

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Методи та підходи оптимізації виробничих процесів
за допомогою нейронних мереж»

на здобуття освітнього ступеня бакалавра
зі спеціальності 124 Системний аналіз

(код, найменування спеціальності)

освітньо-професійної програми Системний аналіз
(назва)

*Кваліфікаційна робота містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело*

(підпис) Валерія ШЕВЧУК
Ім'я, ПРІЗВИЩЕ здобувача

Виконав: здобувачка вищої освіти гр.
САД-41
Валерія ШЕВЧУК
Ім'я, ПРІЗВИЩЕ

Керівник: Михайло КУЗЬМІЧ
доцент кафедри , Ім'я, ПРІЗВИЩЕ
доктор філософії

Рецензент: Вікторія ЖЕБКА
доктор технічних Ім'я, ПРІЗВИЩЕ
наук, професор

Навчально-науковий інститут інформаційних технологій

Освітньо-професійна програма Системний аналіз

«_____» _____ 2025 p.

(прізвище, ім'я, по батькові здобувача)

6. Дата видачі завдання « » _____ 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Збір, аналіз та систематизація наукових джерел з теми.	21.02.2025-08.03.2025	
2	Написання першого (теоретичного) розділу.	10.03.2025-29.03.2025	
3	Аналіз прикладних кейсів оптимізації: графік, енергоспоживання, виявлення дефектів.	30.03.2025-07.04.2025	
4	Написання другого розділу роботи.	10.04.2025-17.04.2025	
5	Аналіз готових моделей класифікації.	17.04.2025-25.04.2025	
6	Написання третього розділу роботи.	25.04.2025-12.05.2025	
7	Написання та оформлення кваліфікаційної роботи.	12.05.2025-20.05.2025	
8	Підготовка доповіді до захисту.	24.05.2025-18.06.2025	

Здобувач(ка) вищої освіти

(підпис)

Валерія ШЕВЧУК

(Ім'я, ПРІЗВИЩЕ)

Керівник
кваліфікаційної роботи

(підпис)

Михайло КУЗЬМІЧ

(Ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня бакалавра: 52 стор., 3 рис., 7 табл., 15 джерел.

Дипломна робота присвячена дослідженню можливостей використання штучних нейронних мереж для оптимізації виробничих процесів. В умовах цифрової трансформації промисловості та необхідності підвищення ефективності, гнучкості й конкурентоспроможності підприємств, особливого значення набувають інтелектуальні системи управління, засновані на технологіях штучного інтелекту. Серед них нейронні мережі вирізняються здатністю до самонавчання, адаптації до змін навколишнього середовища, а також високою точністю у вирішенні складних багатofакторних задач.

Метою роботи є аналіз теоретичних основ нейронних мереж та їх архітектур, дослідження сучасних підходів до оптимізації виробничих процесів із використанням моделей глибокого навчання, а також розробка та апробація прикладної моделі на основі сучасних інструментів машинного навчання.

У першому розділі розкрито поняття, класифікацію та архітектуру штучних нейронних мереж, а також проаналізовано основні принципи їх роботи, методи навчання (з учителем, без учителя та з підкріпленням), функції активації та особливості реалізації згорткових і рекурентних моделей. Розглянуто застосування нейромереж у виробничій сфері, зокрема для задач контролю якості, прогнозного технічного обслуговування, оптимізації технологічних режимів і управління ресурсами.

У другому розділі висвітлено методи оптимізації виробничих процесів із використанням нейронних мереж. Проведено аналіз таких задач, як оптимізація виробничого графіка, зниження енергоспоживання, автоматизоване виявлення дефектів продукції та порівняння ефективності різних методів оптимізації.

Особливу увагу приділено практичним аспектам впровадження моделей штучного інтелекту в реальне виробниче середовище.

Третій розділ містить прикладну частину дослідження, в якій реалізовано і протестовано модель класифікації об'єктів із використанням фреймворку TensorFlow та бібліотеки Object Detection API. Проведено підготовку даних, навчання моделі, її валідацію та оцінку результатів. На основі експерименту зроблено висновки про ефективність запропонованого підходу та можливості його подальшої адаптації до потреб конкретного підприємства.

Робота має як теоретичну, так і практичну цінність, оскільки дозволяє оцінити перспективи застосування нейромережевих технологій у сфері виробництва, забезпечити підвищення якості продукції, зниження витрат та розвиток адаптивних інтелектуальних систем управління.

Ключові слова: штучні нейронні мережі, глибоке навчання, оптимізація виробництва, комп'ютерний зір, машинне навчання, TensorFlow, автоматизація, цифрове виробництво.

ЗМІСТ

ВСТУП.....	9
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ НЕЙРОННИХ МЕРЕЖ ТА ЇХ ЗАСТОСУВАННЯ У ВИРОБНИЧИХ ПРОЦЕСАХ.....	12
1.1. Поняття та класифікація нейронних мереж.....	12
1.2. Одношарові та багатошарові штучні нейронні мережі.....	14
1.3. Основні архітектури та принципи роботи нейромереж.....	17
1.4. Методи навчання нейронних мереж.....	19
1.5. Класифікатори на основі згорткової нейронної мережі.....	24
1.6. Використання нейронних мереж у виробництві.....	26
РОЗДІЛ 2. МЕТОДИ ТА ПІДХОДИ ДО ОПТИМІЗАЦІЇ ВИРОБНИЧИХ ПРОЦЕСІВ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ	31
2.1. Оптимізація виробничого графіка.....	31
2.2. Оптимізація енергоспоживання на підприємстві.....	33
2.3. Автоматизоване виявлення дефектів продукції.....	36
2.4. Порівняння ефективності різних методів оптимізації.....	41
РОЗДІЛ 3. ДОСЛІДЖЕННЯ НЕЙРОННИХ МЕРЕЖ.....	45
3.1 Аналіз прикладного використання базової моделі класифікації.....	45
3.2 Дослідження з TensorFlow Object Detection API.....	49
Висновки.....	56
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	59

ВСТУП

У сучасних умовах динамічного розвитку економіки, цифровізації та глобалізації ринків одним із ключових чинників забезпечення конкурентоспроможності підприємства є ефективна організація виробничих процесів. Оптимізація виробництва дозволяє зменшити витрати, підвищити якість продукції, скоротити терміни виконання замовлень і раціонально використовувати ресурси. Проте зі зростанням складності виробничих систем, варіативністю параметрів і необхідністю адаптації до змін зовнішнього середовища класичні методи оптимізації втрачають ефективність. У цьому контексті актуальності набуває впровадження інтелектуальних систем аналізу та управління, зокрема на основі технологій штучного інтелекту, серед яких провідне місце посідають нейронні мережі.

Нейронні мережі, як одна з ключових технологій сучасного машинного навчання, демонструють високі результати у вирішенні складних задач, пов'язаних із прогнозуванням, класифікацією, оптимізацією, розпізнаванням закономірностей та прийняттям рішень в умовах неповної або нечіткої інформації. На відміну від традиційних аналітичних методів, нейромережеві моделі здатні самонавчатись, узагальнювати досвід і ефективно працювати в реальному часі. Це робить їх особливо цінними у сфері виробництва, де часто необхідно оперативно реагувати на зміни, враховувати велику кількість параметрів і здійснювати складні багатofакторні розрахунки.

Застосування нейронних мереж у виробничій сфері охоплює широкий спектр задач: прогнозування попиту, оптимізація виробничих графіків, керування потоками ресурсів, контроль якості продукції, виявлення аномалій у технологічних процесах, управління технічним обслуговуванням обладнання та інші. Впровадження таких рішень дозволяє підприємствам автоматизувати процеси прийняття рішень, підвищити гнучкість та адаптивність виробничих систем, а також зменшити залежність від людського фактору.

Попри значний науковий і практичний інтерес до теми, процес інтеграції нейронних мереж у виробничі системи потребує подальшого дослідження. Це

пов'язано з низкою викликів, серед яких — складність побудови якісних навчальних вибірок, необхідність обґрунтованого вибору архітектури нейромережі, інтерпретованість моделей і забезпечення стабільності результатів у реальних умовах. У цьому контексті важливим є формування ефективних методів і підходів до розробки та впровадження нейромережових рішень з урахуванням специфіки конкретних виробничих задач.

Таким чином, дослідження методів і підходів оптимізації виробничих процесів із використанням нейронних мереж є актуальним як у науковому, так і в прикладному аспекті. Результати такого дослідження можуть стати основою для розробки інтелектуальних систем управління, що забезпечують якісно новий рівень ефективності сучасних виробництв.

Мета дослідження

Метою роботи є вивчення, аналіз підходів до оптимізації виробничих процесів за допомогою нейронних мереж, з метою підвищення ефективності, гнучкості й адаптивності виробничих систем.

Об'єкт дослідження

Об'єктом дослідження є виробничі процеси підприємств, що характеризуються багатофакторністю, складною структурою і потребою в оптимізації для підвищення ефективності функціонування.

Предмет дослідження

Предметом дослідження є методи і підходи застосування штучних нейронних мереж для оптимізації виробничих процесів, включаючи моделювання, прогнозування та автоматизацію прийняття рішень.

Структура роботи

Робота складається з вступу, трьох розділів, висновків, списку використаних джерел.

У першому розділі розглядається теоретичний аспект оптимізації виробничих процесів та основи нейронних мереж.

У другому розділі проводиться аналіз існуючих підходів та рішень, що використовуються в промисловості.

У третьому розділі описується практична реалізація оптимізаційного підходу.

У завершальній частині подаються висновки та перспективи подальшого розвитку теми.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ НЕЙРОННИХ МЕРЕЖ ТА ЇХ ЗАСТОСУВАННЯ У ВИРОБНИЧИХ ПРОЦЕСАХ

1.1. Поняття та структура нейронних мереж

Штучні нейронні мережі (ШНМ) – це обчислювальні моделі, що імітують принципи роботи біологічних нейронних мереж у людському мозку. Вони складаються з великої кількості взаємопов'язаних штучних нейронів, які обробляють інформацію через вагові коефіцієнти та активаційні функції.

Основним будівельним блоком нейромережі є штучний нейрон – математичний об'єкт, який отримує на вхід кілька значень, обробляє їх за допомогою вагових коефіцієнтів і передає результат далі за мережею.

Типова структура нейронної мережі складається з трьох основних типів шарів: вхідного, прихованих та вихідного. Вхідний шар виконує функцію прийому даних ззовні та передає їх далі вглиб моделі. Кількість нейронів у вхідному шарі відповідає кількості ознак (фіч) у вхідному наборі даних. Наприклад, при обробці зображення 28×28 пікселів вхідний шар містить 784 нейрони. Сам вхідний шар не виконує обчислень, він лише передає значення до наступного шару.

Приховані шари – це основна обчислювальна частина мережі. Вони розташовані між вхідним і вихідним шарами та виконують обробку, фільтрацію і трансформацію інформації. Кожен нейрон прихованого шару отримує значення з попереднього шару, зважує їх, додає зміщення (bias) і передає результат через активаційну функцію далі. Типи прихованих шарів можуть відрізнятися залежно від задачі. Повнозв'язні (fully connected, або dense) шари з'єднують кожен нейрон із кожним у попередньому та наступному шарі і застосовуються в задачах класифікації або регресії. Згорткові (convolutional) шари застосовуються до просторових даних, таких як зображення, і дозволяють виділяти характерні ознаки завдяки операції згортки. Рекурентні шари (recurrent), зокрема LSTM або GRU, використовуються в задачах, де важлива послідовність даних, наприклад, у

часових рядах, тексті чи аудіо. Ці шари мають пам'ять про попередні значення і дозволяють враховувати контекст при прогнозуванні або аналізі.

Вихідний шар формує фінальний результат роботи мережі. Його структура залежить від типу задачі. У випадку класифікації кількість нейронів відповідає кількості класів. Наприклад, для класифікації деталей на три категорії вихідний шар міститиме три нейрони. У випадку регресії, як-от прогноз часу виготовлення продукту, зазвичай використовується один нейрон. Тип активаційної функції у вихідному шарі залежить від задачі: для бінарної класифікації використовується Sigmoid, для багатокласової класифікації – Softmax, а для регресії – лінійна функція або без активації.

Ключовою складовою нейронної мережі є активаційна функція, яка визначає вихід нейрона залежно від його зваженої суми вхідних сигналів. Основні функції включають ReLU (Rectified Linear Unit), яка повертає нуль для від'ємних значень і саме значення для додатних; Sigmoid – дає результат у межах від 0 до 1, корисна для бінарної класифікації; Tanh – подібна до Sigmoid, але має вихід від -1 до 1; Softmax – використовується для багатокласової класифікації, нормалізуючи вихід у вигляді ймовірностей. Вибір структури нейронної мережі, кількості шарів, кількості нейронів у кожному шарі та активаційних функцій визначає її здатність до навчання, узагальнення даних і точного прогнозування. Така структура особливо ефективна у виробничій сфері, де необхідно аналізувати великі обсяги даних, оптимізувати графіки виконання робіт або прогнозувати навантаження на ресурси.

Таблиця 1.1

Функція	Формула/Дія	Область застосування
ReLU (Rectified Linear Unit)	$f(x) = \max(0, x)$	Найбільш поширена для прихованих шарів, швидко навчається, проста у реалізації

Sigmoid	$f(x) = \frac{1}{(1 + e^{-x})}$	Використовується для бінарної класифікації, видає значення від 0 до 1
Tanh (гіперболічний тангенс)	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	Схожа на Sigmoid, але симетрична відносно нуля: значення від -1 до 1
Softmax	$f(x, i) = \frac{e^{(x_i)}}{\sum e^{x_j}}$	Нормалізує вектор виходів у ймовірності (для багатокласової класифікації)

Продовження таблиці 1.1

1.2 Одношарові та багатошарові штучні нейронні мережі

Незважаючи на те, що один окремих штучний нейрон здатен виконувати елементарні функції розпізнавання, ефективність його роботи значно зростає при об'єднанні з іншими нейронами у вигляді мережі. Одношаровою штучною нейронною мережею (ОДШНМ) називають архітектуру, в якій обчислення здійснюється лише на одному обчислювальному шарі, що містить кілька паралельно працюючих нейронів.

