
- для ШНМ на основі RNN: Units – 64; Activation – tanh; Dropout – 0.4, PredictionHorizon – 20;
- для ШНМ на основі TCN: Filters – 64, Activation – relu, Dropout – 0.3, KernelSize – 3, Dilations – 1, 2, 4, 8, 16, 32; PredictionHorizon – 20;
- для ШНМ на основі CNN: Filters – 64, Activation – relu; Dropout – 0,3; KernelSize – 2; Stride – 1; PredictionHorizon – 20;
- глобальними гіперпараметрами були визначені Epoch – 25 та Batch – 10.

Для візуального аналізу якості прогнозування числових рядів обрано конфігурації нейронних мереж, у яких за результатами експериментів було найменше значення MSE для прогнозування на 20 днів уперед при врахуванні згладжування даних у стовпцях new_tests та new_cases:

- LSTM (Units = 8; Dropout = 0.4, Activation = tanh). Досягнуте значення MSE дорівнює 0,3176;
- RNN (Units = 128; Dropout = 0.4, Activation = tanh). Досягнуте значення MSE дорівнює 0,1103;
- TCNN (Filters = 64; KernelSize = 3; Dilations = 1,4,16,64; Dropout = 0.3). Досягнуте значення MSE дорівнює 0,1090;
- CNN (Filters = 128; KernelSize = 2; KernelStride = 1). Досягнуте значення MSE дорівнює 0,3702.

В якості тестових даних з перевірконого набору були обрані 2 проміжки зі зростаючим і спадним трендом.

Мережі на основі архітектур RNN і TCN врахували можливе зростання захворюваності також як і LSTM, проте прогнозована кількість нових випадків захворювання виявилася вищою за реальні дані.

Необхідність підбору значень гіперпараметрів є однією з проблем при використанні моделей на основі ШНМ глибокого навчання, тому були проведені дослідження впливу гіперпараметрів Units і Dropout на середньоквадратичну помилку прогнозування та час, необхідний для навчання ШНМ, результати наведені у табл. 2.

Проаналізувавши результати тестування впливу гіперпараметрів, можна зробити висновок, що зі збільшенням розмірності вихідного простору (гіперпараметр units) у шарах LSTM-мережі починається поступове збільшення значення MSE і часу навчання. Зміна значення Dropout до 0,8 дозволила уповільнити збільшення MSE та процес перенавчання.

При додаванні нового шару значення MSE збільшується на 0,1, а при зменшенні гіперпараметру units у першому шарі до 8, значення MSE зменшилося до 0,3176.

Подальше зменшення гіперпараметру units у першому шарі призвело до збільшення MSE та втрати точності прогнозування.

Таблиця 2 – Результати тестування впливу гіперпараметрів на ШНМ на основі LSTM

Units	Dropout	MSE	Час навчання
2	0,4	0,6467	18
4	0,4	0,4113	18
8	0,4	0,3176	18
16	0,4	0,405	18
32	0,4	0,4397	18
64	0,4	0,5131	36
128	0,4	0,5023	35
64	0,8	0,4037	23

У ході експериментів для аналізу використання архітектури LSTM були також використані історичні дані курсів акцій NASDAQ.

В процесі розрахунків використовувалися техніки масштабування для формування вхідних та вихідних даних для моделі. Метрики помилок прогнозування для отриманих результатів прогнозування на різних кроках у межах горизонту прогнозування наведені у табл. 3.

Таблиця 3 – Реальні та прогнозовані значення, значення абсолютної та абсолютної у відсотках помилки

Фактичне	Прогнозоване	АЕ	АРЕ, %
5220,5	5250,1	29,6	0,56
5039,25	5217,92	178,67	3,54
5086,5	5184,37	97,87	1,92
4936,75	5149,92	213,17	4,31
5224,5	5115	109,5	2,09
5282,75	5080,55	202,2	3,82
5380,5	5047,56	332,94	6,18
5572,5	5017,3	555,2	9,96
5626,5	4991	635,5	11,29
5711,5	4969,81	741,69	12,98
5717,75	4954,63	763,12	13,34
5568,5	4945,88	622,62	11,18
5499	4943,09	555,91	10,10

Наведені розрахунки помилок використані для визначення середніх метрик MAE та MAPE:

$$MAE = \frac{\sum_{i=1}^k |y_i - x_i|}{k}, \quad (8)$$

$$MAPE = \frac{100\%}{k} \sum_{i=1}^k \left| \frac{y_i - x_i}{y_i} \right|, \quad (9)$$

де k – кількість пар фактичних (y) та прогнозованих (x) значень.

Візуалізація отриманих результатів порівняння фактичних та прогнозованих значень курсів акцій NASDAQ представлена на рис. 1.

