

using a tailored inventory management system in the military aviation industry. *Heliyon*. 2020. Vol. 6, iss. 7. <https://doi.org/10.1016/j.heliyon.2020.e04424>.

28. Fedorovich, O., Lutai, L., Kompanets, V., Bahaiev, I. The Creation of an Optimisation Component-Oriented Model for the Formation of the Architecture of Science-Based Products. *Integrated Computer Technologies in Mechanical Engineering - 2023. ICTM 2023*. Lecture Notes in Networks and Systems, Springer, Cham, 2024. Vol. 996. P. 415–426. https://doi.org/10.1007/978-3-031-60549-9_31.

Надійшла до редколегії 14.11.2024 р.

Liudmyla Lutai, Candidate of Technical Sciences, Associate Professor, Associate Professor of the department Mechatronics and Electrical Engineering, National Aerospace University «Kharkiv Aviation Institute», Kharkiv, Ukraine, e-mail: l.lutay@khai.edu, ORCID: <https://orcid.org/0000-0003-1822-8938>.

УДК 004.8

DOI: 10.30837/0135-1710.2024.183.035

І.Ю. ПАНФЬОРОВА, Н.Б. ГАНЖИЛА

ДОСЛІДЖЕННЯ АРХІТЕКТУР НЕЙРОННИХ МЕРЕЖ ДЛЯ ПІДВИЩЕННЯ ТОЧНОСТІ ПРОГНОЗУВАННЯ ПОПИТУ НА ПРОДУКЦІЮ

Розглянуто використання різних архітектур нейронних мереж для задач прогнозування попиту на продукцію. Проаналізовано основні виклики та сучасні проблеми в області прогнозування, що виникають при роботі з великими обсягами даних і сезонними коливаннями. Особливу увагу приділено порівнянню архітектур нейронних мереж за кількома ключовими характеристиками, зокрема, Accuracy, F1-Score, Logarithmic, середньою абсолютною помилкою, коренем середньоквадратичної помилки та AUC-ROC. На основі детального аналізу та проведеного експерименту було визначено архітектуру нейронної мережі, яка забезпечує найвищу точність прогнозування при прогнозуванні попиту на продукцію.

1. Вступ

Прогнозування попиту на продукцію є важливою складовою успішної діяльності будь-якого підприємства, оскільки якість та достовірність прогнозів безпосередньо впливає на ефективність планування виробничих та логістичних процесів. Неточні прогнози можуть призвести до виникнення надлишкових запасів, що збільшує витрати на зберігання, або до нестачі продукції, що може спричинити зниження рівня обслуговування клієнтів і втрату доходу. У сучасних умовах, коли обсяги даних значно зросли, з'явилися нові можливості для застосування передових технологій аналізу, зокрема нейронних мереж. Використання нейронних мереж дозволяє підприємствам виявляти приховані закономірності в історичних даних, прогнозувати зміни попиту та формувати точніші прогнози, що допомагає підприємству своєчасно адаптуватися до ринкових умов [1].

Однак вибір архітектури нейронної мережі є складним завданням, оскільки різні архітектури мають свої переваги та недоліки в залежності від умов застосування. Наприклад, деякі архітектури можуть давати точніші результати при обробці великих обсягів історичних даних, тоді як інші краще справляються з задачами прогнозування в реальному часі або за нестандартними умовами. Незважаючи на значний прогрес у використанні нейронних мереж для задач прогнозування, питання вибору архітектури нейронної мережі, що надає точний результат, залишається недостатньо вивченим. Важливим напрямом сучасних досліджень є вивчення та вдосконалення існуючих архітектур для досягнення найкращих результатів у специфічних умовах, зокрема у виробничих і логістичних процесах, де попит на продукцію може змінюватися сезонно або

залежати від зовнішніх економічних факторів [1]-[2].

Використання нейронних мереж у задачах прогнозування супроводжується низкою викликів. Серед них особливо важливими є необхідність обробки великих обсягів даних, забезпечення стабільності результатів, інтеграція нейронної мережі в існуючі інформаційні системи підприємств та визначення обчислювальних ресурсів для використання мережі. Наприклад, для ефективної роботи з великими даними необхідно враховувати не лише швидкість обробки, а й можливість масштабування системи на майбутнє, щоб забезпечити її результативність при збільшенні обсягів даних. Інтеграція нейронних мереж у вже існуючі інформаційні системи може бути ускладнена вимогами до сумісності з іншими програмними засобами, що використовуються в підприємстві, а також необхідністю налаштування архітектури мережі відповідно до конкретних потреб бізнесу. Крім того, зростає потреба в покращенні обчислювальних ресурсів, що може вплинути на продуктивність та вартість реалізації остаточної системи. Використання нейронних мереж може вимагати значних обчислювальних потужностей, що є важливим аспектом при проектуванні ефективних систем прогнозування. Проведення досліджень, спрямованих на визначення архітектури нейронної мережі, що дає найточніший результат прогнозу, є важливим як для теоретичного розвитку, так і для практичної реалізації систем прогнозування. Це дозволяє не лише підвищити точність моделювання попиту, але й забезпечити надійність і сталість результатів при змінних умовах.