На рисунку 1.1 зображено загальну структуру такої мережі. Вхідні сигнали подаються на умовні вузли, які не виконують обчислень — вони позначені колами для візуального відокремлення від активних обчислювальних елементів (нейронів), які представлені квадратами.

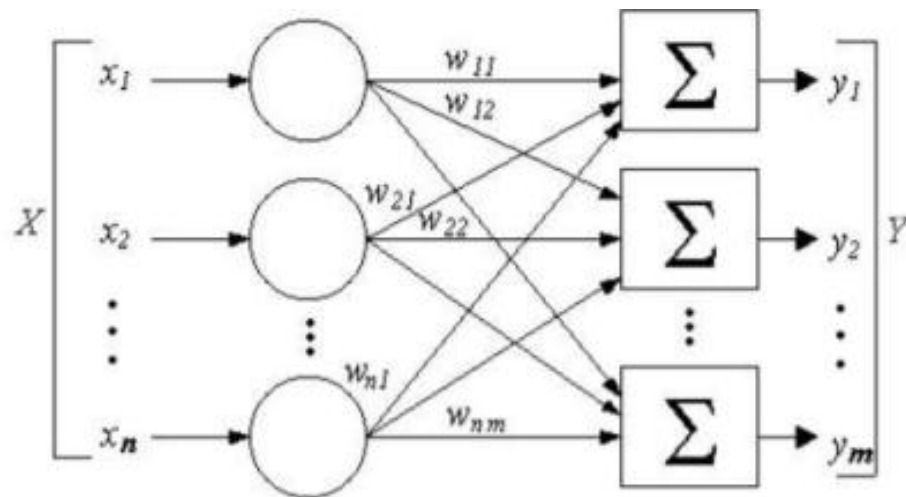


Рисунок 1.1

Кожен вхід із множини $X = \{x_1, x_2, \dots, x_n\}$ з'єднаний окремим ваговим коефіцієнтом із кожним нейроном шару. Таким чином формується повнозв'язна структура, де кожен нейрон приймає на вхід лінійну комбінацію всіх вхідних сигналів, зважених відповідними коефіцієнтами.

В результаті мережа генерує вектор вихідних значень $Y = \{y_1, y_2, \dots, y_m\}$ який може бути передано на наступний рівень обробки або безпосередньо використано для прийняття рішень.

Ця архітектура є найпростішим прикладом нейронної мережі, на основі якої будуються складніші багатошарові моделі. Водночас, вона дозволяє реалізувати базові операції класифікації, регресії та апроксимації у задачах, де немає потреби в глибокому розпізнаванні складних шаблонів.

Багатошарові нейронні мережі (БШНМ) значно перевищують одношарові за обчислювальними можливостями, оскільки здатні моделювати складні, нелінійні залежності між вхідними та вихідними даними.

На сучасному етапі розвитку штучного інтелекту багатошарові архітектури поділяються на дві великі групи:

- статичні (мережі з фіксованою структурою зв'язків і без залежності від часу);
- динамічні (мережі, в яких стан залежить від попередніх кроків або часу).

До статичних архітектур належать моделі прямого поширення сигналу (feedforward networks), у яких інформація передається строго з вхідного шару

через один або кілька прихованих шарів до вихідного без зворотних зв'язків. У таких мережах вихід залежить винятково від поточного вхідного вектора, а не від історії попередніх входів.

Прикладами статичних архітектур є:

- Багатошаровий персептрон (MLP) — класична модель з повнозв'язними шарами та функцією активації;
- Нейронна мережа Кохонена (SOM) — модель без вчителя, яка здійснює кластеризацію та зниження розмірності;
- Когнітрон і неокогнітрон — попередники згорткових мереж, орієнтовані на ієрархічне виділення ознак;
- Сучасні згорткові нейронні мережі (CNN) — архітектури, які показали високу ефективність у задачах комп'ютерного зору.

Усі ці мережі мають спільну рису — відсутність механізмів пам'яті, що робить їх оптимальними для задач, де кожне передбачення не залежить від попередніх (наприклад, класифікація зображень, аналіз сигналів у фіксовані моменти часу).

Динамічні багатошарові нейронні мережі

На противагу статичним, динамічні архітектури (рекурентні нейронні мережі, RNN) мають зворотні зв'язки, завдяки яким вихід моделі залежить не лише від поточного входу, а й від попередніх станів нейромережі. Це дозволяє реалізовувати пам'ять про попередні події та аналізувати послідовності даних — наприклад, часові ряди, текст, аудіо або відео.

Рекурентні мережі часто базуються на багатошаровому персептроні, який доповнюється рекурентними з'єднаннями. До найпоширеніших представників цієї групи належать:

- Класичні RNN (Recurrent Neural Networks);
- LSTM (Long Short-Term Memory) — модифіковані RNN з механізмом збереження довготривалих залежностей;
- GRU (Gated Recurrent Unit) — спрощений варіант LSTM, що зберігає переваги пам'яті при меншій обчислювальній складності.

Рекурентні мережі є незамінними в задачах прогнозування, розпізнавання мовлення, обробки природної мови, а також у випадках, коли результат залежить від контексту або історії подій.

Таким чином, багатошарові штучні нейронні мережі реалізують значно вищу гнучкість і функціональність порівняно з одношаровими, і їх вибір (статичний чи динамічний тип) залежить від типу задачі: статичні моделі — для незалежних входів, динамічні — для послідовнісного аналізу та моделей, що враховують контекст.

1.3. Основні архітектури та принципи роботи нейромереж

Штучні нейронні мережі (ШНМ) є одним із ключових інструментів сучасного інтелектуального аналізу даних та машинного навчання, що дозволяє вирішувати широкий спектр задач — від розпізнавання образів до прогнозування складних процесів. В основі функціонування будь-якої нейронної мережі лежить здатність моделі адаптивно навчатися, виявляти закономірності в даних та формувати нелінійні відображення між вхідною і вихідною інформацією. Розуміння архітектур ШНМ є критично важливим для практичного застосування цих систем у виробничій сфері, зокрема для оптимізації технологічних процесів, логістики, управління ресурсами та прийняття рішень.

Архітектура нейромережі визначає її внутрішню будову: кількість і тип шарів, спосіб з'єднання нейронів, типи передавання сигналів, а також механізми навчання. Умовно архітектури ШНМ поділяють на кілька базових типів, кожен з яких має своє призначення, переваги та обмеження.

1. Прямого розповсюдження (Feedforward Neural Networks, FFNN) – це найпростіша і найпоширеніша архітектура нейромереж. Вона передбачає одностороннє проходження сигналів від вхідного шару до вихідного без зворотного зв'язку. Кожен нейрон наступного шару отримує сигнал від усіх нейронів попереднього шару. Навчання такої мережі зазвичай здійснюється методом зворотного поширення помилки (Backpropagation) із використанням

градієнтного спуску. Feedforward-мережі добре підходять для класифікації, регресії, контролю якості та інших задач, де не враховується часова або послідовна залежність.

2. Згорткові нейронні мережі (Convolutional Neural Networks, CNN) – це архітектура, спеціально створена для обробки структурованих у просторі даних, таких як зображення або тривимірні моделі. Основною її особливістю є використання згорткових шарів, які виконують операції фільтрації (згортки) над локальними ділянками вхідного простору, що дозволяє ефективно виділяти ознаки (features). CNN надзвичайно потужні у завданнях розпізнавання образів, візуального контролю якості продукції, моніторингу стану обладнання за відеоданими тощо. Крім згорткових, така мережа також включає шари підвибірки (subsampling або pooling) і повнозв'язні шари для прийняття остаточного рішення.

3. Рекурентні нейронні мережі (Recurrent Neural Networks, RNN) – архітектура, яка враховує часову складову вхідних даних. Вона має зворотні зв'язки, що дозволяють зберігати інформацію про попередні входи у внутрішньому стані мережі. Це робить RNN незамінними для аналізу часових рядів, прогнозування параметрів виробництва на основі історичних даних, обробки сигналів від датчиків, управління виробничими циклами тощо. Проте традиційні RNN страждають від проблеми затухаючих або вибухаючих градієнтів, тому в реальних задачах найчастіше використовуються їхні вдосконалені варіанти: LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Units). Ці архітектури забезпечують ефективне довготривале збереження інформації, що критично важливо для задач, де вплив попередніх подій зберігається протягом тривалого часу.

4. Мережі зі зворотним зв'язком (Feedback Neural Networks) – ще один тип архітектури, де присутні петлі зворотного зв'язку. Вони формують динамічні системи, стан яких змінюється у відповідь на нові вхідні дані. Такі мережі здатні моделювати складні адаптивні системи та можуть застосовуватись у задачах управління, оптимізації систем реального часу та моделювання фізичних процесів.

5. Самоорганізовувані карти (Self-Organizing Maps, SOM) – це неklasичний тип нейромереж, де навчання здійснюється без вчителя. SOM проєктують багатовимірні дані на площину, зберігаючи їхню топологічну структуру, що дозволяє візуалізувати дані, виявляти кластери або аномалії. У виробництві SOM можуть бути використані для класифікації типових режимів роботи або виявлення нетипових ситуацій.

Загалом, принцип роботи нейромережі полягає в ітеративному проходженні інформації через усі шари моделі з подальшим оновленням ваг з'єднань між нейронами. Навчання ШНМ базується на порівнянні реального виходу з очікуваним (еталонним) і мінімізації функції втрат, що визначає помилку прогнозу. На цьому етапі використовується зворотне поширення помилки та оптимізаційні алгоритми (найчастіше стохастичний градієнтний спуск або його варіації: Adam, RMSprop тощо). Після завершення процесу навчання модель здатна робити узагальнення і видавати точні прогнози або рішення при поданні нових даних.

Таким чином, вибір архітектури нейромережі є визначальним етапом у побудові ефективної інтелектуальної системи. Для задач оптимізації виробничих процесів, де важливо враховувати множинні параметри, взаємозв'язки між етапами виробництва, часові затримки та ресурсні обмеження, доцільно використовувати гібридні підходи, які поєднують елементи згорткових, рекурентних та повнозв'язних мереж. Це дозволяє побудувати адаптивну систему прийняття рішень, що здатна динамічно реагувати на зміни у виробничому середовищі, прогнозувати навантаження на обладнання, оптимізувати графіки та мінімізувати витрати.

1.4 Методи навчання нейронних мереж

Методи навчання нейронних мереж становлять основу функціонування інтелектуальних систем і визначають здатність моделі адаптуватися до змінного середовища, узагальнювати інформацію, робити прогнози та приймати рішення. Навчання мережі полягає у налаштуванні її параметрів (ваг і зміщень) таким

чином, щоб вихідні сигнали відповідали заданим цілям або очікуванням, що задаються через тренувальні приклади. Існує кілька основних підходів до навчання нейронних мереж, які класифікуються залежно від характеру наявних даних і принципів взаємодії мережі з навчальним середовищем.

Першим і найбільш поширеним є метод навчання з учителем (supervised learning), при якому мережі надається набір вхідних даних із відповідними вихідними мітками. Основна мета полягає в тому, щоб знайти таку конфігурацію вагових коефіцієнтів, яка мінімізує функцію втрат (наприклад, середньоквадратичну або крос-ентропію) між фактичними виходами мережі та цільовими значеннями. Базовим алгоритмом, що використовується у цьому контексті, є зворотне розповсюдження помилки (backpropagation), що реалізує метод градієнтного спуску для багат шарових мереж. У процесі зворотного проходження обчислюється градієнт функції помилки відносно кожного з параметрів мережі, після чого ці параметри оновлюються з урахуванням кроку навчання (learning rate). З часом були розроблені вдосконалені варіанти цього методу, такі як стохастичний градієнтний спуск (SGD), а також адаптивні оптимізатори – Adam, RMSProp, AdaGrad, які покращують збіжність та стабільність навчання, особливо в умовах великої кількості параметрів. Середньоквадратична похибка (для регресії):

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1.1)$$

Крос-ентропійна похибка (для класифікації):

$$\mathcal{L}_{\text{CE}} = - \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (1.2)$$

Формула оновлення ваг у градієнтному спуску:

$$w^{(t+1)} = w^t - \eta \cdot \frac{\partial \mathcal{L}}{\partial w} \quad (1.3)$$

Другим важливим напрямом є навчання без учителя (unsupervised learning), що застосовується у випадках, коли вхідні дані не мають відповідних вихідних значень. Мета такого навчання – виявити приховану структуру в даних, знайти закономірності, групи або головні ознаки. До найбільш популярних архітектур цього типу належать автокодери (autoencoders), які складаються з двох частин: енкодера, що стискає вхідні дані у вектор фіксованої довжини, і декодера, який намагається відновити оригінальне представлення. Під час навчання мережа мінімізує похибку реконструкції, що дозволяє виділити суттєві ознаки. Також використовуються самоорганізовані карти Кохонена (SOM), які дозволяють зменшити розмірність даних і кластеризувати їх у просторах меншої розмірності. У більш традиційному підході іноді нейронні мережі комбінуються з класичними алгоритмами кластеризації (наприклад, K-means) для аналізу багатовимірних структур даних.

Енкодер:

$$z = f_{\text{enc}}(x) \quad (1.4)$$

Декодер:

$$\hat{x} = f_{\text{dec}}(z) \quad (1.5)$$

Функція втрат- похибка реконструкції:

$$\mathcal{L}_{\text{AE}} = \|x - \hat{x}\|^2 = \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (1.6)$$

Третім фундаментальним методом є навчання з підкріпленням (reinforcement learning) – це підхід, за якого нейронна мережа діє як агент, що взаємодіє із середовищем, отримуючи зворотний зв'язок у вигляді винагороди або штрафу залежно від обраних дій. На відміну від supervised learning, у навчанні з підкріпленням не існує наперед заданої правильної відповіді – мережа самостійно формує політику поведінки, яка максимізує накопичену винагороду у довгостроковій перспективі. Такий підхід є надзвичайно ефективним у задачах управління, стратегічного планування, ігор (зокрема, відомий приклад AlphaGo), а також у складних сценаріях з динамічними обмеженнями. Для реалізації цього типу навчання використовуються алгоритми на кшталт Q-learning, Deep Q Networks (DQN), Policy Gradient, а також Actor-Critic архітектури, які комбінують оцінку значень станів і прямих стратегій.

