

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Прикладних інформаційних технологій та електроінженерії
(повна назва факультету)

Комп'ютерно-інтегрованих технологій
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр

(назва освітнього ступеня)

на тему: **Застосування методів машинного навчання для класифікації
даних в комп'ютеризованих системах керування технологічними
процесами**

Виконав: студент 6 курсу, групи КТм-62

спеціальності 174 – «Автоматизація, комп'ютерно-

інтегровані технології та робототехніка»

(шифр і назва спеціальності)

(підпис)

Яцишин Р.І.

(прізвище та ініціали)

Керівник

(підпис)

Микитишин А.Г.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Левицький В.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Микитишин А.Г.

(прізвище та ініціали)

Рецензент

(підпис)

Марущак П.О.

(прізвище та ініціали)

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет прикладних інформаційних технологій та електроінженерії
(повна назва факультету)

Кафедра комп'ютерно-інтегрованих технологій
(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Микитишин А.Г

(підпис)

(прізвище та ініціали)

« 29 » жовтня 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня магістр
(назва освітнього ступеня)

за спеціальністю 174 – «Автоматизація, комп'ютерно-інтегровані технології та
робототехніка»
(шифр і назва спеціальності)

студенту Яцишину Роману Ігоровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Застосування методів машинного навчання для класифікації даних в
комп'ютеризованих системах керування технологічними процесами

Керівник роботи Микитишин Андрій Григорович, к.т.н., доцент
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від « 29 » жовтня 2024 року № 4/7-1030

2. Термін подання студентом завершеної роботи 22.12.2024 р.

3. Вихідні дані до роботи методи класифікації, критерії їх оцінки, нейронні мережі,
підсилени дерева, випадкові ліси, датасет

4. Зміст роботи (перелік питань, які потрібно розробити)

Аналітична частина. 2. Технологічна частина. 3. Конструкторська частина. 4. Науково-
дослідна частина. 5. Спеціальна частина. 6. Охорона праці та безпека в надзвичайних
ситуаціях.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1, 2. Актуальність теми та мета роботи. 3, 4, 5. Архітектура нейронних мереж. 6. Алгоритм
підсилених дерев. 7. Метод випадкових лісів. 8. Втомне руйнування. 9. Діаграма
втомного руйнування.

10, 11, 12. Процес збору, обробки та підготовки даних для класифікації. 13-18. Отримані
результати класифікації. 19. Аналіз результатів точності класифікування.

20. Висновки.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці та безпека в надзвичайних ситуаціях	кандидат технічних наук, доцент кафедри КТ Тотосько Олег Васильович		
	проректор з адміністративно-господарської роботи та будівництва		
	Клепчик Василь Михайлович		

7. Дата видачі завдання 29 жовтня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Огляд комп'ютеризованих систем керування технологічними процесами	02.11.2024	Виконано
2.	Сучасні методи машинного навчання для задач класифікації	04.11.2024	Виконано
3.	Огляд характеристик діаграм втомного руйнування алюмінієвого сплаву	16.11.2024	Виконано
4.	Попередня обробка даних: нормалізація, масштабування, очистка	30.11.2024	Виконано
5.	Розробка архітектури комп'ютеризованої системи керування з інтеграцією методів МН	05.12.2024	Виконано
6.	Науково-дослідна частина	08.12.2024	Виконано
7.	Спеціальна частина	12.12.2024	Виконано
8.	Охорона праці та безпека в надзвичайних ситуаціях	15.12.2024	Виконано
9.	Оформлення графічної частини	18.12.2024	Виконано
10.	Захист кваліфікаційної роботи	24.12.2023	Виконано

Студент

(підпис)

Яцишин Р.І.
(прізвище та ініціали)

Керівник роботи

(підпис)

Микитишин А.Г.
(прізвище та ініціали)

АНОТАЦІЯ

Застосування методів машинного навчання для класифікації даних в комп'ютеризованих системах керування технологічними процесами // Кваліфікаційна робота освітнього рівня «Магістр» // Яцишин Роман Ігорович // Тернопільський національний технічний університет імені Івана Пулюя, факультет прикладних інформаційних технологій та електроінженерії, кафедра комп'ютерно-інтегрованих технологій, група КТм-62 // Тернопіль, 2024 // С. 75, рис. – 24, табл. – 0, додат. – 1, бібліогр. – 19.

Ключові слова: методи машинного навчання, штучний інтелект, класифікація, випадкові ліси, підсилені дерева, нейронні мережі, асиметрій циклу навантажування, діаграма втомного руйнування, швидкість РВТ.

Кваліфікаційна робота присвячена застосуванню методів машинного навчання в комп'ютеризованих системах керування технологічними процесами, котрі відкривають нові можливості для оптимізації виробничих процесів, покращення контролю якості, прогнозування несправностей та зменшення витрат. Класифікація даних за допомогою алгоритмів машинного навчання дозволяє автоматизувати аналіз великого обсягу даних, зменшуючи людський фактор та підвищуючи точність і швидкість прийняття рішень.

У першому розділі представлено аналіз сучасного стану досліджень у сфері комп'ютеризованих систем керування технологічними процесами. Розглянуто актуальні методи машинного навчання, такі як нейронні мережі, випадкові ліси та підсилені дерева, що застосовуються для задач класифікації.

У другому розділі кваліфікаційної роботи детально описано характеристики діаграми втомного руйнування, постановку задач класифікації типу навантаження, а також методики підготовки даних для навчання моделей, обробки даних (очищення, нормалізація, масштабування) та вибору критеріїв оцінки якості класифікації.

Третій розділ присвячено розробленню архітектури комп'ютеризованої системи з інтегрованими моделями нейронних мереж, випадкових лісів та підсилених дерев. Описано вибір програмного середовища для реалізації моделей машинного навчання, а також забезпечення надійності та продуктивності системи.