2. Аналіз літературних джерел та визначення проблеми дослідження

У сфері прогнозування попиту на продукцію значну увагу приділяють використанню різних архітектур нейронних мереж, зокрема, глибоких нейронних мереж (DNN – Deep Neural Networks), рекурентних нейронних мереж (RNN – Recurrent Neural Networks), а також згорткових нейронних мереж (CNN – Convolutional Neural Networks). В існуючих дослідженнях пропонується поділяти архітектури на групи залежно від типу навчання: контрольоване навчання та самонавчання [2].

Архітектури нейронних мереж, що використовують контрольоване навчання, працюють з даними, де кожен вхідний елемент супроводжується відповідною міткою або цільовим значенням. Такі мережі навчаються на прикладах з відомими правильними відповідями, що дає можливість прогнозувати нові дані. Такий тип нейронної мережі зазвичай використовується для завдань, де потрібен максимально точний результат, таких як прогнозування попиту. Прикладом таких архітектур є DNN, які часто застосовуються для вирішення складних задач прогнозування з великими обсягами даних [3].

Архітектури з самонавчанням працюють з даними без міток. Метою використання таких мереж є виявлення прихованих структур і зв'язків у даних, що дозволяє знаходити групи схожих об'єктів або виявляти шаблони без заздалегідь визначених цільових значень. Такі мережі застосовуються для кластеризації, зменшення розмірності даних та виявлення аномалій. Одним із таких прикладів є застосування нейронних мереж для кластеризації продукції на основі історичних даних попиту для подальшого прогнозування [4].

CNN є ефективними для аналізу просторових даних, такими як зображення, тоді як RNN та їх варіації, зокрема Long Short-Term Memory (LSTM), демонструють відмінні результати при роботі з послідовними даними, такими як часові ряди попиту. Проте RNN мають обмеження у задачах з довгими часовими рядами, оскільки при роботі з великими історичними даними стикаються з проблемою зникання градієнта, що погіршує їхню ефективність.

Попри позитивні результати використання LSTM для прогнозування попиту, існують проблеми з налаштуванням архітектури для адаптації до нових ринкових умов або змін у поведінці споживачів. Деякі дослідження також вказують на потенціал комбінування

різних типів нейронних мереж для покращення якісних характеристик прогнозів, наприклад, інтеграція CNN для попереднього аналізу даних і LSTM для прогнозування на основі часових рядів [5].

Таким чином, невирішеними залишаються питання щодо вибору архітектури для точнішого прогнозування попиту, яка здатна враховувати всі аспекти змінних даних попиту, включаючи сезонні коливання, тренди та випадкові коливання, що впливають на попит у реальних умовах. Особливу складність становить вибір такої архітектури, яка могла б адаптуватися до специфіки конкретних умов підприємства, таких як масштаби виробництва, циклічність замовлень та непередбачувані зовнішні фактори, що змінюють модель попиту. Це дослідження зосереджується на визначенні архітектури, яка не лише забезпечує високу стабільність та достовірність прогнозів, але й є стійкою до змін вхідних даних та швидко адаптується до нових трендів у попиті. Питання вибору архітектури у контексті забезпечення високої швидкості обробки та економії ресурсів підприємства залишаються актуальними для багатьох компаній, що прагнуть підвищити достовірність своїх прогнозів та мінімізувати ризики надлишкових запасів чи дефіциту продукції [6]-[7].

3. Мета і задачі дослідження

Метою цього дослідження є визначення архітектури нейронної мережі, яка забезпечує точніший результат прогнозування попиту на продукцію. Досягнення цієї мети дозволить визначити найпродуктивніші підходи для точного прогнозування та сприятиме підвищенню достовірності прогнозів. Таке підвищення дозволить підприємствам покращити виробничу ефективність та зменшити ризики, пов'язані з надлишковими запасами або дефіцитом продукції.

Для досягнення поставленої мети пропонується вирішити такі задачі:

- визначити основні переваги та обмеження архітектур нейронних мереж, обраних для задачі прогнозування попиту;
- сформулювати вибірку історичних даних для навчання та тестування обраних нейронних мереж;
- провести порівняльний аналіз архітектур нейронних мереж у контексті задач прогнозування попиту для виявлення архітектури яка забезпечує точніший результат прогнозування.