Окрім трьох основних підходів, існують важливі методології та техніки, що дозволяють покращити якість та стійкість процесу навчання. Наприклад, регуляризація – це сукупність методів, що запобігають перенавчанню моделі на тренувальних даних. До таких методів належать L1/L2 регуляризація, Dropout, при якому випадково "вимикаються" нейрони під час навчання, що сприяє кращій узагальнюваності, а також рання зупинка (early stopping), яка передбачає припинення навчання при досягненні мінімуму похибки на валідаційній вибірці. Значну роль відіграє нормалізація вхідних даних, зокрема стандартизація або масштабування, що дозволяє уникнути домінування певних ознак і стабілізує градієнти. У випадках, коли глибокі моделі важко навчити "з нуля", застосовується підхід переднавчання (pretraining) – спочатку мережа тренується на великій кількості універсальних даних, після чого проводиться трансферне навчання (transfer learning) на конкретному наборі задач.

Також варто згадати про батчеве (batch) навчання, коли дані поділяються на міні-пакети (batches), що дозволяє поєднати переваги стохастичної та повної оптимізації. Маленькі батчі забезпечують швидшу ітерацію, а більші – стабільність градієнтів. Крім того, в сучасних дослідженнях активно вивчаються гібридні методи, такі як самонавчання (self-supervised learning) – проміжна форма між навчанням з учителем і без нього, що дозволяє моделі створювати власні мітки на основі структури даних. Це особливо важливо у випадках, коли розмічені набори даних дорогі або обмежені.

Таким чином, методи навчання нейронних мереж охоплюють широкий спектр підходів, що дозволяють ефективно працювати як із класичними задачами класифікації, регресії, так і з більш складними сценаріями оптимізації поведінки та аналізу неструктурованих даних. Вибір конкретного методу залежить від характеру задачі, типу доступних даних, складності архітектури та цілей моделювання. Сучасні досягнення в галузі глибокого навчання поєднують ці методи, створюючи багаторівневі системи з високою точністю, гнучкістю та здатністю до самонавчання.

Методи навчання нейронних мереж мають свої переваги та недоліки, що визначає їх доцільність у конкретних завданнях:

Таблиця 1.2

Метод	Переваги	Недоліки
Навчання з учителем	Висока точність при наявності розмічених даних	Залежність від якості та кількості даних
Навчання без учителя	Виявлення прихованих закономірностей	Складність інтерпретації результатів
Навчання з підкріпленням	Застосування в складних динамічних системах	Труднощі в налаштуванні параметрів та виборі винагород
Адаптивні методи	Швидка та стабільна	Підбір гіперпараметрів може

	оптимізація	бути складним
--	-------------	---------------

1.5 Класифікатори на основі згорткової нейронної мережі

Згорткові нейронні мережі (Convolutional Neural Networks, CNN) є класом глибоких нейронних архітектур, що ефективно обробляють дані зі структурованою топологією, зокрема зображення, відео або просторові сигнали. Як і в класичних багатошарових персептронах, у CNN класифікаційні шари — зазвичай реалізовані у вигляді повнозв'язних (fully connected) — мають зв'язки між усіма нейронами поточного шару та виходами попереднього. Обчислення активацій у таких шарах відбувається за допомогою матричних операцій (добуток ваг і вхідних значень) із додаванням зсуву (bias), що формально описується виразом:

$$a = f(W_x + b) \quad (1.7)$$

де:

- W - матриця вагових коефіцієнтів,
- x - вхідний вектор активацій з попереднього шару,
- b - вектор зсувів,
- f - функція активації.

Ключова відмінність між згортковими та повнозв'язними шарами полягає в локальності з'єднань: нейрони згорткового шару взаємодіють лише з обмеженою локальною областю вхідного простору (receptive field), що дозволяє ефективно виявляти просторові патерни. Окрім того, згорткові шари мають спільні ваги для кожного фільтра, що істотно знижує кількість параметрів і дозволяє моделі узагальнювати ознаки незалежно від їх розташування на зображенні.

Попри конструктивні відмінності, обидва типи шарів (згорткові й повнозв'язні) реалізують обчислення скалярного добутку між вхідними значеннями та ваговими

коефіцієнтами. Таким чином, функціональна форма їх роботи є ідентичною. Більше того, за необхідності можна здійснити перетворення згорткового шару у повнозв'язний, наприклад, під час переходу до вихідного шару класифікації. На етапі класифікації, нейромережа має сформувати розподіл ймовірностей належності об'єкта до одного з класів, що реалізується шляхом лінійного перетворення та застосування функції softmax.

У практичних застосуваннях (наприклад, класифікація зображень), лінійний класифікатор (рис. 1.4) оцінює клас зображення як зважену суму всіх піксельних значень (можливо, у кількох каналах), і залежно від напрямку й значення ваг кожен піксель по-своєму впливає на результат класифікації. Наприклад, класифікатор може вважати, що наявність великої кількості синіх пікселів з боків зображення підвищує ймовірність того, що на зображенні зображено корабель, оскільки синій колір інтерпретується як вода.

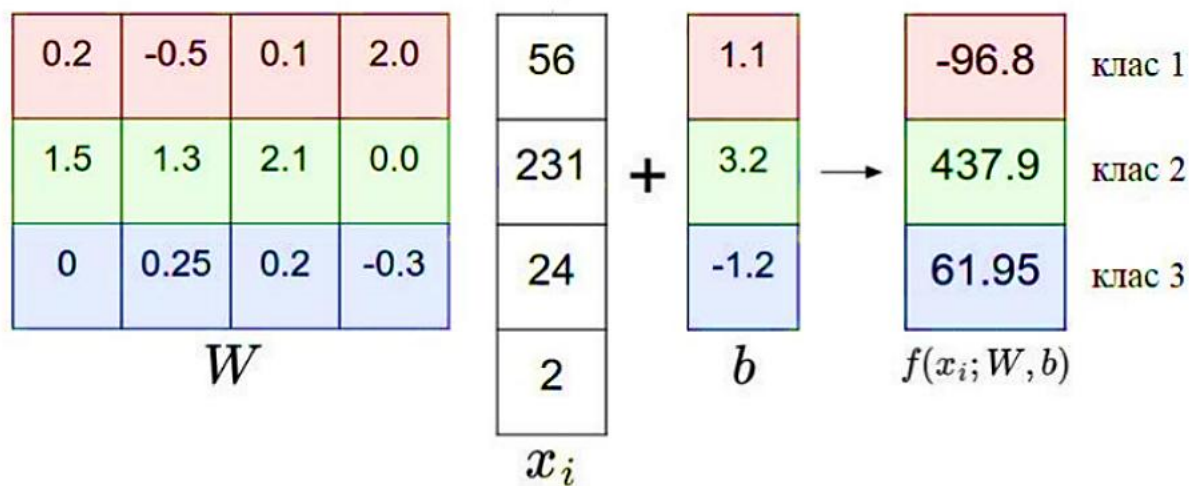


Рисунок 1.2 – Лінійний класифікатор

Шаблонне представлення ваг.

Іншим способом побудови класифікатора є метод шаблонної відповідності (template matching). У цьому випадку кожен рядок матриці ваг W інтерпретується як шаблон (прототип) відповідного класу. Кожне зображення порівнюється із цими шаблонами за допомогою скалярного добутку:

$$s_k = x^T w_k \quad (1.8)$$

де:

- s_k - оцінка відповідності класу k ,
- x - вектор ознак зображення,
- w_k - вектор-шаблон класу k .

Клас із найбільшою відповідністю (найбільшим значенням $s_{k_{sk}}$) вважається передбаченим. Такий підхід дозволяє інтерпретувати навчання мережі як процес побудови системи прототипів, що добре "впізнають" характерні риси певного класу.

Особливості реалізації.

Важливо зазначити, що на відміну від "запам'ятовуючих" моделей (наприклад, методів kkk-ближчих сусідів), класифікатор у CNN не зберігає усі приклади навчального набору, а формує узагальнені представлення через оптимізацію параметрів. Це дозволяє зменшити вимоги до обсягу пам'яті та підвищити швидкість передбачень на етапі тестування.

Класифікаційні шари у згорткових нейронних мережах забезпечують ефективне перетворення ознак, отриманих у процесі згортки, у конкретне рішення щодо належності об'єкта до того чи іншого класу. Незалежно від способу реалізації — чи то як повнозв'язний шар, чи як система шаблонів — такі модулі формують фінальний етап прийняття рішень у задачах комп'ютерного зору та обробки зображень.

1.6 Використання нейронних мереж у виробництві

В умовах цифрової трансформації промисловості застосування нейронних мереж у виробничому середовищі стало одним із ключових факторів підвищення ефективності, гнучкості та конкурентоспроможності підприємств. Нейронні

мережі, особливо в контексті глибокого навчання (deep learning), стали потужним інструментом аналізу даних, оптимізації процесів, автоматизації управління та передбачення подій, які раніше неможливо було моделювати традиційними методами.

Суть використання нейронних мереж у виробництві полягає в тому, щоб шляхом аналізу великих обсягів вхідних даних (датчики, IoT, відеоаналіз, ERP-системи) навчити модель виявляти закономірності, робити прогнози, реагувати на відхилення від норми, приймати рішення або рекомендувати дії оператору. Таким чином, нейромережеві технології стають ядром інтелектуальних систем підтримки рішень, цифрових двійників і автономних кібер-фізичних систем.

1. Контроль якості продукції

Одне з наймасовіших і найрезультативніших впроваджень нейронних мереж стосується автоматизованого виявлення дефектів. Раніше ці процеси покладалися переважно на людський фактор, що зумовлювало суб'єктивність оцінки, втому, низьку масштабованість.

Сьогодні згорткові нейронні мережі (Convolutional Neural Networks, CNN) здатні:

- виявляти мікротріщини;
- аналізувати рівномірність фарбування;
- виявляти порушення геометрії деталей;
- виявляти вклучення, плями, деформації та ін.

Формально, модель функціонально виконує задачу класифікації або сегментації. Наприклад, для задачі бінарної класифікації (дефект/норма), використовують функцію втрат:

$$\mathcal{L}_{\text{binary}} = - [y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})] \quad (1.9)$$

де $\hat{y}$ - прогноз моделі, y - правильна мітка.

2. Прогнозування технічного стану обладнання (Predictive Maintenance)

Прогнозне обслуговування — це концепція, що базується на аналізі часових рядів із датчиків (вібрації, шум, температура, навантаження). Нейронні мережі типу LSTM (Long Short-Term Memory) або GRU використовуються для моделювання послідовностей та виявлення відхилень від нормального функціонування.

Ціль — передбачити ймовірність виходу з ладу або зниження продуктивності, щоб завчасно провести техобслуговування.

Формально модель прогнозує:

$$P_{\text{fail}}(t + \Delta t) = f(x_{t-n}, \dots, x_t) \quad (1.10)$$

де x_t - сенсорні дані, f - навчена нейромережа.

Такі системи значно знижують втрати, пов'язані з аварійними зупинками, подовжують строк служби обладнання та зменшують витрати на ремонт.

3. Оптимізація технологічних процесів

Нейронні мережі дозволяють створювати цифрові моделі складних фізико-хімічних і механічних процесів, які важко або неможливо описати аналітично. У реальному виробництві часто присутні складні, нелінійні, стохастичні взаємозв'язки між десятками параметрів, які впливають на якість та собівартість продукції.

Застосування:

- оптимізація температурного режиму в ливарному виробництві;
- коригування складу суміші в реальному часі;
- управління дозуванням реагентів у хімічному процесі.

Алгоритм навчається знаходити функцію $f(x)=y^{\wedge}f(\mathbf{x})=\hat{y}$, яка мінімізує відхилення між бажаним результатом і фактичним. У виробництві це часто регресійна модель із функцією втрат:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (1.11)$$

де y_i - фактичне значення показника якості, $\hat{y}_i$ - прогноз моделі.

4. Управління виробничими ресурсами та планування

Прогнозування попиту, планування завантаження цехів, розрахунок логістики постачання сировини — всі ці процеси можна оптимізувати із застосуванням нейронних мереж.

Використовуються моделі:

- Feedforward Neural Networks — для прогнозів без часової залежності;
- Transformer — для довготривалих часових залежностей;
- Reinforcement Learning — для оптимізації політики виробництва в умовах змінних параметрів середовища.

5. Роботизація та автономні системи

Сучасні виробництва широко використовують інтелектуальних роботів, які керуються нейромережами:

- об'єктне розпізнавання (візуальні модулі на основі CNN);
- контроль рухів (inverse kinematics + neural control policies);
- адаптація до нестандартних ситуацій (reinforcement learning).

Використовується модель Actor-Critic:

- Actor генерує дію a ;
- Critic оцінює її в контексті цільової функції винагороди RRR.

6. Цифрові двійники (Digital Twins)

Нейронні мережі часто є основним ядром цифрових двійників — динамічних симуляційних моделей реального виробничого об'єкта. Вони дозволяють у режимі реального часу:

- відслідковувати технічний стан;
- симулювати наслідки змін;

- передбачати аварійні стани;
- тестувати нові виробничі сценарії без зупинки процесу.

7. Кібербезпека виробничих систем

Завдяки можливості виявлення аномалій та вторгнень у великих потоках даних, нейронні мережі дозволяють ефективно:

- виявляти саботаж у ПЛК;
- відстежувати нетипову поведінку сенсорів;
- виявляти зловмисне втручання в SCADA-системи.

8. Переваги і виклики

Переваги:

- здатність до навчання зі складних, шумних, багатовимірних даних;
- гнучкість адаптації до нових умов;
- автоматизація аналізу даних у реальному часі.