У четвертому розділі представлено реалізацію та навчання моделей машинного навчання. Виконано тестування моделей на основі експериментальних даних, отриманих з діаграм втомного руйнування. Проведено порівняльний аналіз ефективності моделей за допомогою обраних метрик точності. Обґрунтовано вибір найоптимальнішої моделі для задачі класифікації.

П'ятий розділ містить практичну реалізацію класифікації типу навантаження та перевірку надійності та стійкості системи у різних режимах роботи.

В шостому розділі кваліфікаційної роботи розглянуто забезпечення безпечної роботи з обладнанням.

ANNOTATION

Application of machine learning methods for data classification in computerized control systems for technological processes // Master's degree qualification work / Yatsyshyn Roman Ihorovych // Ternopil Ivan Puluj National Technical University, Faculty of Applied Information Technologies and Electrical Engineering, Department of Computer-Integrated Technologies, group KTm-62 // Ternopil, 2024 // P. 75, Fig. 24, Table 0, Supplement 1, Bibliography 20.

Keywords: machine learning methods, artificial intelligence, classification, random forests, boosted trees, neural networks, stress ratio, fatigue crack growth diagrams.

The qualification work is devoted to the application of machine learning methods in computerized process control systems, which open up new opportunities for optimizing production processes, improving quality control, predicting faults, and reducing costs. Data classification using machine learning algorithms allows automating the analysis of large amounts of data, reducing the human factor and increasing the accuracy and speed of decision-making.

The first chapter presents an analysis of the current state of research in the field of computerized process control systems. The actual machine learning methods, such as neural networks, random forests, and reinforced trees used for classification tasks, are considered.

The second chapter of the qualification work describes in detail the characteristics of the fatigue failure diagram, the setting of load type classification tasks, as well as methods of data preparation for model training, data processing (cleaning, normalization, scaling), and selection of classification quality assessment criteria.

The third chapter is devoted to the development of the architecture of a computerized system with integrated models of neural networks, random forests, and reinforced trees. It describes the choice of software environment for implementing

machine learning models, as well as ensuring the reliability and performance of the system.

The fourth chapter presents the implementation and training of machine learning models. The models are tested on the basis of experimental data obtained from fatigue failure diagrams. A comparative analysis of the models' efficiency is carried out using the selected accuracy metrics. The choice of the most optimal model for the classification task is substantiated.

The fifth chapter contains the practical implementation of load type classification and verification of system reliability and stability in various operating modes.

The sixth chapter of the qualification work deals with ensuring safe operation of equipment.

ПЕРЕЛІК СКОРОЧЕНЬ І ТЕРМІНІВ

КСКТП – комп'ютеризовані системи керування технологічними процесами;

ШІ – штучний інтелект;

МН – машинне навчання;

НМ – нейронні мережі;

ВТ – підсилені дерева;

RF – випадкові ліси;

ДВР – діаграма втомного руйнування;

PVT – ріст втомної тріщини;

KIH – коефіцієнт інтенсивності напружень;

R – асиметрія циклу навантаження;

da/dN – швидкість росту втомної тріщини.

ЗМІСТ

ВСТУП	10
РОЗДІЛ 1. АНАЛІТИЧНА ЧАСТИНА.....	12
1.1 Огляд комп'ютеризованих систем керування технологічними процесами.....	12
1.2 Основні задачі класифікації даних у технологічних процесах ..	16
1.3 Сучасні методи машинного навчання для задач класифікації	17
1.4 Особливості застосування нейронних мереж, випадкових лісів, підсилених дерев	26
РОЗДІЛ 2. ТЕХНОЛОГІЧНА ЧАСТИНА.....	28
2.1 Огляд характеристик діаграм втомного руйнування алюмінієвого сплаву.....	28
2.2 Постановка задачі класифікації типу навантаження за коефіцієнтом асиметрії R	32
2.3 Опис вхідних даних: параметри da/dN , K та R	34
2.4 Попередня обробка даних: нормалізація, масштабування, очистка.....	36
2.5 Вибір метрик для оцінки якості класифікації	38
РОЗДІЛ 3. КОНСТРУКТОРСЬКА ЧАСТИНА.....	41
3.1 Розробка архітектури комп'ютеризованої системи керування з інтеграцією методів МН	41
3.2 Вибір програмного забезпечення для інтеграції методів класифікації.....	43
3.3 Забезпечення надійності та продуктивності системи	46
РОЗДІЛ 4. НАУКОВО-ДОСЛІДНА ЧАСТИНА	48
4.1 Побудова моделей машинного навчання для класифікації	48
4.2 Аналіз впливу параметрів моделей на якість класифікації	53
4.3 Оцінка узгодженості результатів класифікації з експериментальними даними ДВР	56
РОЗДІЛ 5. СПЕЦІАЛЬНА ЧАСТИНА	58
5.1 Практична реалізація класифікації типу навантаження.....	58

5.2 Перевірка надійності та стійкості системи у різних режимах роботи	63
РОЗДІЛ 6. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ	64
6.1 Характеристика шкідливих факторів виробничого середовища .	64
6.2 Забезпечення безпеки життєдіяльності при роботі з ПК	66
ВИСНОВКИ	72
ПЕРЕЛІК ДЖЕРЕЛ.....	74

ДОДАТКИ

ВСТУП

Актуальність теми. З розвитком індустрії та зростанням рівня автоматизації виробництва комп'ютеризовані системи керування технологічними процесами відіграють важливу роль у забезпеченні ефективності, надійності та безпеки технологічних операцій. Одним із головних викликів таких систем є обробка та аналіз великих обсягів даних, що генеруються під час роботи обладнання. Зокрема, у процесах, де матеріали зазнають втомного навантаження, критично важливо визначати типи навантаження та їхній вплив на втомне руйнування для забезпечення довговічності конструкцій і запобігання аварійним ситуаціям.