4. Матеріали і методи дослідження

Об'єктом дослідження є процес прогнозування попиту на продукцію за допомогою нейронних мереж. Предметом дослідження є різні архітектури нейронних мереж, застосування яких слід оцінити за результативністю в контексті задач прогнозування попиту на продукцію. Основною гіпотезою дослідження є те, що різні архітектури нейронних мереж (LSTM, CNN, FNN, RNN, Gated Recurrent Units (GRU), Attention-based models (ABM), Autoencoders (AE)) можуть продемонструвати різну результативність та достовірність результатів у прогнозуванні попиту на продукцію в залежності від особливостей даних і структурного підходу.

Задача порівняльного аналізу є однією з основних задач системного аналізу, що передбачає порівняння кількох альтернативних підходів для вибору найкращого. У контексті задачі прогнозування попиту ця задача набуває особливостей, оскільки для кожної архітектури нейронної мережі потрібно оцінити її здатність до виявлення закономірностей у даних, що можуть мати різну природу (часові, просторові, тренди та цикли). Адаптація цієї задачі до особливостей архітектур нейронних мереж включає в себе визначення метрик оцінки, які дозволяють врахувати різні аспекти роботи моделей, такі як точність прогнозування, стабільність та швидкість навчання.

Методи порівняльного аналізу архітектур нейронних мереж включають:

– аналіз достовірності результатів, тобто порівняння результатів кожної архітектури нейронної мережі за метриками, такими як Accuracy, середня абсолютна похибка, середньоквадратична похибка, Logarithmic Loss, F1-Score, а також AUC-ROC для оцінки загальної якості прогнозування;

– аналіз надійності мереж, а саме визначення стабільності архітектури при використанні різних підмножин даних та перевірка здатності до генералізації на нових наборах даних;

– аналіз швидкості навчання та обробки даних, а саме порівняння часу тренування мереж та здатності до масштабування на великих наборах даних.

Для всіх архітектур використовувалися однакові набори даних, що включають історичні дані про реалізацію продукції, а також зовнішні фактори, які можуть впливати на попит, такі як сезонні коливання, економічні показники або інші зміни на ринку. Перед проведенням порівняльного аналізу було проведено попередню обробку та підготовку вибірки історичних даних, а саме: очищення даних від шуму та помилок, нормалізація значень для забезпечення коректної роботи мереж, а також вибір ключових ознак з використанням методів відбору ознак для підвищення достовірності прогнозування.

Оцінка результатів прогнозування проводилася на основі таких архітектур нейронних мереж: LSTM, CNN, FNN, RNN, GRU, ABM, AE. Вибір цих архітектур зумовлений їхньою здатністю до обробки різних типів даних та специфіки задачі прогнозування попиту. Зокрема, LSTM та GRU були вибрані завдяки їхній ефективності у роботі з часовими залежностями, CNN – для виявлення локальних шаблонів в історичних даних, а FNN, RNN, ABM та AE доповнюють спектр можливих варіантів архітектур для аналізу різних аспектів даних.

Використання запропонованого методу дослідження дозволило визначити архітектуру нейронної мережі, яка забезпечує точніше прогнозування попиту, а також дало змогу глибше зрозуміти, як різні архітектури нейронних мереж, можуть бути застосовані в реальних виробничих процесах для підвищення ефективності їхньої діяльності, зокрема, в плануванні виробництва та управлінні запасами, що дозволяє прискорити прийняття рішень і покращити достовірність прогнозів.

5. Результати дослідження

У ході дослідження було виділено основні переваги та недоліки вибраних варіантів архітектур нейронних мереж

Рекурентні нейронні мережі (RNN) здатні обробляти послідовні дані, такі як часові ряди. Вони мають внутрішні цикли, що дозволяють зберігати інформацію про попередні етапи обробки, але часто стикаються з проблемою зникаючих градієнтів. Це обмежує здатність моделювати довготривалі залежності в даних. Вони ефективні на коротших послідовностях, але їхня результативність знижується на складніших завданнях [8]-[9].

Long Short-Term Memory (LSTM) є спеціалізованою архітектурою RNN, що вирішує проблему зникаючих градієнтів. Вона використовує клітинні стани для зберігання інформації на довгі періоди, що робить її ефективною для задач з тривалими залежностями, такими як сезонні коливання попиту. Однак її складна структура потребує більших обчислювальних ресурсів, що уповільнює процес навчання порівняно з іншими методами [9].

Конволюційні нейронні мережі (CNN) традиційно використовуються для обробки зображень, але інколи також застосовуються до часових рядів, де вони виявляють локальні шаблони, що можуть бути корисними для короткотермінових коливань попиту. CNN зазвичай мають вищу швидкість навчання та менші вимоги до обчислювальних ресурсів, але їхня здатність моделювати довготривалі тренди обмежена, що знижує їхню

ефективність для складніших прогнозувальних завдань.

Feedforward нейронні мережі (FNN) не мають зворотних зв'язків і тому не можуть враховувати залежності від попередніх значень. Вони можуть бути швидкими в навчанні і здатні давати прийнятні результати в простих задачах, де залежності між даними не є значними. Однак для задач прогнозування попиту, де врахування історичних даних є критично важливим, ефективність таких мереж є обмеженою [10].