Виклики:

- потреба у великих обсягах якісних даних для навчання;
- інтерпретованість результатів (нейромережі — «чорні скриньки»);
- складність у налаштуванні архітектур і гіперпараметрів;
- потреба в високих обчислювальних потужностях.

Нейронні мережі стали ключовим елементом сучасного виробничого ландшафту. Вони дозволяють реалізувати концепції Smart Manufacturing, забезпечити гнучке управління, зменшити втрати, підвищити якість і адаптивність підприємств до нестабільного ринку. Завдяки швидкому розвитку інфраструктури штучного інтелекту, хмарних технологій і промислового Інтернету речей (IIoT), нейромережеві рішення поступово стають стандартом у промисловості XXI століття.

РОЗДІЛ 2. МЕТОДИ ТА ПІДХОДИ ДО ОПТИМІЗАЦІЇ ВИРОБНИЧИХ ПРОЦЕСІВ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

2.1. Оптимізація виробничого графіка

В умовах стрімкого розвитку промисловості та зростання конкуренції підприємства стикаються з необхідністю підвищення ефективності своїх виробничих процесів. Одним із ключових аспектів оптимізації є планування та управління виробничим графіком. Від правильності його побудови залежить раціональне використання ресурсів, мінімізація простоїв обладнання, зменшення витрат і, як наслідок, підвищення прибутковості компанії.

Традиційні методи планування виробничого процесу, які базуються на жорстко заданих алгоритмах або евристичних підходах, часто виявляються неефективними в умовах змінного попиту, невизначеності в постачанні сировини, можливих поломок обладнання чи людського фактора. У зв'язку з цим зростає потреба у впровадженні сучасних технологій, які дозволяють адаптивно керувати процесами, передбачати ризики та приймати оптимальні рішення в режимі реального часу.

Одним із найперспективніших інструментів для вирішення цих задач є штучний інтелект, зокрема нейронні мережі. Вони здатні аналізувати великі обсяги даних, визначати приховані закономірності та пропонувати ефективні стратегії управління виробничими процесами.

Оптимізація виробничого графіка – це складний багатофакторний процес, який охоплює кілька ключових аспектів:

- Нерівномірне завантаження обладнання. У виробничих умовах часто виникають ситуації, коли частина обладнання працює з перевантаженням, тоді як інші виробничі одиниці простоюють. Це призводить до неефективного використання ресурсів і зниження загальної продуктивності.
- Людський фактор. Робота персоналу також відіграє важливу роль у виробничих процесах. Планування змін, контроль робочого часу, уникнення

перевантаження співробітників – усе це впливає на ефективність виробництва.

- Різні пріоритети виконання замовлень. У реальних умовах не всі замовлення мають однакову важливість. Деякі з них потребують прискореного виконання, інші можуть бути виконані в більш гнучкі терміни. Визначення пріоритетності та балансування навантаження є важливим завданням при складанні виробничого графіка.
- Непередбачувані обставини. Поломки обладнання, затримки в постачанні сировини, зміни ринкового попиту – усе це фактори, які можуть суттєво впливати на виробничий процес і потребують оперативного коригування графіка.

У зв'язку з цими викликами виникає потреба у використанні технологій, які здатні швидко адаптуватися до змін і пропонувати оптимальні рішення на основі реальних даних.

Нейронні мережі дозволяють аналізувати історичні дані, прогнозувати навантаження та автоматично коригувати виробничий графік. Основні підходи до їх використання включають:

- Прогнозування навантаження. На основі аналізу минулих виробничих циклів нейромережа може передбачати пікові навантаження на обладнання, що дозволяє завчасно коригувати планування.
- Автоматичне коригування розкладу. За допомогою алгоритмів глибокого навчання система може у режимі реального часу змінювати послідовність операцій, перенаправляти ресурси та пропонувати альтернативні варіанти виконання завдань.
- Оптимізація розподілу ресурсів. Нейромережа аналізує наявність обладнання, стан виробничих ліній, завантаженість персоналу та підбирає найкращий спосіб їх використання.

Для вирішення завдань оптимізації виробничого графіка використовуються різні архітектури нейронних мереж:

- Рекурентні нейронні мережі (RNN, LSTM, GRU). Вони ефективні для аналізу часових рядів і прогнозування змін у виробничих процесах.
- Глибокі нейронні мережі (DNN). Дозволяють працювати з великими обсягами даних, визначати закономірності та будувати складні прогностичні моделі.
- Гібридні підходи. Поєднання нейронних мереж із генетичними алгоритмами та методами оптимізації, що дозволяє покращити адаптивність і швидкість ухвалення рішень.

Застосування нейронних мереж для оптимізації виробничого графіка має низку переваг:

- Підвищення продуктивності виробничих ліній за рахунок рівномірного розподілу навантаження.
- Мінімізація простоїв та неефективного використання ресурсів.
- Зниження витрат на позапланове обслуговування обладнання.
- Гнучкість у плануванні, що дозволяє швидко реагувати на зміни в умовах виробництва.
- Покращення точності прогнозування термінів виконання замовлень.

У результаті використання штучного інтелекту для управління виробничими процесами дозволяє значно підвищити ефективність роботи підприємств, зменшити витрати та забезпечити стабільне виконання замовлень у задані терміни.

2.2. Оптимізація енергоспоживання на підприємстві

Енергоспоживання є одним із ключових чинників, що безпосередньо впливає на економічну ефективність, стабільність і екологічну відповідальність промислових підприємств. У структурі операційних витрат енергоносії - зокрема електроенергія, газ, тепло- можуть становити значну частину постійних витрат, особливо в енергоємних галузях, таких як металургія, хімічна промисловість, машинобудування або харчове виробництво. Зі зростанням цін на енергоресурси, запровадженням вуглецевих квот, екологічних податків та стандартів

енергоефективності, питання оптимізації енергоспоживання набуває стратегічного значення.

Крім безпосередніх економічних витрат, значною проблемою є нестабільність енергетичних тарифів. Залежність від зовнішніх постачальників, коливання валютних курсів, зміни у державному регулюванні та геополітичні чинники можуть викликати різкі коливання вартості енергоресурсів, що унеможливорює довгострокове фінансове планування. Водночас, підвищення екологічної свідомості суспільства та вимоги до зменшення викидів CO₂ зобов'язують підприємства дотримуватись принципів сталого розвитку, що включає мінімізацію споживання енергії та впровадження чистих технологій. Традиційні підходи до управління енергоспоживанням – зокрема створення графіків роботи обладнання, ручна оптимізація виробничих режимів або використання стандартних енергетичних норм – часто виявляються недостатньо гнучкими. Вони ґрунтуються на статичних правилах, які не враховують динамічну природу виробничих процесів і низку змінних факторів, що впливають на ефективність енергоспоживання.

Сучасні технології на основі штучного інтелекту, зокрема нейронні мережі, дозволяють аналізувати великі масиви даних і знаходити оптимальні стратегії управління енергоспоживанням. Вони здатні прогнозувати витрати енергії, виявляти аномалії та автоматично коригувати режими роботи обладнання, що сприяє значному зниженню витрат.

Основні проблеми енергоспоживання на підприємствах.

Під час виробничого процесу підприємства стикаються з низкою проблем, пов'язаних із використанням енергоресурсів:

- Нерівномірне навантаження на енергосистему. Багато підприємств споживають енергію нерівномірно протягом доби, що призводить до пікових навантажень та підвищених тарифів на електроенергію.
- Неefективне використання обладнання. Більшість систем працюють у стандартному режимі без урахування реальних потреб у споживанні

електроенергії. Наприклад, вентиляційні системи, освітлення або виробничі лінії можуть працювати навіть тоді, коли це не потрібно.

- Втрати енергії. Через зношеність обладнання, недосконалість технологій або погане управління процесами значна частина енергії витрачається марно.
- Екологічні вимоги та обмеження. Сучасні підприємства повинні зменшувати викиди вуглекислого газу та впроваджувати енергоефективні технології, щоб відповідати екологічним стандартам.

Роль нейронних мереж у оптимізації енергоспоживання.

Нейронні мережі здатні аналізувати історичні дані про споживання електроенергії, враховувати зовнішні фактори (температуру, навантаження на мережу, години пік) і пропонувати оптимальні стратегії управління. Основні напрями їх застосування включають:

- Прогнозування енергоспоживання. Нейромережа аналізує попередні дані та прогнозує, скільки електроенергії знадобиться підприємству в конкретний час. Це дозволяє уникати пікових навантажень і скорочувати витрати.
- Автоматичне керування обладнанням. Інтелектуальні системи можуть вмикати та вимикати пристрої залежно від рівня споживання, а також оптимізувати роботу техніки з урахуванням навантаження.
- Виявлення аномалій. Нейромережі здатні знаходити відхилення у споживанні енергії та сигналізувати про можливі несправності обладнання або неефективне використання ресурсів.

Методи реалізації.

Для оптимізації енергоспоживання можуть використовуватися такі архітектури нейронних мереж:

- Рекурентні нейронні мережі (RNN, LSTM). Використовуються для аналізу послідовностей даних та прогнозування майбутнього споживання.
- Глибокі нейронні мережі (DNN). Дозволяють обробляти великі масиви даних та знаходити складні закономірності.

- Гібридні системи. Поєднання нейронних мереж із алгоритмами оптимізації (генетичними алгоритмами, методами рою часток), що дає змогу знаходити найкращі стратегії управління енергоспоживанням.

Очікувані результати впровадження.

Використання нейронних мереж для оптимізації енергоспоживання дозволяє досягти таких результатів:

- Зменшення витрат на електроенергію за рахунок точного прогнозування та адаптивного керування споживанням.
- Зниження навантаження на обладнання та продовження терміну його експлуатації.
- Автоматичне реагування на зміни у споживанні, що зменшує ризик аварійних ситуацій.
- Відповідність екологічним стандартам та зниження викидів парникових газів.

Таким чином, впровадження нейромереж у систему управління енергоспоживанням дозволяє не лише знизити витрати, а й підвищити загальну ефективність роботи підприємства.

2.3 Автоматизоване виявлення дефектів продукції

У сучасному світі, де темпи розвитку промисловості, технологій і вимог до якості зростають щодня, питання забезпечення стабільного контролю якості продукції стає надзвичайно актуальним. Конкурентоспроможність підприємства на ринку багато в чому залежить не лише від вартості продукції чи швидкості виробництва, а й від її відповідності встановленим стандартам якості. В умовах великосерійного або безперервного виробництва ручна перевірка кожної одиниці продукції є економічно не вигідною та технічно складною. Людський фактор, зокрема, суб'єктивність оцінки та фізичні обмеження не дозволяють досягти стабільної, високої точності виявлення дефектів. Саме тому в останні роки широкого поширення набули технології автоматизованого виявлення дефектів –

сучасний підхід до організації контролю якості, заснований на використанні комп'ютерного зору, штучного інтелекту, глибинного навчання та автоматизованих систем обробки даних.

Автоматизоване виявлення дефектів продукції – це складний, багаторівневий процес, що передбачає виявлення, аналіз і класифікацію можливих відхилень від нормативних характеристик виробів за допомогою технічних і програмних засобів. У загальному розумінні, система автоматичного контролю якості функціонує як поєднання апаратної частини (засоби зчитування, сенсори, камери, освітлення тощо) та програмного забезпечення, яке виконує аналіз отриманих даних, виявляє потенційні дефекти та передає інформацію про результат перевірки далі – або до систем сортування, або на екран оператора, або в систему збору статистики. При цьому система працює в реальному часі, тобто здійснює аналіз без уповільнення виробничого процесу, що особливо важливо в умовах швидкісних ліній або потокового виробництва.

Особливу роль у таких системах відіграє комп'ютерний зір – напрям ІТ, який дозволяє «навчити» машину бачити, інтерпретувати та аналізувати зображення так, як це робить людина, але при цьому — набагато швидше, точніше і без емоційних або фізичних відхилень. Наприклад, камера фіксує поверхню деталі або виробу під різними кутами або при різному освітленні, після чого зображення проходить через певні алгоритми обробки — це можуть бути як класичні методи (фільтрація, контурний аналіз, зіставлення з еталоном), так і новітні підходи, побудовані на глибинному навчанні та нейронних мережах. Останні особливо ефективні у випадках, коли дефекти мають складну структуру, слабо виражені ознаки або варіативність проявів, як-от подряпини, мікротріщини, зміщення елементів, деформації або сторонні включення.

Завдяки використанню методів машинного навчання, зокрема згорткових нейронних мереж (CNN), системи стають здатними не лише виявляти відхилення, але й навчатися на реальних даних, адаптуючись до нових типів дефектів, навіть якщо вони раніше не зустрічалися у виробництві. Це принципово відрізняє сучасні автоматизовані системи від жорстко налаштованих алгоритмів, які

працюють лише в межах заданих параметрів. Водночас це відкриває перспективи для розробки самонавчальних систем, які з часом підвищують свою точність і ефективність без втручання людини.

Не менш важливим компонентом таких рішень є модулі інтеграції та зв'язку з іншими елементами виробничої системи. Автоматизоване виявлення дефектів часто не є ізольованим етапом – воно інтегрується у ширшу архітектуру автоматизованої системи управління підприємством (АСУП), де результати контролю впливають на прийняття рішень про перенаправлення продукції, зупинку виробничого процесу, автоматичну зміну параметрів обладнання тощо. Такі інтеграційні можливості відкривають шлях до реалізації концепції «розумного виробництва» або «Індустрії 4.0», де всі елементи – від датчиків до аналітики – працюють у єдиному цифровому просторі.

Сфера застосування автоматизованих систем контролю якості надзвичайно широка. Вони успішно використовуються у металургії, де важливо виявити найменші тріщини або відшарування на прокаті; у харчовій промисловості — для контролю упаковки, маркування та зовнішнього вигляду продукції; у фармацевтичній галузі – для перевірки герметичності, цілісності та відсутності сторонніх часток; у текстильній та швейній промисловості – для виявлення дефектів плетіння або зміни кольору. Сучасна електроніка, автомобілебудування, виробництво медичних пристроїв – усі ці галузі мають високі вимоги до якості та безпеки продукції, що робить автоматизований контроль не просто бажаним, а обов'язковим елементом виробничого процесу.