Методи машинного навчання є потужним інструментом для вирішення задач аналізу та класифікації даних у комп'ютеризованих системах керування технологічними процесами. Завдяки здатності моделювати складні залежності між параметрами та обробляти великі масиви інформації, методи машинного навчання можуть значно підвищити точність і швидкість ухвалення рішень. Це дозволяє автоматизувати моніторинг стану системи, визначати критичні стани обладнання та класифікувати типи навантажень, що особливо важливо для матеріалознавства та діагностики процесів втомного руйнування.

Мета і задачі дослідження: метою дослідження є розробка та впровадження методів машинного навчання для класифікації типів навантажень у комп'ютеризованих системах керування технологічними процесами на основі даних діаграм втомного руйнування.

Об'єкт, методи та джерела дослідження. Об'єктом дослідження є важливі елементи конструкцій.

Практичне значення одержаних результатів. Практичне значення магістерської роботи полягає у розробці та впровадженні методів машинного навчання для автоматичної класифікації типу навантаження на основі діаграм втомного руйнування в комп'ютеризованих системах керування технологічними процесами. Результати роботи спрямовані на підвищення ефективності

управління, моніторингу стану систем та забезпечення надійності конструкцій у промислових середовищах. Зокрема, запропоновані методи нейронних мереж, випадкових лісів та підсилених дерев дозволяють автоматизувати процес класифікації типів навантаження на основі коефіцієнта асиметрії циклу R . А також, розроблені моделі машинного навчання дозволяють проводити моніторинг стану матеріалів та визначати їхню поведінку за умов різних циклічних навантажень.

Завдяки впровадженню алгоритмів машинного навчання у комп'ютеризовані системи керування технологічними процесами стало можливим автоматизувати процес прийняття рішень щодо керування параметрами технологічного процесу, оптимізувати робочі режими для зниження навантажень, які призводять до втомного руйнування конструкцій, а також підвищити ефективність управління завдяки швидкому виявленню та класифікації типу навантаження.

Отже, практичне значення магістерської роботи полягає у створенні ефективного інструменту для класифікації типу навантаження за допомогою алгоритмів машинного навчання. Розроблені моделі можуть інтегруватися у комп'ютеризовані системи керування технологічними процесами для підвищення надійності обладнання, оптимізації процесів та автоматизації контролю втомного руйнування матеріалів. Результати роботи мають широкий спектр застосування у промисловості та забезпечують економічний ефект шляхом підвищення продуктивності та зниження витрат на технічне обслуговування.

РОЗДІЛ 1. АНАЛІТИЧНА ЧАСТИНА

1.1 Огляд комп'ютеризованих систем керування технологічними процесами

Комп'ютеризовані системи керування технологічними процесами (КСКТП) є комплексом апаратно-програмних засобів, призначених для автоматизації управління виробничими процесами з мінімальним втручанням людини. Основним завданням КСКТП є підтримка стабільної та ефективної роботи технологічного обладнання через моніторинг, контроль, аналіз даних та оперативне коригування параметрів технологічного процесу [1].

Типова структура комп'ютеризованих систем керування складається з наступних основних елементів:

- датчики та сенсори – забезпечують збір даних про параметри технологічного процесу, тобто, температура, тиск, вібрація, швидкість (рис.1.1);



Рисунок 1.1 – Базові компоненти системи

- контролери та обчислювальні пристрої – аналізують дані з датчиків, виконують керуючі команди та забезпечують зворотний зв'язок із системою (рис.1.2);

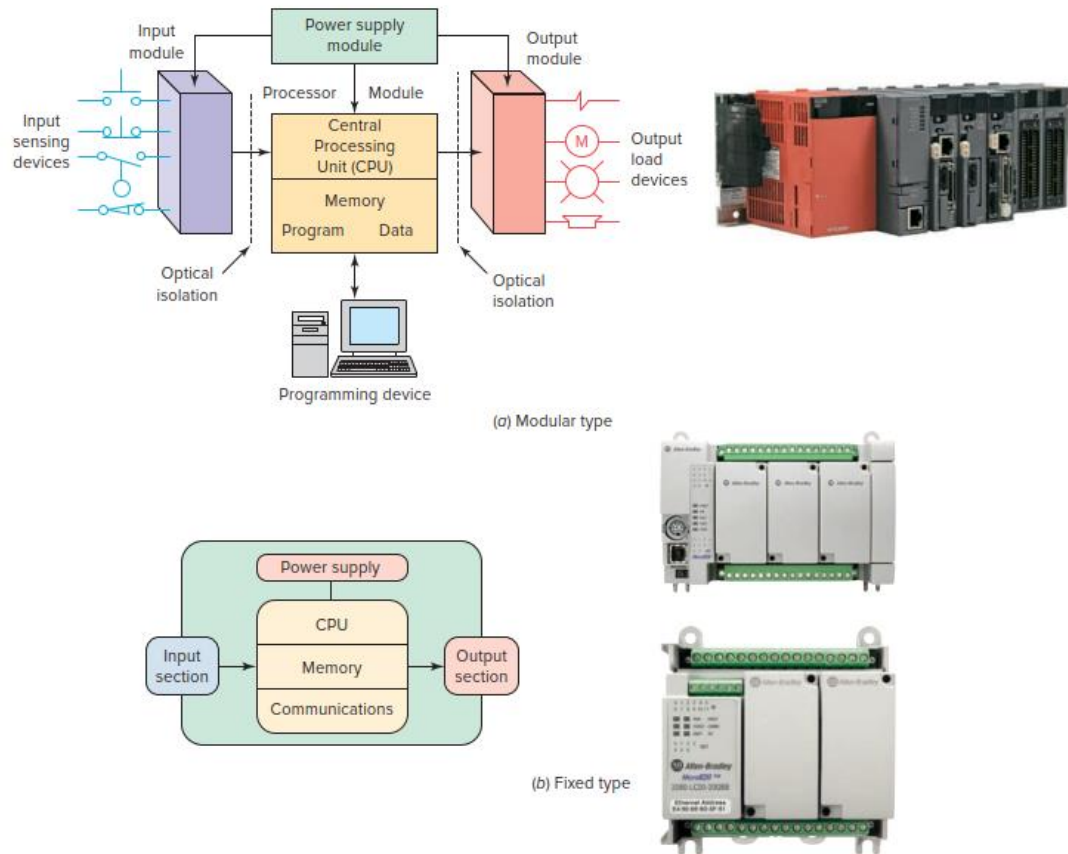


Рисунок 1.2 – Типові частини програмованого логічного контролера.
(а) зображення надано Mitsubishi Automation; (б) зображення надано Rockwell Automation [1]