Gated Recurrent Units (GRU) є спрощеною версією LSTM, що має меншу кількість параметрів та простішу архітектуру. Вони використовують два основні механізми «оновлення» та «заборону», що дозволяє ефективно зберігати довготривалі залежності в даних з меншими обчислювальними витратами, ніж у LSTM. Це робить GRU ефективними для задач прогнозування попиту при меншій потребі в обчислювальних ресурсах.

Моделі з механізмами уваги (ABM) є потужними для задач прогнозування, оскільки дають змогу фокусуватися на важливих частинах вхідних даних, враховуючи не лише часову інформацію, але й взаємозв'язки між різними елементами. Це робить їх ефективними для складних завдань прогнозування попиту [11].

Autoencoders (AE) використовуються для зменшення розмірності та виявлення латентних властивостей даних. Вони можуть застосовуватися для передобробки даних перед основним етапом прогнозування, виділяючи основні шаблони з великих обсягів даних.

Для проведення дослідження обраних архітектур нейронних мереж для прогнозування попиту на продукцію було використано навчальні та тестові набори даних, що включають історичні дані про реалізацію продукції, а також додаткові атрибути, пов'язані з продукцією. Історичні дані про реалізацію було отримано з внутрішніх систем підприємства, що спеціалізується на виробництві та продажу абразивних матеріалів, таких як шліфувальні та полірувальні круги, круги пелюсткові, абразивні стрічки тощо. Дані містять інформацію про продажі продукції за різні періоди, що дозволяє детально аналізувати попит на продукцію за часом і визначати основні тренди та сезонні коливання. Для отримання додаткових атрибутів продукції, таких як ціна, категорія, розмір, вага тощо, було використано додаткові інформаційні ресурси підприємства, включаючи файли у форматі .csv, що містять характеристики кожного товару.

Для вирішення проблеми неповних або необроблених вхідних даних застосовувався метод імпутації даних, який дозволяє заповнювати відсутні або невизначені значення шляхом генерації нових даних на основі наявної інформації. Це дозволяє зберегти корисну структуру даних і покращити навчання мережі. В рамках цього методу використовувалась техніка стохастичних пертурбацій, що полягає у введенні випадкових змін у значення вхідних даних. Ці зміни моделюють потенційні варіації даних, які можуть виникати в реальних умовах, і таким чином допомагають зробити мережу стійкішою до неповних або нечітких вхідних даних. Застосування цього методу дозволяє не лише заповнювати пропуски в даних, а й знижувати ризик перенавчання мережі, оскільки вона навчиться працювати з різноманітнішими варіантами вхідних значень.

Реалізацію кожної архітектури нейронних мереж для вирішення задачі прогнозування попиту було виконано за допомогою мови програмування Python та бібліотек TensorFlow і Keras. Для архітектури RNN було використано стандартні компоненти Keras, зокрема шар RNN для обробки часових рядів, де налаштовувалась кількість шарів і нейронів для врахування довготривалих залежностей. Для архітектури LSTM було застосовано спеціалізовані клітинні стани для зберігання інформації, що дозволяє моделювати довготривалі залежності з налаштуванням гіперпараметрів, таких як кількість епох та розмір пакетів. Для архітектури CNN використовувались шари 1D-конволюції для виявлення локальних шаблонів у часових рядах, а також пулінгові шари для зменшення

розмірності. Архітектуру GRU було реалізовано аналогічно LSTM, але з меншою кількістю параметрів, що дозволяє досягти результатів за менших обчислювальних витрат. АВМ було реалізовано за допомогою відповідних шарів для фокусування на важливих частинах вхідних даних, в той час як для архітектури Autoencoders застосовувалися шари для зменшення розмірності та виявлення латентних властивостей даних.

Дослідження проводилося на портативному комп'ютері з процесором AMD Ryzen 5 1600 @ 3,20 ГГц, 16 Гб оперативної пам'яті DDR4 і графічним процесором AMD RX 580 8 Гб DDR5, що забезпечує необхідну обчислювальну потужність для роботи з великими обсягами даних і тренуванням нейронних мереж.

Для порівняння результатів прогнозування різних архітектур нейронних мереж використовувались такі метрики, як Accuracy, F1-Score, Logarithmic Loss, середня абсолютна похибка, середнє квадратичне відхилення та AUC-ROC.

Accuracy (точність) вимірює, яка частка всіх прогнозів була правильною, та дає загальне уявлення про Accuracy результатів прогнозування мережі в контексті всіх можливих прогнозів. Формула розрахунку Accuracy має вигляд:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

де TP (True Positives) – кількість правильно передбачених позитивних класів; TN (True Negatives) – кількість правильно передбачених негативних класів; FP (False Positives) – кількість хибнопозитивних передбачень; FN (False Negatives) – кількість хибнонегативних передбачень.