На виробничій лінії встановлена камера, яка знімає продукцію. Зображення передаються на нейронну мережу, що аналізує їх та автоматично визначає, чи є дефекти на продукції. У разі виявлення дефекту продукція автоматично відбраковується.

Автоматизоване виявлення дефектів продукції дозволяє значно підвищити якість виробів та скоротити витрати на контроль якості. Використання нейронних мереж забезпечує високу точність та швидкість аналізу.

Таким чином, автоматизоване виявлення дефектів продукції можна розглядати не лише як технічне нововведення, а як стратегічну інвестицію в якість, репутацію та ефективність підприємства. У міру розвитку технологій ці системи стають доступнішими, гнучкішими й адаптивнішими, відкриваючи нові можливості для підприємств будь-якого масштабу. З кожним роком автоматизовані рішення удосконалюються, стають розумнішими, самонавчальними, здатними працювати в умовах невизначеності, змін і нових викликів. І саме в цьому криється їх головна перевага – здатність забезпечити безперервний, точний і надійний контроль якості там, де людський ресурс уже вичерпує свої можливості.

Один із базових підходів, який застосовується в системах автоматизованого виявлення дефектів, передбачає порівняння отриманого зображення з еталонним зразком. Для цього можуть використовуватись класичні методи обчислення відхилень, наприклад, середньоквадратичне відхилення (MSE), яке обчислюється за формулою:

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (2.1)$$

де x_i – значення інтенсивності пікселя у перевірюваному зображенні, $\hat{x}_i$ – відповідне значення еталонного зразка, а n – загальна кількість пікселів. Значне відхилення від еталону свідчить про потенційний дефект.

У випадку застосування нейронних мереж, зокрема згорткових (CNN), для класифікації або сегментації зображень, основна мета мережі – мінімізувати функцію втрат (loss function), наприклад, бінарну крос-ентропію:

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (2.2)$$

де y_i – реальне значення (1 – дефект, 0 – норма), $\hat{y}_i$ – ймовірність дефекту, яку прогнозує модель. Мережа навчається на численних прикладах, і з часом здатна з великою точністю розпізнавати найдрібніші порушення.

Для оцінки ефективності такої моделі використовуються метрики, що базуються на категоріях класифікації (істинно позитивні, хибно позитивні тощо):

- Precision (Точність):

$$\frac{TP}{TP + FN} \quad (2.3)$$

- Recall (Повнота):

$$\frac{TP}{TP + FN} \quad (2.4)$$

- F1-міра (гармонійне середнє Precision і Recall):

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (2.5)$$

TP– кількість випадків, коли дефект правильно виявлено (true positive),

FP– кількість помилкових спрацювань (false positive),

FN– кількість пропущених дефектів (false negative).

Завдяки таким математичним та алгоритмічним підходам автоматизовані системи можуть бути адаптовані до будь-яких умов виробництва, масштабуватись і вдосконалюватись із мінімальним втручанням людини.

2.4. Порівняння ефективності методів оптимізації виробничих процесів

Оптимізація виробничих процесів є ключовим завданням сучасного промислового підприємства, яке прагне досягти високої продуктивності, ефективного використання ресурсів і зниження операційних витрат. Серед наявних підходів до оптимізації можна виділити кілька основних груп: класичні математичні методи, експертні системи, евристичні алгоритми (метаевристики), а також інтелектуальні моделі на основі штучних нейронних мереж. Кожен із зазначених підходів має свої переваги, обмеження та галузі застосування. Порівняння ефективності цих методів є важливим етапом вибору оптимального рішення для конкретних виробничих умов.

1. Класичні математичні методи

До цієї групи відносяться методи лінійного та нелінійного програмування, динамічного програмування, мережевих моделей (наприклад, графи, критичний шлях, транспортні задачі). Вони дозволяють формалізувати задачу оптимізації у вигляді математичної моделі та знайти точне або наближене рішення.

Переваги:

- строгість формулювання;
- можливість аналітичного аналізу рішень;
- гарантії збіжності для певного класу задач.

Недоліки:

- низька масштабованість при великій кількості змінних і обмежень;
- чутливість до неточностей вхідних даних;
- обмеження на тип функцій (лінійність, опуклість).

2. Евристичні та метаевристичні алгоритми

Сюди входять генетичні алгоритми, імітація відпалу, алгоритми рою частинок, мурашині алгоритми тощо. Вони широко використовуються для складних оптимізаційних задач, у яких класичні методи неефективні або не застосовні.

Переваги:

- висока гнучкість;
- ефективність у випадках із багатьма локальними екстремумами;
- можливість застосування до задач із великою кількістю змінних.

Недоліки:

- відсутність гарантії знаходження глобального оптимуму;
- потреба у великій кількості ітерацій;
- складність налаштування параметрів (наприклад, швидкість мутації, розмір популяції).

3. Методи штучного інтелекту (нейронні мережі)

Нейронні мережі, зокрема глибокі згорткові (CNN), рекурентні (RNN, LSTM) та моделі з підкріпленням (Reinforcement Learning), дедалі активніше використовуються у виробничих системах для вирішення задач прогнозування, управління та оптимізації. Вони дозволяють виявляти складні закономірності у великих обсягах неструктурованих даних, адаптуватися до змінних умов та навчатися на основі зворотного зв'язку.

Переваги:

- здатність до самоорганізації і навчання з даних;
- ефективність у задачах з високою розмірністю ознак;
- висока точність прогнозування та прийняття рішень у складних динамічних системах;
- можливість роботи в режимі реального часу.

Недоліки:

- потреба у великій кількості навчальних даних;
- складність інтерпретації результатів (модель — «чорна скринька»);

- вимоги до обчислювальних ресурсів.

4. Емпіричне порівняння ефективності

Для порівняння ефективності в умовах виробничої задачі зазвичай використовують такі метрики оцінки:

Таблиця 2.1

Показник	Класичні методи	Евристики	Нейромережі
Час обчислення	Низький (для малих задач)	Середній	Високий (під час навчання)
Якість рішення	Висока (для точних моделей)	Висока (за налаштуванням)	Дуже висока (після навчання)
Масштабованість	Низька	Висока	Висока
Гнучкість	Низька	Висока	Висока
Стійкість до шуму	Низька	Середня	Висока
Інтерпретованість	Висока	Середня	Низька

Приклад: у задачі прогнозування енергоспоживання нейромережева модель LSTM показала середню абсолютну помилку (MAE) на рівні 4.3%, тоді як генетичний алгоритм дав похибку 7.9%, а лінійна регресія – 12.1%.

5. Гібридні підходи

На практиці все частіше застосовуються комбіновані рішення, наприклад:

- Попереднє формування обмежень за допомогою класичних моделей – подальша оптимізація нейромережею.
- Використання результатів евристики як стартових значень для нейромережевого агента.

- Постійне підкріплення нейромережі експертними правилами або обмеженнями із виробничих стандартів.

Такі підходи дозволяють досягати вищої точності, надійності та адаптивності, ніж кожен з методів окремо.

Порівняльний аналіз засвідчує, що жоден метод не є універсальним для всіх типів виробничих задач. Класичні методи доцільні для формалізованих, маломасштабних задач із відомими функціями. Евристики ефективні в умовах складної комбінації змінних. Нейронні мережі найбільш перспективні для динамічних, багатофакторних, нечітких і даноцентричних задач, де традиційні підходи не справляються.

У контексті індустрії 4.0 нейромережі стають основою гнучких, самонавчальних і масштабованих виробничих систем, що працюють в режимі реального часу. Їх поєднання з іншими методами відкриває шлях до створення повноцінних систем автономної оптимізації виробничих процесів.

РОЗДІЛ 3. ДОСЛІДЖЕННЯ НЕЙРОННИХ МЕРЕЖ

3.1 Аналіз прикладного використання базової моделі класифікації

У промислових умовах тротуарна плитка є типовим об'єктом для візуального контролю якості, оскільки вона має низку потенційних дефектів (тріщини, відколи, деформації), які можуть бути автоматично виявлені за допомогою комп'ютерного зору. Завдання класифікації полягає в аналізі зображення об'єкта і присвоєнні йому однієї з двох категорій:

- "Цілий" зразок (відповідає стандарту),
- "Дефектний" зразок (виявлено пошкодження або відхилення).

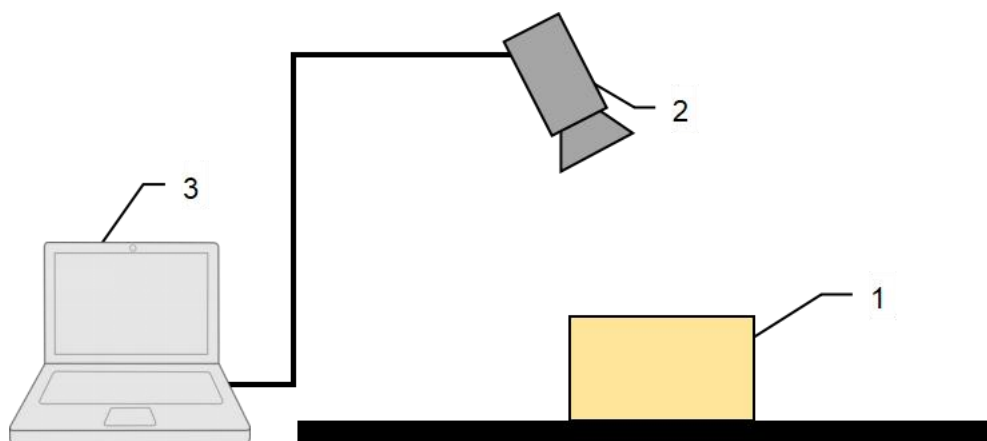


Рисунок 3.1. Схема експериментальної установки: 1 – Об'єкт класифікації; 2 – Камера; 3 – ПК.

Модель за результатами навчання продемонструвала середню точність класифікації 97,98% на тестовій вибірці, при середньому часі обробки одного зображення – 2,517 секунди. Ці результати є свідченням високої ефективності моделі навіть при її базових налаштуваннях, а також демонструють потенціал використання машинного навчання для задач технічного контролю якості продукції на підприємствах.

Після навчання базової моделі, яка продемонструвала задовільні результати як у плані точності класифікації, так і продуктивності, наступним кроком стало вдосконалення моделі шляхом оптимізації її архітектури. Основна мета цього етапу – знайти найкраще співвідношення між точністю класифікації та часом виконання, що є критичним для застосування в умовах реального виробництва.

Для досягнення цього було проведено серію експериментів, у яких змінювалися два параметри архітектури моделі:

1. WidthMultiplier (множник ширини)

Цей параметр керує кількістю фільтрів (каналів) у згорткових шарах моделі.

Зменшення значення множника ширини дозволяє:

- знизити кількість обчислень;
- пришвидшити виконання моделі;
- зменшити розмір самої моделі.

Однак, така оптимізація може потенційно призвести до втрати точності класифікації, оскільки модель оперує меншою кількістю параметрів і меншою глибиною розпізнавання ознак.

У ході дослідження були протестовані чотири варіанти множника ширини:

1.0 (базове значення), 0.75, 0.5 та 0.25.

Для кожного значення виконувалась обробка 10 контрольних зображень.

Результати експерименту наведено в таблиці 3.1, де фіксувалась як точність класифікації, так і час обробки одного зображення.

Таблиця 3.1

№ фот о	Точність, %				Час виконання, секунди			
	Множник ширини							
	0.25	0.5	0.75	1.0	0.25	0.5	0.75	1.0
1	100,00%	72,218%	98,08%	92,51%	0,359	0,829	1,453	2,484
2	100,00%	56,06%	99,29%	99,99%	0,344	0,812	1,453	2,515
3	100,00%	100,00%	81,90%	98,52%	0,375	0,812	1,5	2,515
4	100,00%	99,67%	99,85%	100,00%	0,375	0,797	1,453	2,515
5	100,00%	98,63%	25,78%	100,00%	0,359	0,812	1,453	2,562
6	100,00%	100,00%	100,00%	100,00%	0,375	0,812	1,453	2,578
7	100,00%	100,00%	100,00%	100,00%	0,359	0,812	1,453	2,5

8	100,00%	100,00%	99,99%	88,82%	0,359	0,828	1,50	2,5
9	100,00%	100,00%	99,94%	100,00%	0,359	0,812	1,453	2,5
10	99,93%	100,00%	100,00%	100,00%	0,375	0,812	1,437	2,5
Сер.	100,00%	92,66%	90,48%	97,98%	0,364	0,814	1,461	2,517

Основні спостереження:

- Час виконання прямо пропорційний множнику ширини: чим менше значення параметра, тим швидше працює модель.
- Точність не завжди зменшується при зменшенні ширини. Наприклад, при множнику 0.25 точність у середньому залишалась на рівні 100%, що свідчить про достатню здатність моделі виділяти ключові ознаки навіть при зменшеній кількості фільтрів.
- У ряді випадків точність при 0.75 виявилася нижчою, ніж при 0.5 чи 0.25, що може бути результатом випадкових флуктуацій або недостатнього навчання конкретної архітектури.

Найефективнішою з точки зору компромісу між точністю та швидкодією стала модель з множником 0.25, яка при середній точності 100% демонструвала середній час виконання 0,364 секунди, що в 6–7 разів швидше базової конфігурації. Такий результат робить її надзвичайно перспективною для реального впровадження в системи контролю якості.

2. ResolutionMultiplier (множник розширення на вході)

Цей параметр визначає розмір вхідного зображення, яке подається на модель.

Зменшення розширення дозволяє знизити обсяг обчислень, оскільки модель працює з меншою кількістю пікселів. Водночас існує ризик втрати дрібних або малококонтрастних ознак, важливих для точної класифікації.