- апаратне забезпечення – включає комп'ютери, сервери, мікроконтролери та інші пристрої для обчислень і управління;
- програмне забезпечення – реалізує алгоритми керування, моніторингу та аналізу даних. Сюди відносяться системи SCADA, HMI, MES та алгоритми машинного навчання.
- комунікаційна система – забезпечує передачу даних між датчиками, контролерами та центральними обчислювальними системами за допомогою протоколів передачі даних (Modbus, Profibus, OPC,

Ethernet). Внутрішньозаводське обладнання, як правило, з'єднане між собою через локальну мережу (LAN), використовуючи Ethernet іншу високошвидкісну систему зв'язку. Деякі системи SCADA можуть виходити за межі фізичної будівлі на віддалені об'єкти. Ці об'єкти також потребують певної форми зв'язку з головним об'єктом. Топологія шини, як показано на рисунку 1.3, складається з апаратно-програмного з'єднання між усіма вузлами системи;

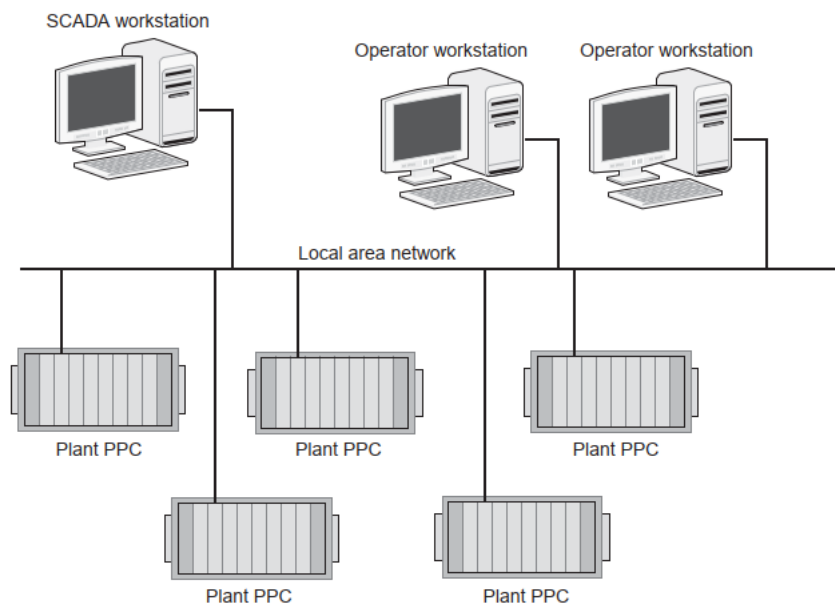


Рисунок 1.3 – З'єднання між усіма вузлами системи [2]

- інтерфейси користувача – дозволяють операторам моніторити та контролювати процеси через графічні візуалізації, сигнали тривоги та звіти.

Комп'ютеризовані системи керування виконують такі функції як збір, обробка та відображення даних про стан технологічного обладнання у реальному часі, а також управління технологічними параметрами для забезпечення стабільності та оптимальності процесів, обробку великих обсягів даних для виявлення трендів, аномалій та оптимізації процесів, передбачення відхилень від норми та діагностика причин несправностей на основі аналізу даних, оперативне коригування параметрів на основі даних, отриманих із сенсорів, для усунення

небажаних відхилень та виявлення критичних станів та аварійних ситуацій для запобігання несправностям і небезпечним умовам роботи.

Загалом, попри значні переваги, сучасні комп'ютеризовані системи керування технологічними процесами стикаються з такими викликами:

- збільшення обсягів даних, у технологічних процесах генерується величезна кількість даних, які потребують ефективної обробки;
- виявлення аномалій та нестандартних ситуацій, традиційні методи аналізу не завжди дозволяють своєчасно ідентифікувати відхилення та потенційні загрози;
- оптимізація управління, відсутність інтелектуальних систем, які адаптуються до зміни умов роботи процесів у реальному часі;
- безпека даних, потреба у захисті інформації від несанкціонованого доступу та кібератак.

Методи машинного навчання, зокрема, нейронні мережі, випадкові ліси та підсилені дерева є важливим інструментом для підвищення ефективності КСКТП завдяки аналізу даних для виявлення відхилень і прийняття рішень, прогнозуванню стану обладнання та виявлення критичних ситуацій, класифікації параметрів процесу, таких як тип навантаження, умови експлуатації та фактори, що впливають на поведінку матеріалів та оптимізації управління шляхом адаптивного налаштування параметрів процесу на основі отриманих даних.

Отже, перспективність дослідження полягає у використанні методів нейронних мереж, випадкових лісів та підсилених дерев для автоматичної класифікації типу навантаження, що дозволить підвищити надійність та безпеку роботи технологічного обладнання.

1.2 Основні задачі класифікації даних у технологічних процесах

Класифікація даних у КСКТП передбачає розподіл об'єктів або станів системи на певні категорії чи класи на основі аналізу вхідних параметрів. Ці параметри можуть включати фізичні величини (температура, тиск, швидкість росту тріщини тощо), технологічні змінні (продуктивність, час роботи, коефіцієнт навантаження) та інші дані, отримані від датчиків і систем збору інформації. Тому основними цілями класифікації є:

- виявлення нормального чи аномального стану системи;
- ідентифікація типу режиму роботи технологічного процесу;
- контроль відповідності параметрів якості продукції заданим стандартам;
- прогнозування стану системи на основі класифікації історичних даних.