F1-Score – середнє гармонійне між Precision (чіткість) та Recall (повнота), яке дозволяє отримати збалансованішу оцінку мережі, зважаючи на дві основні характеристики: Precision та Recall. Precision – це метрика, яка показує частку правильних позитивних передбачень серед усіх випадків, класифікованих як позитивні, тобто, скільки з передбачених позитивних результатів дійсно є позитивними. Recall – це метрика, яка вимірює, скільки з усіх реальних позитивних випадків було правильно передбачено мережею. Це показник здатності мережі виявляти всі позитивні випадки.

Формула розрахунку F1-Score має вигляд:

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad (2)$$

Precision розраховується за формулою:

$$Precision = \frac{TP}{TP + FP}, \quad (3)$$

Recall розраховується за формулою:

$$Recall = \frac{TP}{TP + FN}, \quad (4)$$

Logarithmic Loss (Log Loss) використовується для оцінки ймовірнісних прогнозів мережі і розраховується за формулою:

$$LogLoss = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)], \quad (5)$$

де N – кількість зразків у наборі даних; y_i – справжнє значення для i -го зразка; p_i – ймовірність передбаченого класу для i -го зразка.

Середня абсолютна похибка (MAE) вимірює середнє значення абсолютних помилок між прогнозами та реальними значеннями попиту.

Середнє квадратичне відхилення (RMSE) оцінює середнє квадратне відхилення між

прогнозами та реальними значеннями попиту.

AUC-ROC вимірює здатність мережі правильно класифікувати позитивні та негативні випадки і розраховується за формулою:

$$AUCROC = 1 - \frac{1}{2} \sum_{i=1}^N (TPR_i - FPR_i), \quad (6)$$

де TPR (True Positive Rate) – частка дійсно позитивних результатів, що правильно класифіковані; FPR (False Positive Rate) – частка хибнопозитивних результатів серед усіх негативних результатів.

Таблиця 1 демонструє результати прогнозування попиту різних архітектур нейронних мереж за Accuracy, F1-Score, Log Loss, MAE, RMSE та AUC-ROC.

Таблиця 1

Порівняння результатів прогнозування різними архітектурами за ключовими метриками

Архітектура	Accuracy	F1-Score	Log Loss	MAE	RMSE	AUC-ROC
LSTM	0.87	0.89	0.32	0.10	0.14	0.91
CNN	0.83	0.85	0.37	0.12	0.16	0.89
FNN	0.84	0.86	0.35	0.11	0.15	0.88
RNN	0.81	0.83	0.40	0.13	0.17	0.87
GRU	0.85	0.87	0.34	0.11	0.15	0.90
ABM	0.85	0.90	0.30	0.09	0.13	0.92
AE	0.82	0.84	0.38	0.12	0.16	0.80

Оцінка надійності мереж, а також аналіз швидкості навчання та обробки даних є важливими аспектами для повного порівняння архітектур нейронних мереж. Цей аналіз дозволяє отримати уявлення про точність результатів прогнозування кожної мережі в процесі навчання та прогнозування.

Оцінка надійності мереж передбачає вимірювання стабільності результатів на основі багатократного тренування мережі з однаковими початковими умовами та параметрами. Для отримання стабільності результатів проводяться кілька тренувань мережі із збереженням усіх вхідних параметрів та початкових умов, після чого вимірюється стандартне відхилення отриманих результатів. Це дозволяє встановити, наскільки передбачуваними та повторюваними є результати при багатьох запусках, що, в свою чергу, є критично важливим для забезпечення високої надійності мережі в реальних умовах, де можуть виникати зміни в даних або зовнішні фактори, що впливають на прогнози. Така стабільність дозволяє переконатися в тому, що мережа здатна виробляти точні прогнози навіть за умов, коли дані можуть бути неповними або варіативними.

Швидкість навчання, в свою чергу, оцінюється часом, необхідним для тренування кожної мережі до досягнення бажаного рівня Accuracy, що не є меншим 80 %. Час тренування вимірюється від початку процесу навчання до досягнення вказаного рівня Accuracy. Цей показник важливий для порівняння достовірності результатів прогнозування різних архітектур у реальному часі, оскільки в багатьох практичних застосуваннях важливо не тільки отримати точні прогнози, а й зробити це в найкоротші терміни. Час навчання також дозволяє оцінити, наскільки швидко мережа адаптується до нових умов і даних, що важливо для безперервного процесу прогнозування в динамічних умовах.

Таблиця 2 демонструє додаткові результати проведення прогнозування, що стосуються надійності, швидкості навчання та обробки даних для кожної з обраних архітектур нейронних мереж.