У дослідженні розглядалися чотири варіанти вхідного розширення:

224 (базове значення), 192, 160 та 128 пікселів.

Таблиця 3.2

№ фот о	Точність, %				Час виконання, секунди			
	Розширення на вході							
	128	160	192	224	128	160	192	224
1	97,759%	100,00%	98,783%	92,510%	2,538	2,572	3,013	2,484
2	17,630%	100,00%	72,544%	99,993%	2,503	2,492	2,919	2,515
3	99,992%	100,00%	99,991%	98,520%	2,516	2,54	2,919	2,515
4	99,997%	90,776%	99,834%	100,00%	2,526	2,496	2,953	2,515
5	100,000 %	69,459%	100,000 %	100,000 %	2,485	2,479	2,801	2,562
6	100,00%	100,00%	99,999%	100,00%	2,568	2,528	2,866	2,578
7	100,00%	100,00%	99,898%	100,00%	2,481	2,524	2,819	2,5
8	99,999%	93,668%	88,967%	88,818%	2,481	2,559	2,72	2,5
9	100,00%	100,00%	11,033%	100,00%	2,492	2,503	2,908	2,5
10	99,770%	99,553%	100,00%	100,00%	2,499	2,562	2,733	2,5
Сер.	91,515%	95,346%	87,105%	97,984%	2,5089	2,5255	2,8653	2,5169

Основні спостереження:

- Зменшення розширення дозволяє незначно покращити продуктивність, зниження часу виконання до 2,48–2,50 секунд у порівнянні з 2,5169 секунд для базового значення (224).
 - Водночас точність значно варіюється і виявляє нелінійний характер залежності від розміру вхідного зображення. Наприклад, при розширенні 128 середня точність впала до 91,52%, що є значущим погіршенням.
 - Деякі приклади (зображення №2 при розширенні 128) показали критично низькі значення точності – лише 17,63%, що свідчить про втрату ключових ознак класифікації через агресивне зменшення розміру вхідного зображення.
- Зменшення розширення на вході не дає настільки суттєвого приросту швидкодії, як множник ширини, і часто супроводжується нестабільними або різко зниженими показниками точності. Таким чином, використання базового значення

(224 пікселі) є більш доцільним, особливо у випадках, коли точність є критичним показником.

Оптимізація моделі за допомогою параметрів `WidthMultiplier` та `ResolutionMultiplier` дозволяє значно покращити продуктивність без істотної втрати якості класифікації. Однак, вплив цих параметрів різний:

- Множник ширини (`WidthMultiplier`) має стабільний, передбачуваний і позитивний вплив на час виконання, з незначними коливаннями точності, тому є ключовим параметром для оптимізації.
- Множник розширення (`ResolutionMultiplier`) має нелінійний, менш прогнозований вплив, і в багатьох випадках його зменшення призводить до неприйнятної втрати точності.

З огляду на отримані результати, найкращим варіантом для задачі контролю якості тротуарної плитки виявилася модель з параметрами:

- `WidthMultiplier` = 0.25
- `Input Resolution` = 224×224

3.2 Дослідження з TensorFlow Object Detection API

У сучасних виробничих системах, де контроль якості продукції потребує обробки поточкових зображень із кількома об'єктами, традиційна класифікація виявляється недостатньою. У таких умовах ключовим стає завдання багатооб'єктного розпізнавання (object detection), яке передбачає не лише визначення класу кожного об'єкта, а й його точну локалізацію на зображенні. Це дозволяє системам технічного зору відстежувати якість продукції в реальному часі без необхідності модифікації виробничої лінії або ручного сортування.

Метою даного етапу дослідження було виявлення оптимального типу нейронної архітектури, здатного одночасно забезпечувати високу точність розпізнавання та швидкодію, необхідну для використання у виробничих умовах. Основним критерієм при відборі моделей була їх придатність до обробки відеопотоку в реальному часі, без суттєвих затримок у робочому циклі.

Для цього було використано фреймворк TensorFlow Object Detection API, який надає доступ до широкого спектру попередньо навчених моделей з різними характеристиками продуктивності, точності та архітектурної складності. Його перевагою є універсальність та підтримка об'єктної локалізації, що критично важливо для автоматизованого технічного контролю.

Порівнювані архітектури та принципи їх роботи. У межах аналізу були обрані дві фундаментально різні архітектури:

- Faster R-CNN (Region-based Convolutional Neural Network) – архітектура, що містить додатковий модуль RPN (Region Proposal Network), який генерує регіони інтересу для подальшої класифікації.
- SSD (Single Shot Detector) – компактніша модель, у якій функції регресії координат та класифікації виконуються безпосередньо після кожного шару ознак (feature map), що забезпечує вищу швидкодію.

Принципова відмінність полягає в тому, що SSD обробляє об'єкти на основі щільної сітки, що генерується згортковими шарами, тоді як Faster R-CNN виконує додаткову генерацію регіонів, що підвищує точність (наприклад, згідно з COCO mAP), але знижує швидкість обробки через збільшення кількості обчислень.

Це дозволило оцінити ефективність різних архітектур у однакових умовах та зробити висновки щодо доцільності їх впровадження у виробниче середовище. Навчання виконувалося на датасеті з 516 зображень тротуарної плитки, які були вручну розмічені відповідно до наявності дефектів (відколи, тріщини, деформації). Умови моделювали реальне виробниче середовище з рухомими об'єктами на конвеєрі, що дозволяло оцінити ефективність моделей у практичному застосуванні.

Таблиця 3.3

Model name	Speed (ms)	COCO mAP
ssd_mobilenet_v1_0.75_depth_coco	26	18
ssd_mobilenet_v1_ppn_coco	26	20

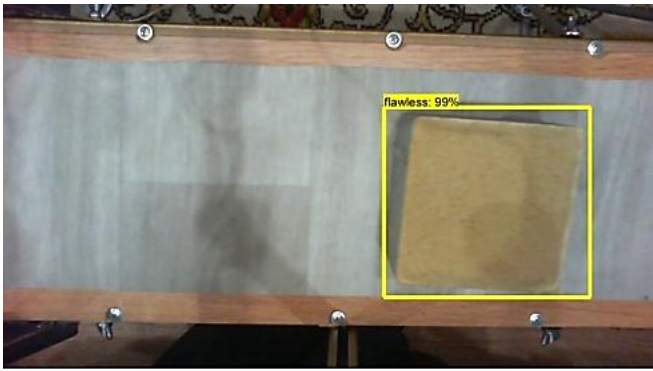
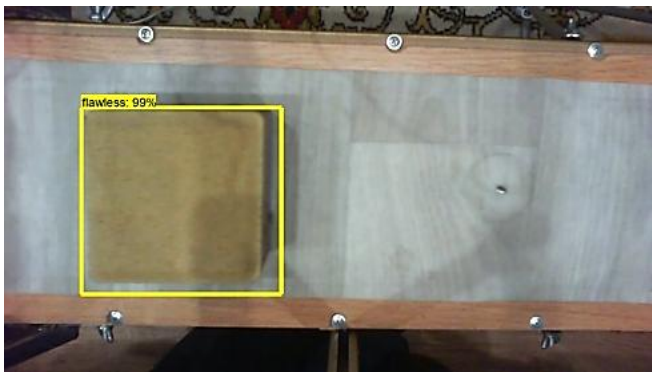
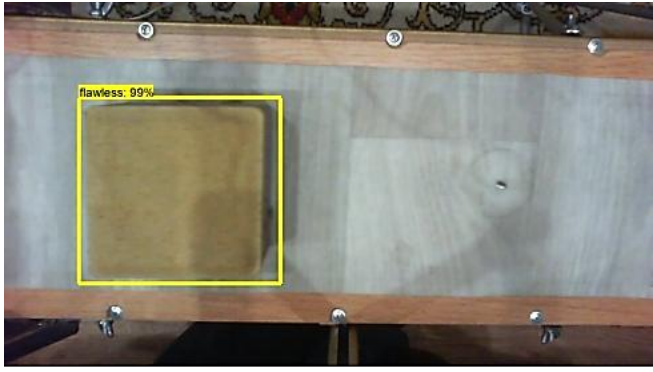
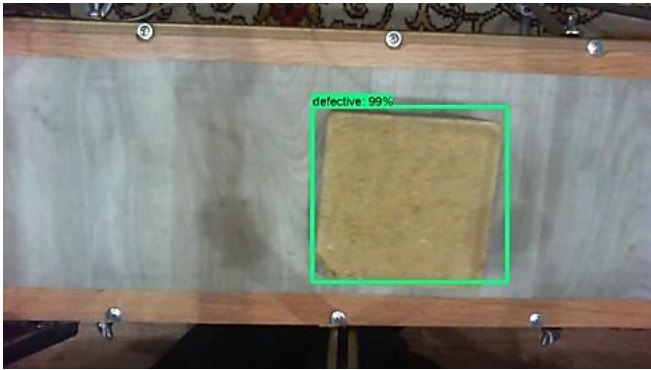
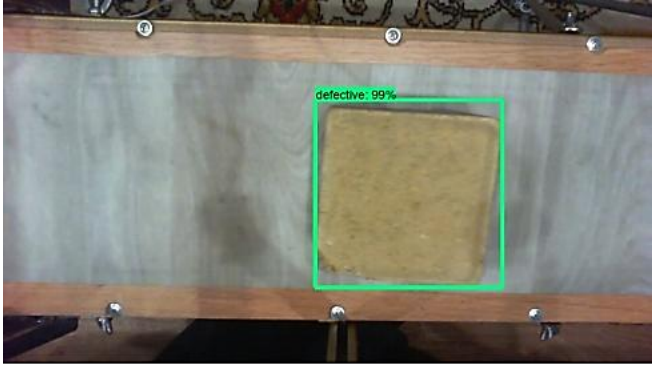
ssdlite_mobilenet_v2_coco	27	22
ssd_mobilenet_v1_0.75_depth_quantized_coco	29	16
ssd_mobilenet_v1_quantized_coco	29	18
ssd_mobilenet_v2_quantized_coco	29	22
ssd_mobilenet_v1_coco	30	21
ssd_mobilenet_v2_coco	31	22
ssd_inception_v2_coco	42	24
faster_rcnn_inception_v2_coco	58	28
faster_rcnn_resnet50_lowproposals_coco	64	
faster_rcnn_resnet101_lowproposals_coco	82	
faster_rcnn_resnet50_coco	89	30
faster_rcnn_resnet101_coco	106	32
faster_rcnn_inception_resnet_v2_atrous_lowproposals_coco	241	
faster_rcnn_nas_lowproposals_coco	540	
faster_rcnn_inception_resnet_v2_atrous_coco	620	37
faster_rcnn_nas	1833	43

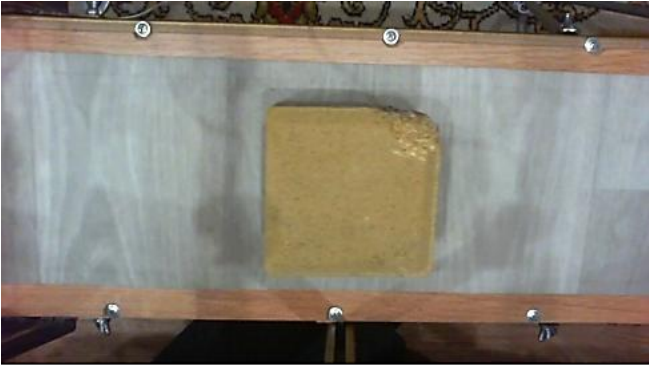
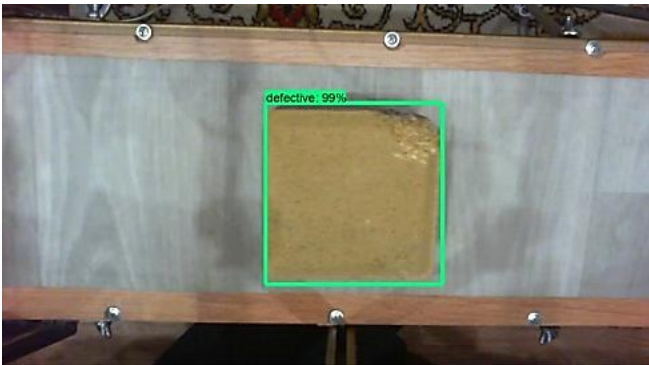
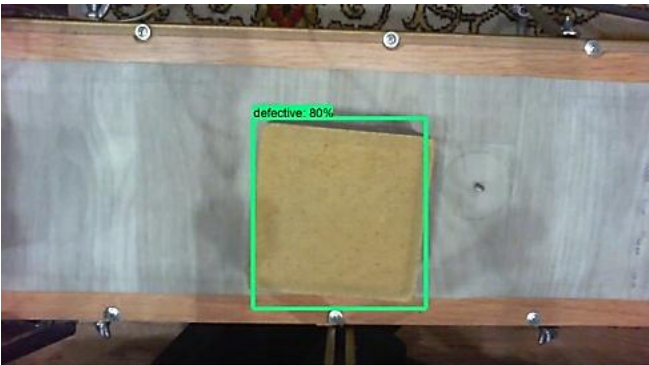
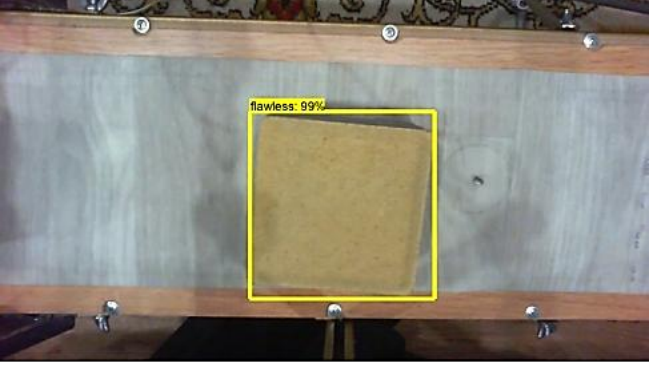
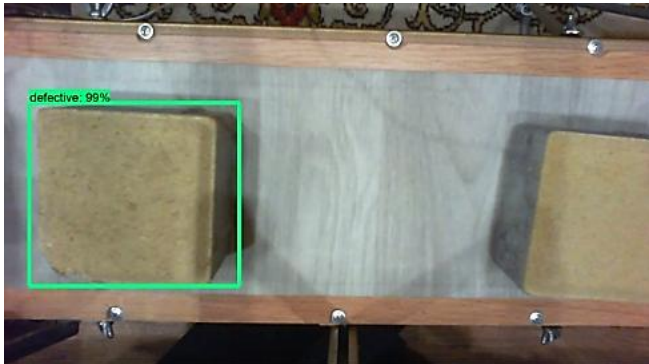
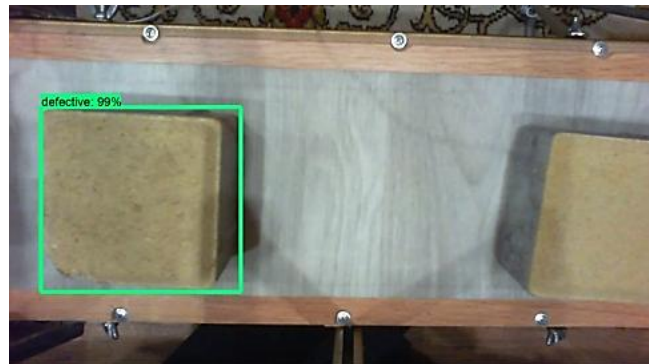
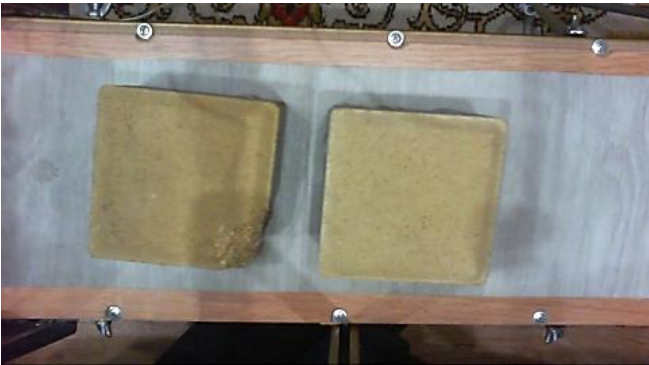
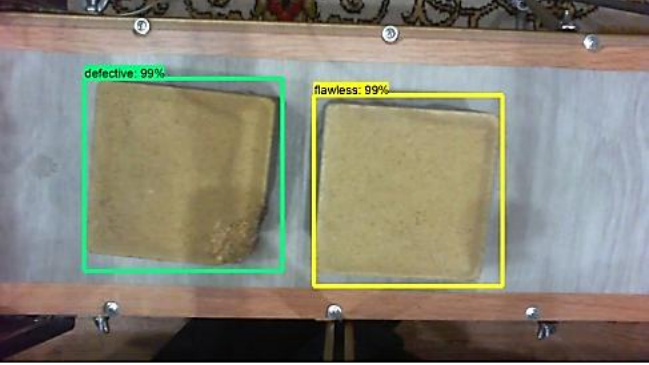
Продовження таблиці 3.3