У цій роботі основний акцент зроблено на задачі класифікації типу навантаження (коефіцієнта асиметрії R) для аналізу стану матеріалів, що зазнають втомного руйнування. Завдяки класифікації за типом навантаження можливо визначити ризик розвитку втомного руйнування та спрогнозувати поведінку матеріалу під різними умовами експлуатації чи автоматизувати моніторинг стану конструкцій на основі даних із датчиків та експериментальних характеристик матеріалу.

Викликами задач класифікації є:

- великий обсяг даних, тобто, зростання обсягу даних у технологічних процесах вимагає ефективних алгоритмів обробки;
- висока залежність від параметрів, зокрема, точність класифікації значно залежить від якості підготовки даних, вибору релевантних ознак та налаштувань моделей;
- складність взаємозв'язків, а саме, нелінійні залежності між параметрами ускладнюють аналіз і класифікацію за допомогою традиційних методів.

Застосування методів машинного навчання дає змогу ефективно вирішувати задачі класифікації даних у технологічних процесах завдяки моделюванню складних нелінійних залежностей між параметрами процесу, адаптації до великих обсягів даних з високою точністю класифікації та можливості роботи у реальному часі для моніторингу та прогнозування стану системи.

Класифікація даних у технологічних процесах за допомогою машинного навчання дозволяє автоматизувати контроль якості, прогнозування відмов, оптимізацію процесів, моніторинг та керування в реальному часі, виявлення аномалій та підтримку прийняття рішень. Це дозволяє підвищити точність, швидкість та автоматизувати прийняття рішень у виробничих системах.

1.3 Сучасні методи машинного навчання для задач класифікації

Сучасні методи машинного навчання є потужним інструментом для вирішення задач класифікації в різних сферах, зокрема у комп'ютеризованих системах керування технологічними процесами (КСКТП). Класифікація даних є однією з найбільш поширених задач, де застосовують машинне навчання, і це автоматично визначає категорії чи класу для нових спостережень на основі навчальних даних. Загалом, алгоритми машинного навчання використовують різні підходи та методи для досягнення високої точності і ефективності [3-8].

Нейронні мережі є моделлю, що імітує роботу нервової системи людини для обробки та аналізу даних [3]. Вони складаються з багатьох шарів нейронів, кожен з яких виконує певну математичну операцію. Штучні нейронні мережі складаються з входу, прихованих шарів і виходу, де кожен нейрон передає сигнал до наступного після виконання нелінійної операції (рис.1.4). Зокрема, вони здатні моделювати складні та нелінійні взаємозв'язки між вхідними і вихідними даними. Тому, часто їх застосовують для задач, де є велика кількість

параметрів, що взаємодіють між собою, для прикладу, класифікація типу навантаження на основі даних втомного руйнування.

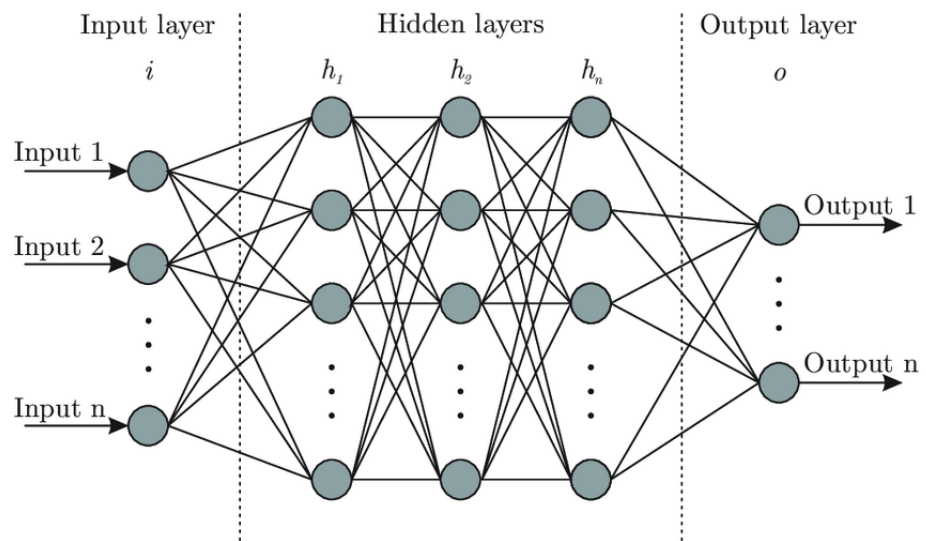


Рисунок 1.4 – Структурна схема нейронної мережі

Їх перевагами є:

- можливість працювати з великими обсягами даних та складними, нелінійними залежностями;
- можливість автоматичного навчання та адаптації моделі до нових даних;
- підходять для задач, де важко вручну формулювати правила для класифікації.

Тоді як недоліками є:

- високі вимоги до обчислювальних потужностей, особливо для глибоких нейронних мереж;
- тривалий процес навчання на великих вибірках;
- моделі є "чорними ящиками", що ускладнює інтерпретацію результатів.

Нейронна мережа обчислює вихідне значення через лінійні та нелінійні трансформації в кожному шарі. Тому, формулою для вихідного нейрона є:

$y = f(\sum_{i=1}^n w_i x_i + b)$ де:

- y – вихід нейрона;
- x_i – вхідні сигнали (вхідні ознаки);
- w_i – ваги нейрона;
- b – зсув (bias);
- f – активаційна функція (наприклад, сигмоїд, ReLU).

Загалом, нейронні мережі добре підходять для задач класифікації складних, багатовимірних даних, таких як аналіз діаграм втомного руйнування матеріалів або виявлення типу навантаження.

Процес класифікації нейронною мережею починається з подачі навчальних даних у модель (рис.1.5). Навчальні дані складаються з вхідних ознак, які відповідають певним класам або категоріям. Алгоритм машинного навчання, використовує ці дані для налаштування своїх параметрів (ваг і зсувів) в процесі навчання. Далі, нейронна мережа приймає вхідні дані, передає їх через кілька шарів (вхідний, приховані, вихідний) і використовує математичні функції для обчислення прогнозу, який є ймовірністю належності кожного класу. На основі цих прогнозів НМ робить рішення про класифікацію. Однак, спочатку ці рішення можуть бути не зовсім точними.