Таблиця 2

Оцінка стабільності, надійності, швидкості навчання та обробки даних для обраних архітектур

Архітектура	Час навчання (години)	Час обробки даних (сек.)	Стабільність результатів (стандартне відхилення)	Надійність (R^2)
LSTM	12	2.2	0.01	0.98
CNN	8	3.3	0.03	0.88
FNN	1	1.5	0.04	0.85
RNN	5	1.4	0.06	0.87
GRU	6	2.3	0.02	0.95
ABM	17	5.9	0.02	0.96
AE	3	3.7	0.05	0.84

Результати, представлені у таблиці 1, дозволяють здійснити порівняльний аналіз застосування обраних архітектур нейронних мереж для вирішення задачі прогнозування попиту на продукцію за ключовими метриками.

LSTM демонструє найвищий показник Ассурасу (0.87), що вказує на її високу здатність правильно прогнозувати дані в задачі прогнозування попиту. Ця архітектура випереджає інші архітектури за показником Ассурасу, зокрема ABM (0.85), яка показує майже однаковий результат, але з трохи кращими показниками F1-Score (0.90) і Log Loss (0.30), що свідчить про ефективніше балансування між точністю та відгуком, а також вищу здатність до прогнозування ймовірностей. GRU (0.85) також показує добрі результати, з F1-Score 0.87, але поступається LSTM та ABM за показниками Ассурасу і Log Loss.

CNN і FNN показують скромніші результати в порівнянні з LSTM та ABM, зі значенням Ассурасу 0.83 і 0.84 відповідно. Вони також мають дещо гірші значення за метриками MAE і RMSE, що свідчить про більші помилки в прогнозах порівняно з LSTM і ABM. Однак, вони все ще продовжують демонструвати високий показник Ассурасу у прогнозуванні, разом з F1-Score 0.85 та 0.86.

RNN зі значенням Ассурасу 0.81 демонструє найгірші результати серед всіх архітектур, що свідчить про її обмежену ефективність у задачі прогнозування попиту. Це підтверджується низькими значеннями по метриці AUC-ROC (0.87) та F1-Score (0.83).

AE, незважаючи на гарні показники Log Loss (0.38) та Ассурасу (0.82), має найнижчий результат за AUC-ROC (0.80), що вказує на недостатню здатність цієї архітектури до точної класифікації позитивних і негативних класів.

Оцінка надійності та стабільності результатів під час багатократного тренування мереж із однаковими початковими умовами підтверджує лідерство LSTM серед інших архітектур. LSTM демонструє найвищі показники стабільності (стандартне відхилення 0.01) та надійності ($R^2 = 0.98$), що свідчить про її здатність до стабільних і передбачуваних результатів. Водночас GRU показує хороші результати стабільності (стандартне відхилення 0.02) і надійності ($R^2 = 0.95$), але поступається LSTM.

ABM має високу надійність ($R^2 = 0.96$) і стабільність (стандартне відхилення = 0.02), що робить її ефективною альтернативою для прогнозування попиту, зважаючи на її здатність до стабільних результатів. У той час як AE показує порівняно низькі результати за цими метриками (стандартне відхилення 0.05, $R^2 = 0.84$), що робить її менш ефективною для задач, де важлива достовірність та стабільність.

Швидкість навчання у FNN є найкращою з усіх архітектур (1 година навчання), за нею йде AE (3 години), далі RNN та GRU (по 5 та 6 годин відповідно), в той час як CNN потребує 8 годин. Навчання LSTM триває 12 годин, що робить її однією із найповільніших,

але це не впливає на її перевагу за іншими метриками. АВМ як найскладніша архітектура потребує 17 годин для навчання, що значно більше, ніж значення цього показника для інших архітектур, що може бути важливим фактором при виборі архітектури для задач з обмеженим часом вирішення.

Таким чином, LSTM залишається найкращою архітектурою для прогнозування попиту завдяки високим результатам за показниками стабільності, надійності та Accuracy. GRU демонструє хороші результати, але поступається LSTM. АВМ є сильною альтернативою LSTM за надійністю, але значний час її навчання може бути важливим обмеженням. АЕ, хоча і має певні обмеження в стабільності та надійності, залишається корисною для специфічних умов, де важлива швидкість вирішення задачі. Архітектури CNN, RNN та FNN мають переваги в часі навчання, але поступаються LSTM та GRU за іншими показниками.

6. Обговорення результатів дослідження

Проведене дослідження показало, що вибір архітектури нейронної мережі є одним із ключових факторів для досягнення високої результативності вирішення задачі прогнозування попиту на продукцію. Кожна архітектура має свої переваги та обмеження, які залежать від специфіки задачі, що вимагає порівняння різних архітектур для вибору найкращого рішення.