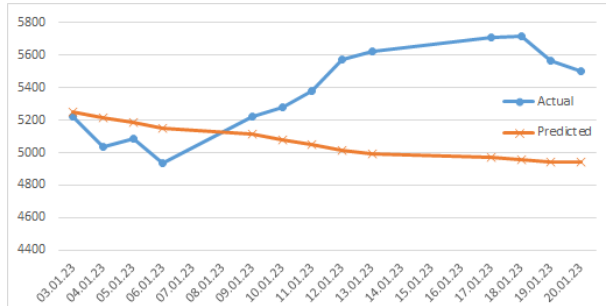


Рис. 1. Результати прогнозування значень курсів акцій NASDAQ за допомогою ШНМ на основі LSTM

Отримані середні значення метрик (MAE – 387,53, MAPE – 7,02) свідчать про успішність застосування нейронних мереж на основі LSTM, принаймні, для короткострокового прогнозу курсів акцій NASDAQ. У ході експериментального дослідження також були використані дані середньорічної температури повітря, прогнозування цих часових рядів є важливою задачею в галузі кліматології та метеорології. Для порівняння були обрані архітектури LSTM та MLP, вхідні дані для аналізу обмежені періодом 2009-2023 років, результати наведені у табл. 4.

Таблиця 4 – Реальні та прогнозовані значення середньорічної температури (градуси за Фаренгейтом) для ШНМ на основі архітектур MLP та LSTM

Рік	Фактичне значення	Прогноз (MLP)	Прогноз (LSTM)
2009	54,4	53,651	53,503
2010	53,4	53,738	53,558
2011	51,1	53,607	53,495
2012	53,5	53,177	53,234
2013	52,3	53,517	53,164
2014	53,2	52,995	53,016
2015	54,8	53,497	52,930
2016	54,4	52,998	53,024
2017	54,7	53,159	53,038
2018	53,7	53,567	53,158
2019	53,5	53,102	53,184
2020	53,1	53,282	53,153
2021	54,5	53,901	53,135
2022	54,0	53,478	53,358
2023	53,9	53,934	53,395

Аналізуючи значення метрик ME, MAE, MPE, MAPE, SMAPE, SSE, MSE та RMSE, що наведені у табл. 5, можна визначити, що модель MLP виявилася ефективнішою у короткостроковому прогнозуванні середньорічної температури повітря порівняно з використанням моделі глибокого навчання.

Таблиця 5 – Результати прогнозування часових рядів середньорічної температури

Метрика помилки	ШНМ на основі MLP	ШНМ на основі LSTM
ME	0,19	0,4
MAE	0,76	0,87
MPE	0,32%	0,73%
MAPE	1,43%	1,63%
SMAPE	1,42%	1,63%
SSE	15,46	18,49
MSE	1,03	1,23
RMSE	1,01	1,11

За результатами дослідження, модель MLP продемонструвала низьку MAE (0,76) та MAPE (1,43%), вказуючи на прийнятну точність прогнозування середньорічної температури повітря. Графічне відображення порівняння прогнозів за допомогою MLP та LSTM наведено на рис. 2.

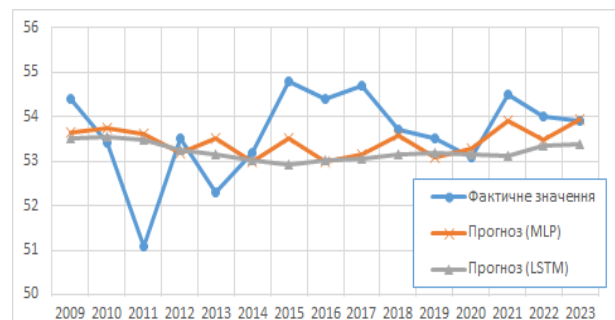


Рис. 2. Результати прогнозування середньорічної температури за допомогою ШНМ на основі LSTM та MLP

Результати експериментальних досліджень вказують на те, що модель на основі MLP виявляється ефективнішою при короткостроковому прогнозуванні необроблених даних середньорічної температури повітря. Це підтверджується низькими значеннями MAE, MAPE, ME та MPE, а також меншими значеннями SMAPE порівняно з LSTM.

З огляду на отримані результати прогнозування даних щодо поширення COVID-19 та курсів акцій NASDAQ можна дійти висновку, що найкращі результати показує мережа на основі архітектури LSTM, але вимагає попередньої обробки вихідних даних. Результати мережі на основі TCNN мають меншу точність в порівнянні з результатами LSTM. Мережі RNN та CNN застосовувати недоцільно, бо результати, отримані за допомогою цих мереж, значно відрізняються від реальних даних.

Висновки

В роботі проведений порівняльний аналіз використання моделей штучних нейронних мереж глибокого навчання (LSTM, TCN, CNN, та RNN) для прогнозування часових рядів різного походження: дані щодо поширення COVID-19, історичні дані курсів акцій NASDAQ та часовий ряд середньорічної температури повітря.