Після отримання результатів класифікації мережа порівнює свої прогнози з реальними мітками (фактичними класами) навчальних даних, оцінюючи точність через функцію втрат (наприклад, середньоквадратичну помилку або крос-ентропію). Якщо прогнози не збігаються з реальними класами, мережа здійснює корекцію своїх параметрів через метод зворотного поширення помилки. Це дозволяє мінімізувати помилки класифікації, поступово покращуючи точність моделі.

Процес перенавчання повторюється протягом кількох етапів (ітерацій), доки не буде досягнуто бажаного рівня точності. Кожна ітерація коригує ваги і зсуви таким чином, щоб нейронна мережа ставала все точнішою в класифікації

нових, невідомих даних. Завдяки цьому нейронна мережа може поступово підвищувати свою здатність до правильного розпізнавання класів і класифікації даних з часом.

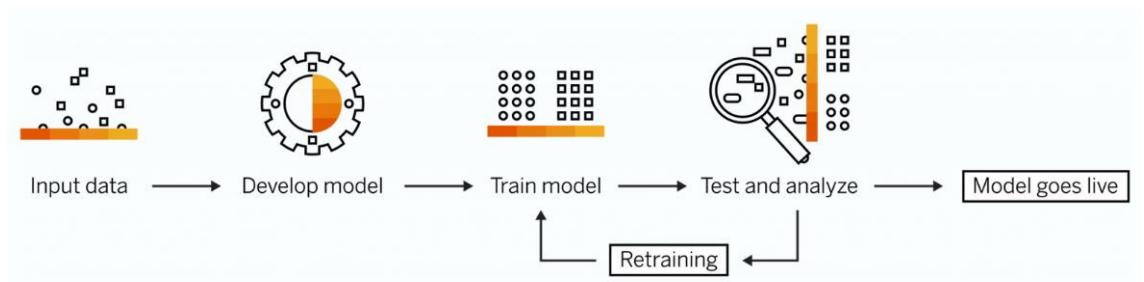


Рисунок 1.5 – Процес навчання нейронної мережі

Важливими є також функції активації, тобто, математичні функції, які використовуються в нейронних мережах для визначення вихідного значення нейрона на основі сумарного входу, що надійшов через ваги. Вони виконують роль фільтра, який допомагає моделювати нелінійні взаємозв'язки в даних. Вибір функції активації має важливе значення для ефективності навчання нейронної мережі та її здатності моделювати складні патерни.

Функція активації визначає, чи буде нейрон активний, тобто чи буде передавати сигнал до наступного шару нейронів. У нейронних мережах використовуються різні типи функцій активації, і кожна з них має свої характеристики та застосування:

- лінійна функція активації є найпростішою, де вихід нейрона є лінійною функцією від суми вхідних значень. Це виглядає так:

$$f(x) = wx + b$$

де:

- w — вага;
 - x — вхід;
 - b — зсув (bias).
- сигмоїд є однією з найпоширеніших функцій активації, яка перетворює вхідне значення в діапазон між 0 і 1, що робить її корисною для задач

класифікації, де ймовірності повинні бути в межах цього діапазону.

Формула для функції сигмоїда:

$$f(x) = \frac{1}{1 + e^{-x}}$$

де e — це число Ейлера, яке приблизно дорівнює 2.718.

- функція активації $\tanh$ схожа на сигмоїд, але її вихід знаходиться в діапазоні від -1 до 1, що робить її більш стабільною, оскільки не має проблеми "згладжування" при великих значеннях.

Формула для $\tanh$:

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$

де e — число Ейлера.

- функція активації ReLU є однією з найбільш популярних функцій активації в сучасних глибоких нейронних мережах. Вона використовує лінійний функціонал для позитивних значень і фіксує нульову активацію для від'ємних значень.

Формула для ReLU:

$$f(x) = \max(0, x)$$

- функція активації Softmax використовується на виході нейронної мережі для багатокласової класифікації. Вона перетворює вхідні значення в ймовірності, що всі сума яких дорівнює 1.

Формула для Softmax:

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}}$$

де:

- x_i — вхідні значення для кожного класу;
- K — кількість класів.

Загалом, вибір функції активації є критичним для ефективної роботи нейронної мережі. Ось кілька важливих висновків:

- сигмоїд та $\tanh$ зазвичай використовуються в простих моделях і для задач з малими даними, але можуть бути чутливими до проблеми "vanishing gradient";
- ReLU є одним із найефективніших варіантів для глибоких нейронних мереж завдяки швидкому навчанні та уникненню проблеми згасаючого градієнта;
- Softmax використовується для багатокласової класифікації, перетворюючи вихідні значення в ймовірності, що дає можливість інтерпретувати результат як ймовірність належності до кожного класу.

Отже, нейронні мережі є гнучкими та потужними інструментами, котрі здатні моделювати складні взаємозв'язки в даних.

Випадкові ліси є ансамблевим методом машинного навчання, який використовує безліч дерев рішень для класифікації [4-5]. Кожне дерево приймає рішення незалежно, і кінцевий результат класифікації визначається шляхом голосування за найбільш ймовірний клас (рис.1.6).