У дослідженні було порівняно кілька популярних архітектур для задач прогнозування попиту, зокрема LSTM, CNN, RNN, FNN, GRU, АВМ та АЕ. Результати дослідження показали, що LSTM продемонструвала найкращий результат для показника Accuracy, досягнувши значення цього показника, вищого за значення для інших архітектур. Це свідчить про високу надійність LSTM у контексті автоматизованих систем прогнозування попиту, здатних ефективно враховувати сезонні тренди та циклічність даних.

LSTM має здатність обробляти довготривалі залежності в даних, що є критично важливим для задач прогнозування попиту, де можуть спостерігатися сезонні коливання або довготривалі тренди. Ця особливість дозволяє LSTM виявляти складні шаблони в історичних даних, які можуть бути важко ідентифіковані іншими архітектурами, такими як RNN або FNN. Зокрема, LSTM має високий рівень надійності, що підтверджує її ефективність для задач прогнозування попиту.

Архітектура CNN, хоча й поступається LSTM за Accuracy, продемонструвала гарні результати у задачах прогнозування попиту. CNN виявилася ефективною для роботи з просторовими даними, що є корисним, коли необхідно виявляти локальні шаблони в історичних даних, зокрема в ситуаціях, коли тренди повторюються, але без значної залежності від довготривалих циклів.

GRU та RNN показали добрі результати за Accuracy, але їхня здатність до обробки складних часових залежностей є менш ефективною в порівнянні з LSTM, що відображається в дещо нижчих значеннях R^2 та Accuracy. FNN також продемонстрував задовільні результати, але його здатність до аналізу складних шаблонів не досягла рівня LSTM та GRU.

АВМ та АЕ, хоча й не досягли таких високих значень показника Accuracy, як LSTM, демонструють здатність до стабільних прогнозів за певних умов. АВМ з високим рівнем надійності є гарним варіантом для складних систем, а АЕ може бути корисним для виявлення аномалій у даних.

Незважаючи на свої переваги, архітектура LSTM має кілька суттєвих обмежень. Одним із основних обмежень є висока обчислювальна складність, що потребує значних обчислювальних ресурсів, особливо при роботі з великими обсягами даних. Це може значно ускладнити використання LSTM в реальних умовах, де обмежені можливості для

обробки даних у режимі реального часу.

Крім того, архітектура LSTM може бути неефективною в нестабільних умовах або в ситуаціях, де дані не мають чіткої циклічності. У таких випадках її здатність до навчання на тривалих часових рядах, залежності яких не змінюються, може бути надмірною, що призводить до неефективної обробки даних.

Ці обмеження потребують додаткової оптимізації та адаптації LSTM для застосування в умовах реальних завдань, де не завжди можливо забезпечити великі обсяги даних або високу обчислювальну потужність.

З огляду на обмеження досліджених архітектур, в майбутньому можна розглядати такі напрямки для подальших досліджень:

- модифікація архітектур нейронних мереж, а саме, оптимізація та модифікація архітектур LSTM і GRU для підвищення їхньої ефективності при обробці великих даних з високою варіативністю;

- використання гібридних моделей, які поєднують переваги різних архітектур, таких як LSTM та CNN, для досягнення кращих результатів у прогнозуванні попиту на продукцію;

- дослідження ефективності нейронних мереж для малих і неповних даних, які можуть призвести до розробки нових методів прогнозування, які б не вимагали великих обсягів інформації.

Отже, результати дослідження підтверджують, що LSTM є найкращим вибором для прогнозування попиту, забезпечуючи високу точність, стабільність та надійність. Інші архітектури, зокрема CNN та GRU, також можуть бути використані, хоча й поступаються LSTM за показниками. ABM та AE можуть бути корисними в специфічних випадках, де необхідно враховувати складні взаємозв'язки або аномалії у даних. Таким чином, хоча поточні результати підтверджують ефективність LSTM для задач прогнозування попиту, подальші дослідження повинні зосередитися на подоланні існуючих обмежень і пошуку нових шляхів для ефективнішого використання нейронних мереж у таких складних задачах.

7. Висновки

Проведене дослідження, що мало на меті визначення архітектури нейронної мережі, яка забезпечує найточніший результат прогнозування попиту на продукцію, дозволило досягти важливих результатів, що сприяють підвищенню достовірності прогнозів та зменшенню ризиків, пов'язаних з надлишковими запасами або дефіцитом продукції.

Однією з основних задач дослідження було визначення переваг та обмежень архітектур нейронних мереж, обраних для вирішення задачі прогнозування попиту. Як показали результати, вибір LSTM як найпродуктивнішої архітектури був обґрунтований здатністю нейронної мережі з такою архітектурою ефективно обробляти складні часові залежності, такі як сезонні коливання, і забезпечувати високі показники точності результатів прогнозування попиту. Однак, незважаючи на її ефективність, LSTM потребує значних обчислювальних ресурсів, що може обмежити її використання в реальних умовах, де необхідна швидка обробка великих обсягів даних. Крім того, вона може бути менш ефективною в нестабільних умовах або коли дані не мають чіткої циклічності.