У випадку з прогнозуванням числа захворілих під час пандемії COVID-19 найкращий результат показала LSTM мережа, яка має високу ефективність і при прогнозуванні даних NASDAQ. Але запропонована модель прогнозування на основі архітектури LSTM є чутливою до якості вихідних даних, також

недостатнє налаштування гіперпараметрів може призводити до значного погіршення точності, що було підтверджено низькими результатами прогнозування середньорічної температури повітря.

Дослідження підтвердило, що моделі глибокого навчання, такі як LSTM, TCNN, CNN, та RNN, можуть ефективно вирішувати завдання прогнозування часових рядів, проте важливо враховувати, що успішність залежить від правильного вибору моделі, оптимальних значень гіперпараметрів, належного навчання та наявності відповідної кількості даних.

Отримані результати можуть служити підґрунтям для розвитку ефективних моделей прогнозування та підтримки прийняття рішень.

СПИСОК ЛІТЕРАТУРИ

1. Chatfield C. (2000), "Time-Series Forecasting", Chapman and Hall, P. 280.
2. Lin Y., Korsinska Y., Rana M. (2021), "Temporal Convolutional Attention Neural Networks for Time Series Forecasting", International Joint Conference on Neural Networks, pp. 236-238.
3. Pedro H., Coimbra C. (2012), "Assessment of forecasting techniques for solar power production with no exogenous inputs", Solar Energy, № 86(7), pp. 2017-2028.
4. Amin Salih Mohammed, Ivashchenko H., Filimonchuk T., Ivanisenko I., Barkovska O. (2020) "Green Hybrid Models Based on Clonal Selection and Case-based Reasoning for Short-term Time Series Forecasting", Journal of Green Engineering, vol. 10, issue 5, pp. 2139-2154.
5. Shahin A. (2021), "Deep Learning BiLSTM Encoding-Decoding Model for COVID-19 Pandemic Spread Forecasting", Fractal and Fractional, №4. pp. 175-176.
6. Marzouka M. (2021), "Deep learning model for forecasting COVID-19 outbreak in Egypt", Process Safety and Environmental Protection, № 153, pp. 363-375.
7. Rangapuram S., Seeger M., Gasthaus D., Stella L., Wang Y, Januschowski T. (2018), "Deep state space models for time series forecasting", Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), P. 10.
8. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser Ł., Polosukhin I. (2017), "Attention is all you need", Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), P. 11.
9. Derbentsev V., Matviychuk A., Datsenko N., Bezkorovainyi V., Azaryan A. (2020), "Machine learning approaches for financial time series forecasting", Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy, № 2713, Odessa, Ukraine, pp. 434-450.
10. Hemmati A., Abdoos M., Akbar A. (2015), "Short term load forecasting using a hybrid intelligent method", Knowledge-Based System, № 76, pp. 139-147.
11. Voyant C., Nivet M., Paoli C., Muselli M., Notton G. (2014), "Meteorological time series forecasting based on MLP modelling using heterogeneous transfer functions", Journal of Physics: Conference Series, № 574, Madrid, Spain, pp. 28-31
12. Khan S., Rahmani H., Ali Shakh S. A., Bennamoun M (2018), "Guide to Convolutional Neural Networks for Computer Vision", Morgan & Claypool Publishers, P. 207.
13. Goodfellow I., Bengio Y., Courville A (2016), "Deep Learning (Adaptive Computation and Machine Learning series)", The MIT Press, P. 775.
14. Greff, K., Srivastava, R., Koutnik, J., Steunebrink, B., Schmidhuber, J. (2017), "LSTM: A Search Space Odyssey", IEEE Transactions on Neural Networks and Learning Systems, № 28, pp. 2222-2232.
15. Breiman L. (2001), "Random forests", Machine learning, № 45(1), pp. 5-32.

Received (Надійшла) 11.12.2023

Accepted for publication (Прийнята до друку) 31.01.2024

Deep learning models for time series forecasting

Heorhii Ivashchenko, Daria Tymoshenko, Oleksandr Blyzniuk, Oleksandr Kononenko

Abstract. Topicality. Time series forecasting is one of the important tools for various spheres of human activity, as it allows analyzing past trends, understanding the dynamics of events, and making substantiated decisions based on previously collected historical data. In recent years, deep learning artificial neural network models have demonstrated significant potential in the field of time series forecasting. **The goal of this work** is to analyze the use of deep learning models for short-term forecasting of time series of various origins and with possible presence of distortions. **The object of research** is the process of time series forecasting. **The subject of research** is the use of models based on CNN, RNN, TCNN and LSTM architectures for time series forecasting. **Results.** Experimental research has shown that forecasts of non-stationary time series using an artificial neural network based on LSTM architecture are closer to real data, compared to other deep learning models. **Conclusions.** The obtained results in most cases confirm the advantage of using models based on LSTM over other considered deep learning models for time series forecasting.

Keywords: time series forecasting, machine learning, artificial neural network, deep learning models, convolutional networks, long short-term memory.