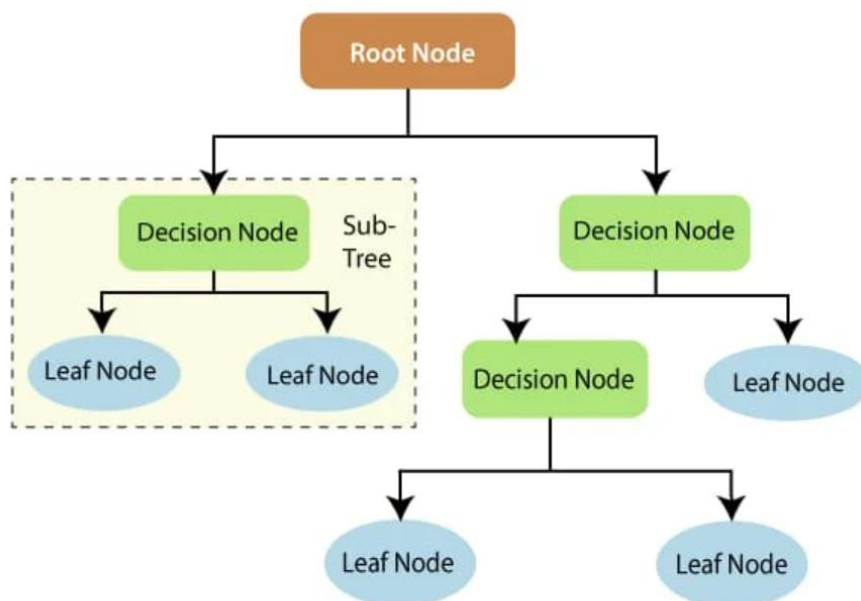


Рисунок 1.6 – Приклад побудови алгоритму випадкових лісів

Зокрема, випадкові ліси використовують метод *bagging*, що полягає в побудові кількох дерев на різних підмножинах даних. Кожне дерево будується на випадковій підмножині ознак, що дозволяє знизити кореляцію між деревами та покращити результати класифікації. Зокрема, рішення на кожному етапі дерево приймає за допомогою простих порогових функцій. Наприклад, для даної ознаки x дерево рішень може використовувати умову:

$$\text{decision} = \begin{cases} \text{class}_1 & \text{if } x_i \leq \theta \\ \text{class}_2 & \text{if } x_i > \theta \end{cases}$$

де:

- x_i — значення ознаки;
- θ — порогове значення;
- $\text{class}_1, \text{class}_2$ — класи, що визначаються деревом.

Випадкові ліси добре підходять для класифікації типів навантаження, де кожне дерево оцінює входні параметри R , а результат класифікації визначається за допомогою голосування між деревами. Їх перевагами є стійкість до перенавчання та легкість в інтерпретації важливості ознак. Тоді як недоліками, високі вимоги до обчислювальних ресурсів для великих наборів даних, а також моделі можуть бути громіздкими.

Загалом, випадкові ліси ідеально підходять для задач класифікації з великими наборами даних та багатьма ознаками, де інтерпретація важливості ознак є важливою.

Методи підсилених дерев побудовані на основі ідеї адаптивного навчання моделей [6], де кожен новий етап фокусується на виправленні помилок попередніх (рис.1.7).

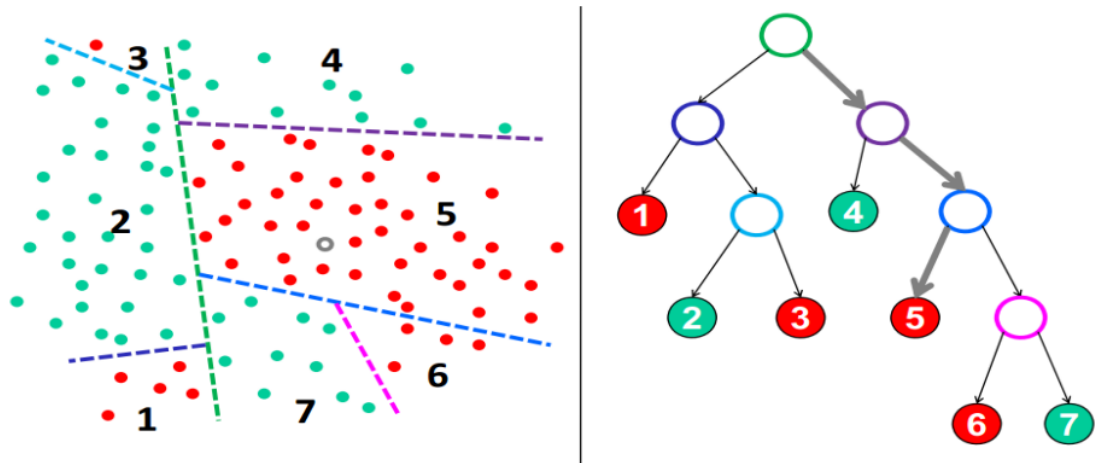


Рисунок 1.7 – Приклад побудови алгоритму підсилених дерев

Зокрема, кожне нове дерево будується на помилках попереднього, при цьому кожен новий етап покращує точність класифікації за рахунок виправлення помилок попередніх дерев. Підсилені дерева можуть працювати з різними типами даних і моделями помилок. Формулою для даного методу є:

Нехай $F_t(x)$ — прогноз на етапі t , тоді новий прогноз на етапі $t + 1$ виглядає так:

$$F_{t+1}(x) = F_t(x) + \eta \cdot h_t(x)$$

де:

- $F_t(x)$ — попередній прогноз;
- $h_t(x)$ — нове дерево, яке додається;
- η — коефіцієнт навчання (learning rate).

Їх перевагами є висока точність при малих обсягах даних, вони добре працюють навіть з неструктурованими даними та мають сильну здатність до адаптації і підвищення точності. Тоді як недоліками є:

- чутливі до шуму в даних, що може призвести до перенавчання;
- потрібно ретельно налаштовувати гіперпараметри моделі (наприклад, кількість дерев, швидкість навчання).

Загалом, підсилені дерева застосовуються для задач, де потрібно досягти високої точності класифікації та де є важливість контролю за помилками в

складних даних. У табл.1 показано порівняння різних методів машинного навчання, їх переваги та недоліки.