Під час вирішення задачі формування вибірки історичних даних для навчання та тестування мереж було акцентовано увагу на необхідності точності та якості даних. Це має критичне значення для досягнення високих результатів прогнозування, оскільки будь-які помилки або невідповідності в даних можуть суттєво вплинути на точність роботи нейронної мережі.

Порівняльний аналіз таких архітектур, як LSTM, CNN, GRU, ABM та AE, показав, що

LSTM забезпечує найвищу точність результатів прогнозування попиту на продукцію серед усіх обраних архітектур. Проте альтернативні архітектури, такі як CNN, продемонстрували свою ефективність при роботі з даними, що містять короточасні залежності або просторові структури. Окремо варто виділити АВМ та АЕ, які хоча й не досягли таких високих показників, як LSTM, але можуть виявитися корисними в специфічних випадках, наприклад, для виявлення аномалій або моделювання складних систем, де важливі не лише точність прогнозування, а й стабільність та здатність до ідентифікації аномальних ситуацій.

Таким чином, порівняльний аналіз не лише визначив найрезультативніші архітектури для задачі прогнозування попиту, але й виявив їхні недоліки та обмеження. Результати дослідження показали, що для досягнення ще точніших результатів прогнозування необхідно модифікувати існуючі архітектури або поєднувати їхні сильні сторони в рамках гібридних моделей, що дозволить максимально використовувати переваги кожної з архітектур.

Перелік посилань

1. Parthasarathy N.M.K., Rastogi S., K A. Demand Forecasting in Supply Chain Management using CNN-LSTM Hybrid Model. *Proceedings of the 2023 14th International Conference on Computing Communication and Networking Technologies*, 2023, Delhi, India. doi: 10.1109/ICCCNT56998.2023.10307665
2. Jahin M. A., Shahriar A., Amin M. A. MCDNF: Supply Chain Demand Forecasting via an Explainable Multi-Channel Data Fusion Network Model Integrating CNN, LSTM, and GRU. *IEEE Xplore*. 2024, doi: 10.48550/ARXIV.2405.15598
3. Gassar A.A.A. Short-Term Energy Forecasting to Improve the Estimation of Demand Response Baselines in Residential Neighborhoods: Deep Learning vs. Machine Learning. *Buildings*. 2024. Vol. 14(7). P. 2242. doi: <https://doi.org/10.3390/buildings14072242>
4. Li W., Law K.L.E. Deep Learning Models for Time Series Forecasting: A Review. *IEEE Access*, 2024. Vol. 12. C. 92306-92327. doi: 10.1109/ACCESS.2024.3422528
5. Shiri M. F., Perumal T., Mustapha N., Mohamed R. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. *Cornell University*. 2023. 61 p. doi: <https://doi.org/10.48550/arXiv.2305.17473>
6. Nguyen V., Duong M., Le M. Electricity Demand Forecasting for Smart Grid Based on Deep Learning Approach. *Proceedings of International Conference on Green Technology and Sustainable Development, 2020*. P. 353-357. doi: 10.1109/ICSES.2024.9303164
7. Oukassi H., Hasni M., Layeb S. Long Short-Term Memory Networks for Forecasting Demand in the Case of Automotive Manufacturing Industry. *International Conference on Advanced Systems and Emergent Technologies, 2023*. doi: 10.1109/ICAR.2024.10150543
8. Aguiar-Pérez, J. M., Pérez-Juárez, M. A. An Insight of Deep Learning Based Demand Forecasting in Smart Grids. *Sensors*. 2023. 23(3). P. 1467. doi: 10.3390/s23031467
9. Capone, V., Iannuzzo, G., Camastra, F. Deep Learning for Time Series Forecasting: Advances and Open Problems. *Information*. 2023. 14(11). P. 598. doi: 10.3390/info14110598
10. Oucheikh, R., Benmoussa, O., Mabrouki, C. Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review. *Applied System Innovation*. 2024. 7(5). P. 93. doi: 10.3390/asi7050093
11. Benti N., Chaka M., Semie A. Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects. *Energy Transition for Climate-Inclusive Growth and Sustainable Environments*. 2023. 15(9). P. 7087. doi: <https://doi.org/10.3390/su15097087>

Надійшла до редколегії 25.11.2024 р.

Панфьорова Ірина Юрївна, кандидат технічних наук, доцент, професор кафедри ІУС ХНУРЕ, м. Харків, Україна, e-mail: iryna.panforova@nure.ua. ORCID: <https://orcid.org/0000-0001-7032-9109> (науковий керівник здобувача вищої освіти Ганжили Н.Б.).

Ганжила Нікіта Борисович, здобувач вищої освіти, група ІУСТМ-23-1, факультет комп'ютерних наук ХНУРЕ, м. Харків, Україна, e-mail: nikita.hanzhyla@nure.ua.