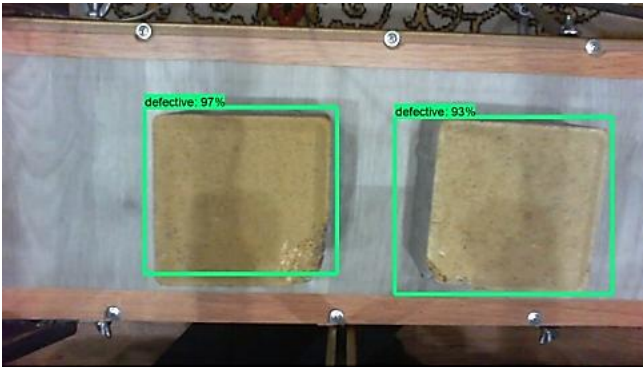
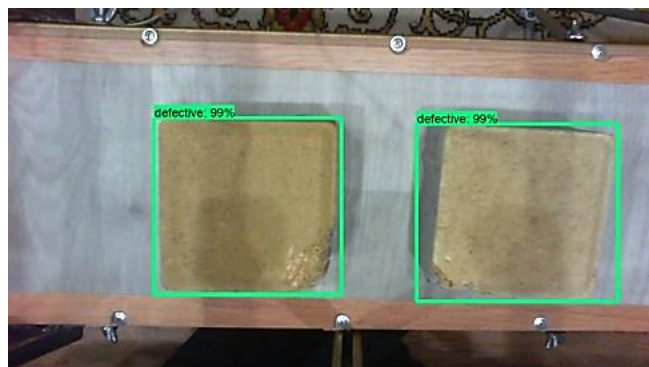
У процесі розпізнавання відеопотоку модель `ssd_inception_v2_coco` досягає швидкості обробки до 20 кадрів в секунду, в той час модель `faster_rcnn_inception_v2_coco` обробляє лише 5 кадрів.

В таблиці 3.4 представлені результати розпізнавання зображень обома моделями. Модель `ssd_inception_v2_coco` має достатньо великі втрати в точності, як розпізнавання (приклади 1, 4, 7), так і класифікації (приклад 5).

Таблиця 3.4

	ssd_inception_v2_coco	faster_rcnn_inception_v2_coco
1	 Не розпізнано	 Розпізнано правильно
2	 Розпізнано правильно	 Розпізнано правильно
3	 Розпізнано правильно	 Розпізнано правильно

4		
	Не розпізнано	Розпізнано правильно
5		
	Неправильний результат	Розпізнано правильно
6		
	Розпізнано правильно	Розпізнано правильно
7		
	Не розпізнані	Розпізнано правильно

8	 Розпізнано правильно	 Розпізнано правильно
---	---	---

Продовження таблиця 3.4

Примітка: Після успішного розпізнавання об'єкта модель візуально позначає його за допомогою кольорової рамки: жовтою у випадку виявлення якісної плитки та зеленою, якщо виявлено дефектний зразок.

У результаті проведених експериментів було встановлено, що, попри підвищену обчислювальну складність процесу розпізнавання об'єктів у порівнянні зі звичайною класифікацією, використання фреймворку TensorFlow Object Detection API у поєднанні з моделлю `faster_rcnn_inception_v2_coco` дозволяє ефективно здійснювати контроль якості продукції в режимі реального часу. Це робить даний підхід придатним для промислових умов без суттєвої затримки у виробничому циклі.

У випадках, коли отриманої точності моделі `faster_rcnn_inception_v2_coco` виявляється недостатньо, можливе застосування інших моделей сімейства Faster R-CNN, які забезпечують більш високу якість розпізнавання за рахунок зниження швидкодії. Таким чином, вибір конкретної моделі повинен враховувати баланс між вимогами до точності та обмеженнями за продуктивністю.

Щодо моделі `ssd_inception_v2_coco`, то вона продемонструвала незадовільні результати в низці випадків (наприклад, у прикладах 1, 4, 5 і 7). Основною причиною цього є архітектурні особливості SSD-моделей, які фокусуються на швидкодії, проте поступаються за точністю локалізації та розпізнавання дрібних або складних деталей. Виходячи з цього, можна припустити, що й інші моделі

типу SSD будуть менш ефективними в задачах, пов'язаних із контролем якості, де критично важливо виявляти навіть незначні дефекти.

У підсумку, обидва підходи – як класифікація, так і розпізнавання об'єктів виявилися працездатними та потенційно придатними для реалізації задач контролю якості. Проте між ними існує принципова різниця щодо практичного застосування. Зокрема, класифікація демонструє високу точність, але обмежується одним об'єктом на зображенні, що знижує її придатність у виробничих середовищах із потоковою обробкою продукції. Натомість, розпізнавання об'єктів є більш адаптивним до реалій виробництва, оскільки дозволяє аналізувати кілька об'єктів одночасно без необхідності модифікації виробничої лінії. Саме тому даний метод є доцільним вибором за замовчуванням для уже функціонуючих підприємств, де модернізація технічного процесу є небажаною або економічно невиправданою. Лише у випадках, коли точність виявлення виявиться критично недостатньою, варто розглядати застосування класичних класифікаційних моделей як альтернативу.

ВИСНОВКИ

У межах виконання дипломної роботи було досліджено теоретичні та прикладні аспекти використання штучних нейронних мереж у задачах оптимізації виробничих процесів. Результати дослідження підтверджують, що сучасні методи глибокого навчання можуть слугувати ефективним інструментом для автоматизації, підвищення точності контролю та зниження операційних витрат на підприємствах, що працюють у режимі масового виробництва.

У теоретичній частині роботи було систематизовано ключові поняття, класифікації та принципи функціонування нейронних мереж, що лежать в основі сучасних інтелектуальних систем. Проаналізовано особливості різних архітектур, зокрема згорткових нейронних мереж (CNN), автоенкодерів, моделей навчання з підкріпленням, а також гібридних підходів, які поєднують кілька видів навчання. Значну увагу було приділено методам навчання — як класичним (з учителем, без учителя), так і сучасним адаптивним алгоритмам (SGD, Adam, RMSProp), що забезпечують ефективну оптимізацію складних моделей.

У другому розділі дипломної роботи було окреслено ключові напрями застосування нейромереж у виробництві:

- оптимізація виробничого графіка;
- зниження енергоспоживання на підприємстві;
- автоматизоване виявлення дефектів продукції;
- порівняльна оцінка ефективності класичних і нейромережевих методів.

Зокрема, було детально проаналізовано проблему енергоефективності: виявлено обмеження традиційних алгоритмів у реальних динамічних умовах виробництва та обґрунтовано доцільність впровадження інтелектуальних моделей прогнозування споживання енергії, які здатні адаптуватися до змін робочих режимів і зовнішніх факторів (наприклад, коливань навантаження чи погодних умов).

Практична частина дослідження була присвячена аналізу й оцінці ефективності застосування моделей нейронних мереж для автоматизованого контролю якості продукції на прикладі тротуарної плитки. Було розглянуто два підходи:

- класифікація одного об'єкта на зображенні (на прикладі моделі MobileNet);
- розпізнавання кількох об'єктів одночасно (на прикладі моделей SSD та Faster R-CNN із використанням TensorFlow Object Detection API).

На підставі результатів, отриманих у процесі аналітичного огляду реалізованих моделей, було встановлено:

- Модель MobileNet, навіть у базовій конфігурації, здатна досягати високої точності (понад 97%) при короткому часі обробки, що робить її доцільною для задач класифікації одиничних зразків.
- Модель Faster R-CNN Inception V2 COCO продемонструвала найкращий баланс між точністю та швидкістю у задачі багатооб'єктного розпізнавання, що робить її придатною до промислового застосування у реальному часі.
- Модель SSD Inception V2 COCO, хоча і забезпечує вищу швидкість (до 20 кадрів/с), виявила недостатню точність при локалізації дефектів, що обмежує сферу її практичного використання у виробничих умовах.

Головні переваги використання нейронних мереж у виробництві, підтверджені дослідженням:

- здатність до обробки неструктурованих візуальних даних без необхідності ручного втручання;
- можливість виявляти складні дефекти, які не завжди ідентифікуються традиційними методами;
- гнучкість у адаптації до нових умов (нові типи продукції, зміни середовища);
- реальна перспектива інтеграції у системи розумного підприємства (Smart Factory, Industry 4.0).

Перспективи подальших досліджень

Отримані результати відкривають широкі можливості для подальшого вдосконалення систем штучного інтелекту у виробництві. Зокрема, доцільно дослідити:

- тонке донавчання (fine-tuning) моделей на специфічних виробничих даних;
- застосування моделей з трансформерною архітектурою (Vision Transformer, DETR);
- використання мультимодальних підходів, які комбінують візуальні та сенсорні дані;
- впровадження систем самоконтролю та адаптації, які коригують свої параметри залежно від змін у процесі виробництва;
- глибоку інтеграцію в цифрову інфраструктуру підприємства через API, SCADA, MES або ERP-системи.

Таким чином, результати дипломного дослідження свідчать про високу ефективність нейронних мереж у задачах оптимізації виробничих процесів і підтверджують їхню практичну значущість для створення нових поколінь автоматизованих систем технічного контролю та управління.

Список використаних джерел

1. Гудфеллоу І., Бенджіо Й., Курвілль А. Глибоке навчання. – Київ: Наука, 2020. – 640 с.
2. Хайкин С. Нейронні мережі: повний курс. – М.: Вільямс, 2006. – 1104 с.
3. Chollet F. Deep Learning with Python. – 2nd ed. – Manning Publications, 2021. – 504 p.
4. Bishop C. M. Pattern Recognition and Machine Learning. – New York: Springer, 2006. – 738 p.
5. LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. – 2015. – Vol. 521(7553). – P. 436–444.
6. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // IEEE Trans. on Pattern Analysis and Machine Intelligence. – 2017. – Vol. 39(6). – P. 1137–1149.
7. Liu W., Anguelov D., Erhan D. et al. SSD: Single Shot MultiBox Detector // In: Computer Vision – ECCV 2016. – Springer, 2016. – P. 21–37.
8. TensorFlow Object Detection API. Documentation – https://github.com/tensorflow/models/tree/master/research/object_detection
9. Ковальов В. В., Шишкін О. П. Штучний інтелект: теорія і практика. – Харків: ХНУРЕ, 2021. – 298 с.
10. Андросов В. В., Павлов С. І. Інтелектуальні системи в управлінні виробництвом. – К.: НАУ, 2019. – 214 с.
11. Ковтун В. О., Лазаренко О. І. Методи оптимізації в задачах машинного навчання. – Київ: КПІ ім. Ігоря Сікорського, 2020. – 192 с.
12. ISO 22400-2:2014 Automation systems and integration – Key performance indicators (KPIs) for manufacturing operations management – Part 2: Definitions and descriptions.
13. Васильєв С. О. Застосування згорткових нейронних мереж у задачах технічного контролю // Вісник ХНУРЕ. – 2021. – №1. – С. 45–52.

14. Huang G., Liu Z., Maaten L., Weinberger K. Densely Connected Convolutional Networks // Proceedings of the IEEE CVPR. – 2017. – P. 4700–4708.
15. Karpathy A., Fei-Fei L. Deep Visual-Semantic Alignments for Generating Image Descriptions // Proceedings of the IEEE CVPR. – 2015. – P. 3128–3137.

ДЕМОНСТРАЦІЙНИЙ МАТЕРІАЛ (Презентація)

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО - КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА НА ТЕМУ:

МЕТОДИ ТА ПІДХОДИ ОПТИМІЗАЦІЇ ВИРОБНИЧИХ ПРОЦЕСІВ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

Виконала студентка групи САД-41 Шевчук Валерія Сергіївна
Науковий керівник: Кузьміч Михайло Юрійович

Актуальність теми

У сучасному виробництві постійно зростають вимоги до швидкості, якості та гнучкості виробничих процесів.

Нейронні мережі, як складова штучного інтелекту, відкривають нові можливості:

- аналіз великих обсягів даних у реальному часі;
- виявлення закономірностей, неочевидних для людини;
- адаптація до змін без перепрограмування систем.

Застосування нейронних мереж у виробництві дозволяє не лише покращити якість продукції та зменшити витрати, а й зробити процеси більш автономними, стійкими й ефективними.

Таким чином, дослідження методів оптимізації виробничих процесів із використанням нейронних мереж є надзвичайно актуальним для підприємств, які прагнуть впроваджувати інновації та цифрову трансформацію.

Мета, об'єкт, предмет дослідження

○ Мета дослідження:

Вивчення та аналіз підходів до оптимізації виробничих процесів за допомогою нейронних мереж для підвищення ефективності, гнучкості й адаптивності виробничих систем.

○ Об'єкт дослідження:

Виробничі процеси підприємств, що мають складну структуру, багатофакторність і потребують оптимізації.

○ Предмет дослідження:

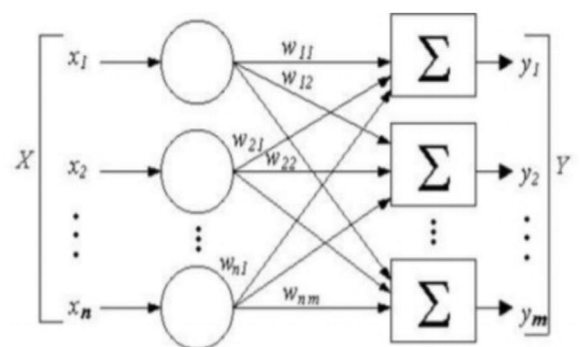
Методи застосування нейронних мереж для оптимізації виробництва: моделювання, прогнозування та автоматизація прийняття рішень.

Приклад структури одношарової нейронної мережі

Штучна нейронна мережа (ШНМ) — це обчислювальна модель, що імітує роботу біологічного мозку.

На зображенні представлено схему одношарової ШНМ, яка складається з:

- Вхідного вектора (X) — це набір ознак, що подаються в мережу (наприклад, пікселі зображення або технічні параметри виробу).
- Мережі ваг (w) — кожен вхід з'єднаний з кожним нейроном шару, причому кожен зв'язок має власну вагу.
- Нейронів (Σ) — кожен з них виконує підсумовування вхідних сигналів з урахуванням ваг і передає результат далі (іноді через активаційну функцію).
- Вихідного вектора (Y) — результат роботи мережі, наприклад: клас «норма» або «дефект».



Активаци́йні функції задають нерівномірну реакцію нейрона на вхідні дані. Вони дозволяють моделі: 1) виявляти складні залежності у даних; 2) вводити нелінійність у мережу; 3) керувати передачею сигналу між шарами.

Функція	Формула/Дія	Область застосування
ReLU (Rectified Linear Unit)	$f(x) = \max(0, x)$	Найбільш поширена для прихованих парів, швидко навчається, проста у реалізації
Sigmoid	$f(x) = \frac{1}{(1 + e^{-x})}$	Використовується для бінарної класифікації, видає значення від 0 до 1
Tanh (гіперболічний тангенс)	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	Схожа на Sigmoid, але симетрична відносно нуля: значення від -1 до 1
Softmax	$f(x, i) = \frac{e^{x_i}}{\sum e^{x_j}}$	Нормалізує вектор виходів у ймовірності (для багатокласової класифікації)

Методи навчання визначають здатність нейронних мереж адаптуватися, узагальнювати дані та приймати рішення. Існують три основні підходи:

- **Навчання з учителем** — модель навчається за мітками правильних відповідей. Найпоширеніший метод — градієнтний спуск із зворотним розповсюдженням помилки. Використовуються оптимізатори: SGD, Adam, RMSProp.
- **Навчання без учителя** — застосовується без міток. Мета — виявити приховану структуру даних. Використовуються автоенкодери, карти Кохонена, кластеризація.
- **Навчання з підкріпленням** — мережа взаємодіє з середовищем, отримуючи винагороду за дії. Застосовується у складних задачах управління. Приклади: Q-learning, DQN, Actor-Critic.

Метод	Переваги	Недоліки
Навчання з учителем	Висока точність при наявності розмічених даних	Залежність від якості та кількості даних
Навчання без учителя	Виявлення прихованих закономірностей	Складність інтерпретації результатів
Навчання з підкріпленням	Застосування в складних динамічних системах	Труднощі в налаштуванні параметрів та виборі винагород
Адаптивні методи	Швидка та стабільна оптимізація	Підбір гіперпараметрів може бути складним

Оптимізація виробничого графіка

Ефективне планування виробничих процесів — ключ до зменшення витрат, уникнення простоїв та рівномірного використання ресурсів.

Проблеми перевантаження обладнання, змінності попиту, людського фактору та затримок в постачанні потребують адаптивних рішень. Нейронні мережі дозволяють прогнозувати навантаження, автоматично коригувати розклади, розподіляти ресурси та враховувати пріоритети замовлень. На відміну від класичних алгоритмів, ШНМ працюють з динамічними даними в режимі реального часу.

Для таких задач застосовують RNN, LSTM, DNN та гібридні моделі, які здатні враховувати часові залежності, змінні умови виробництва та множинні обмеження.

Завдяки використанню нейронних мереж підприємства підвищують гнучкість планування, знижують операційні втрати та забезпечують стабільне виконання замовлень у строк.

Оптимізація енергоспоживання на підприємстві

Енерговитрати — один із найважливіших чинників у структурі собівартості продукції.

Нестабільні тарифи, пікові навантаження, втрати енергії та екологічні обмеження змушують підприємства шукати нові шляхи оптимізації.

Традиційні методи енергоменеджменту часто базуються на статичних графіках і не враховують динаміку виробництва. Нейронні мережі дозволяють аналізувати історичні та поточні дані, прогнозувати споживання, виявляти аномалії та автоматично коригувати режими роботи обладнання.

Серед найбільш ефективних рішень — рекурентні (RNN, LSTM), глибокі (DNN) та гібридні мережі, що враховують внутрішні та зовнішні фактори (навантаження, температуру, графіки роботи).

Застосування ШН дає змогу знизити витрати на електроенергію, продовжити ресурс обладнання, зменшити аварійність і відповідати сучасним стандартам сталого розвитку.

Автоматизоване виявлення дефектів продукції

У сучасному виробництві якість є критичним фактором конкурентоспроможності.

Ручний контроль — повільний, суб'єктивний і не масштабується. Автоматизовані системи на основі комп'ютерного зору й нейронних мереж дозволяють виявляти дефекти швидко, точно й у реальному часі.

Ключову роль у таких рішеннях відіграють згорткові нейронні мережі (CNN), здатні розпізнавати складні типи дефектів (подряпини, мікротріщини, деформації) навіть у складних умовах. Мережа навчається на реальних прикладах і з часом підвищує свою точність. Система включає камеру, обчислювальний модуль і алгоритм прийняття рішень. Вона може інтегруватися в загальну систему управління виробництвом (АСУП), автоматично сортувати продукцію або змінювати параметри обладнання.

Завдяки високій точності, швидкодії та здатності до адаптації такі системи стають основою концепції «розумного виробництва», забезпечуючи стабільний контроль якості та зниження витрат без людського втручання.

Аналіз базової моделі MobileNet для класифікації тротуарної плитки

Було проведено аналітичне тестування вже навчених моделей MobileNet для задач контролю якості тротуарної плитки (класифікація «ціла» / «дефектна»).

Базова модель показала точність 97,98% при середньому часі обробки 2,5 с на зображення.

Для оцінки придатності в реальному виробництві досліджено вплив двох параметрів:

- WidthMultiplier — зменшення кількості фільтрів у шарах
- ResolutionMultiplier — зменшення розміру вхідного зображення

Найкращий баланс досягнуто при WidthMultiplier = 0.25 та Resolution = 224×224.

№ фото	Точність, %				Час виконання, секунди			
	Множник ширини							
	0.25	0.5	0.75	1.0	0.25	0.5	0.75	1.0
1	100,00%	72,218%	98,08%	92,51%	0,359	0,829	1,453	2,484
2	100,00%	56,06%	99,29%	99,99%	0,344	0,812	1,453	2,515
3	100,00%	100,00%	81,90%	98,52%	0,375	0,812	1,5	2,515
4	100,00%	99,67%	99,85%	100,00%	0,375	0,797	1,453	2,515
5	100,00%	98,63%	25,78%	100,00%	0,359	0,812	1,453	2,562
6	100,00%	100,00%	100,00%	100,00%	0,375	0,812	1,453	2,578
7	100,00%	100,00%	100,00%	100,00%	0,359	0,812	1,453	2,5
8	100,00%	100,00%	99,99%	88,82%	0,359	0,828	1,50	2,5
9	100,00%	100,00%	99,94%	100,00%	0,359	0,812	1,453	2,5
10	99,93%	100,00%	100,00%	100,00%	0,375	0,812	1,437	2,5
Сер.	100,00%	92,66%	90,48%	97,98%	0,364	0,814	1,461	2,517

№ фото	Точність, %				Час виконання, секунди			
	Розширення на вхід							
	128	160	192	224	128	160	192	224
1	97,759%	100,000%	98,783%	92,510%	2,538	2,572	3,013	2,484
2	17,630%	100,000%	72,544%	99,993%	2,503	2,492	2,919	2,515
3	99,992%	100,000%	99,991%	98,520%	2,516	2,54	2,919	2,515
4	99,997%	90,776%	99,834%	100,000%	2,526	2,496	2,953	2,515
5	100,000%	69,459%	100,000%	100,000%	2,485	2,479	2,801	2,562
6	100,000%	100,000%	99,999%	100,000%	2,568	2,528	2,866	2,578
7	100,000%	100,000%	99,898%	100,000%	2,481	2,524	2,819	2,5
8	99,999%	93,668%	88,967%	88,818%	2,481	2,559	2,72	2,5
9	100,000%	100,000%	11,033%	100,000%	2,492	2,503	2,908	2,5
10	99,770%	99,553%	100,000%	100,000%	2,499	2,562	2,733	2,5
Сер.	91,515%	95,346%	87,105%	97,984%	2,5089	2,5255	2,8653	2,5169

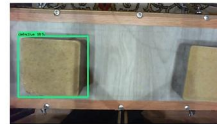
Об'єктне розпізнавання на базі TensorFlow API

Проведено порівняльний аналіз двох типів моделей: SSD та Faster R-CNN.

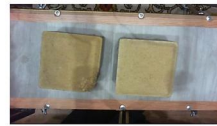
- SSD досягала 20 кадрів/сек, але мала помилки в розпізнаванні.
- Faster R-CNN працювала повільніше (5 кадрів/сек), але демонструвала стабільно високі результати.

Дані моделі не навчалися вручну, а оцінювались з точки зору їхньої придатності до промислового використання. Аналіз базувався на уже існуючих архітектурах і результатах роботи на доступному датасеті з зображеннями тротуарної плитки.

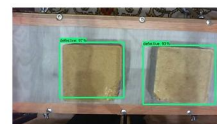
SSD (Single Shot Detector)



Розпізнано правильно



Не розпізнані

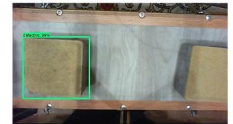


Розпізнано правильно

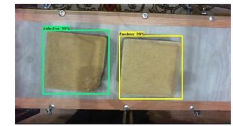


Неправильний результат

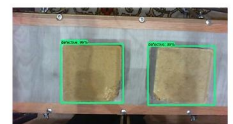
Faster R-CNN



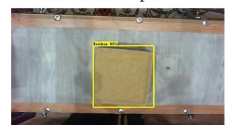
Розпізнано правильно



Розпізнано правильно



Розпізнано правильно



Розпізнано правильно

Примітка: жовта рамка у випадку виявлення якісної плитки; зелена рамка - дефектний зразок.

Висновки

У межах дипломної роботи було досліджено можливості застосування нейронних мереж для оптимізації виробничих процесів. Розглянуто теоретичні основи ШНМ, класифікацію архітектур, принципи навчання та їхню ефективність у виробничому середовищі.

Проведено аналітичний огляд у трьох прикладних напрямках:

- Оптимізація виробничого графіка — зменшення простоїв, підвищення гнучкості;
- Зниження енергоспоживання — точне прогнозування навантажень і економія ресурсів;
- Автоматизоване виявлення дефектів — підвищення якості та швидкості контролю.

У практичній частині проаналізовано роботу попередньо навчених моделей (MobileNet, SSD, Faster R-CNN) для задач класифікації та розпізнавання дефектів тротуарної плитки. Найкращий баланс між точністю та швидкістю показала модель Faster R-CNN Inception V2.

Отримані результати підтверджують, що нейронні мережі є ефективним інструментом для автоматизації виробництва, адаптації до змінних умов та інтеграції в концепцію Smart Factory.



ДЯКУЮ ЗА ВАШУ УВАГУ!