Таблиця 1. Порівняння різних методів машинного навчання

Метод	Переваги	Недоліки	Застосування
Нейронні мережі	Підходять для складних задач, хороші для великих даних	Високі вимоги до обчислювальних ресурсів, "чорний ящик"	Класифікація складних даних, виявлення патернів
Випадкові ліси	Стійкість до переобучення, інтерпретованість	Висока обчислювальна складність, громіздкість	Класифікація з великими наборами даних, вибір важливих ознак
Підсилені дерева	Висока точність, адаптивність	Чутливість до шуму, потребують налаштування гіперп	

Порівняння методів машинного навчання для задач класифікації виявило їхні сильні та слабкі сторони, що дозволяє вибирати найбільш підходящий метод в залежності від конкретних вимог та типу даних. Зокрема, нейронні мережі є дуже потужним інструментом для обробки складних, нелінійних залежностей у великих наборах даних. Вони добре справляються з багатими на інформацію, високорозмірними даними, але їхні потреби у обчислювальних ресурсах і складність інтерпретації результатів можуть бути обмеженням у певних випадках. Тоді як випадкові ліси надають стабільні й точні результати, особливо при наявності великої кількості вхідних ознак і шуму в даних. Вони стійкі до перенавчання і можуть надавати інтерпретовані результати, проте їх обчислювальні вимоги зростають при збільшенні кількості дерев. Зокрема, підсилені дерева демонструють високу точність у класифікації, особливо при малих наборах даних. Вони добре адаптуються до змін у даних, але потребують уважного налаштування гіперпараметрів і можуть бути чутливими до шуму.

Загалом, вибір методу класифікації залежить від конкретної задачі: для складних і великих даних ідеальним вибором є нейронні мережі, для великих наборів даних з високою кількістю ознак - випадкові ліси, а для високоточних

класифікацій при малих даних - підсилені дерева. Врахування специфіки завдання дозволяє досягти оптимальних результатів при мінімальних витратах ресурсів.

1.4 Особливості застосування нейронних мереж, випадкових лісів, підсилених дерев

Методи машинного навчання, зокрема нейронні мережі, випадкові ліси та підсилені дерева, є потужними інструментами для вирішення задач класифікації даних у комп'ютеризованих системах керування технологічними процесами. Кожен із цих методів має свої особливості, переваги та обмеження, що визначають їх доцільність у конкретних задачах.

Нейронні мережі є потужним інструментом для класифікації складних і нелінійних залежностей у даних. Вони активно використовуються для задач, де існує велика кількість ознак, взаємозв'язки між якими є складними. Нейронні мережі є ідеальним інструментом для задач класифікації з нелінійними залежностями, оскільки вони можуть моделювати такі зв'язки через багатошарову структуру з нелінійними функціями активації. А також, мережі добре справляються з великими наборами даних і можуть автоматично визначати важливі ознаки, не потребуючи ручного вибору характеристик. Нейронні мережі здатні виявляти складні патерни в даних, що можуть бути неочевидними для традиційних методів аналізу.

Загалом, нейронні мережі часто застосовуються для задач класифікації в технічних та інженерних системах, таких як прогнозування втомного руйнування матеріалів, де потрібно виявляти складні взаємозв'язки між різними параметрами.

Випадкові ліси є ансамблевим методом, що використовує кілька дерев рішень для класифікації. Кожне дерево рішень будується на випадкових підвибірках даних і ознак, що дозволяє створити модель, яка є більш стійкою до

шуму та перенавчання, ніж одне дерево рішень. Випадкові ліси дуже стійкі до шуму в даних, оскільки використовують багато дерев, що дозволяє зменшити вплив випадкових помилок або аномалій у даних. Зокрема, завдяки паралельній обробці дерев, випадкові ліси можуть ефективно працювати з великими обсягами даних. Випадкові ліси дозволяють легко оцінити важливість ознак для класифікації, що є важливою особливістю при роботі з інженерними та науковими задачами.

Випадкові ліси можуть бути ефективними для задач класифікації, таких як виявлення типу навантаження або прогнозування результатів втомного руйнування матеріалів, де важливо точно оцінити вплив різних параметрів.

Підсилені дерева використовують ітераційний підхід, де кожне нове дерево будується для виправлення помилок попередніх. Зокрема, кожне дерево будується таким чином, щоб виправити помилки попередніх дерев, що дозволяє поступово покращувати точність моделі. А також, підсилені дерева можуть ефективно адаптуватися до змін у даних, оскільки кожен етап навчання зосереджується на виправленні попередніх помилок та дають дуже високі результати в задачах класифікації завдяки своїй здатності знижувати похибки.

Цей метод широко застосовується в задачах класифікації, де важлива висока точність, наприклад, для класифікації типу навантаження на основі параметрів втомного руйнування, де кожне нове дерево намагається коригувати помилки попередніх.

РОЗДІЛ 2. ТЕХНОЛОГІЧНА ЧАСТИНА

2.1 Огляд характеристик діаграм втомного руйнування алюмінієвого сплаву

Втомне руйнування є процесом накопичення мікропошкоджень у матеріалі під впливом циклічних навантажень, що врешті-решт призводить до утворення тріщин і руйнування. Цей процес є одним із найбільш поширених причин виходу з ладу конструкційних елементів у багатьох галузях, таких як авіація, автомобільна промисловість, енергетика та будівництво [9-11].

Термін служби важливих елементів конструкцій, які піддаються втомі, залежить від періоду росту макротріщини, від її зародження до критичного розміру. Графічне зображення вказаної залежності називається діаграмою втомного руйнування, яка визначає характеристики циклічної тріщиностійкості матеріалів та відображає залежність швидкості росту тріщини da/dN від коефіцієнта інтенсивності напружень K при різних значеннях коефіцієнта асиметрії циклу R .

Термін служби елемента конструкції є кількістю циклів до його руйнування. Довжина тріщини a залежить від відповідної кількості циклів навантаження N на момент її вимірювання (рис. 2.1). Тобто в експериментальних дослідженнях швидкість РВТ вимірюється послідовно через певні проміжки часу за довжиною тріщини a при відповідних значеннях циклів навантаження N , а їх результати містять інформацію щодо руйнування зразка з тріщиною у вигляді залежності $a-N$.

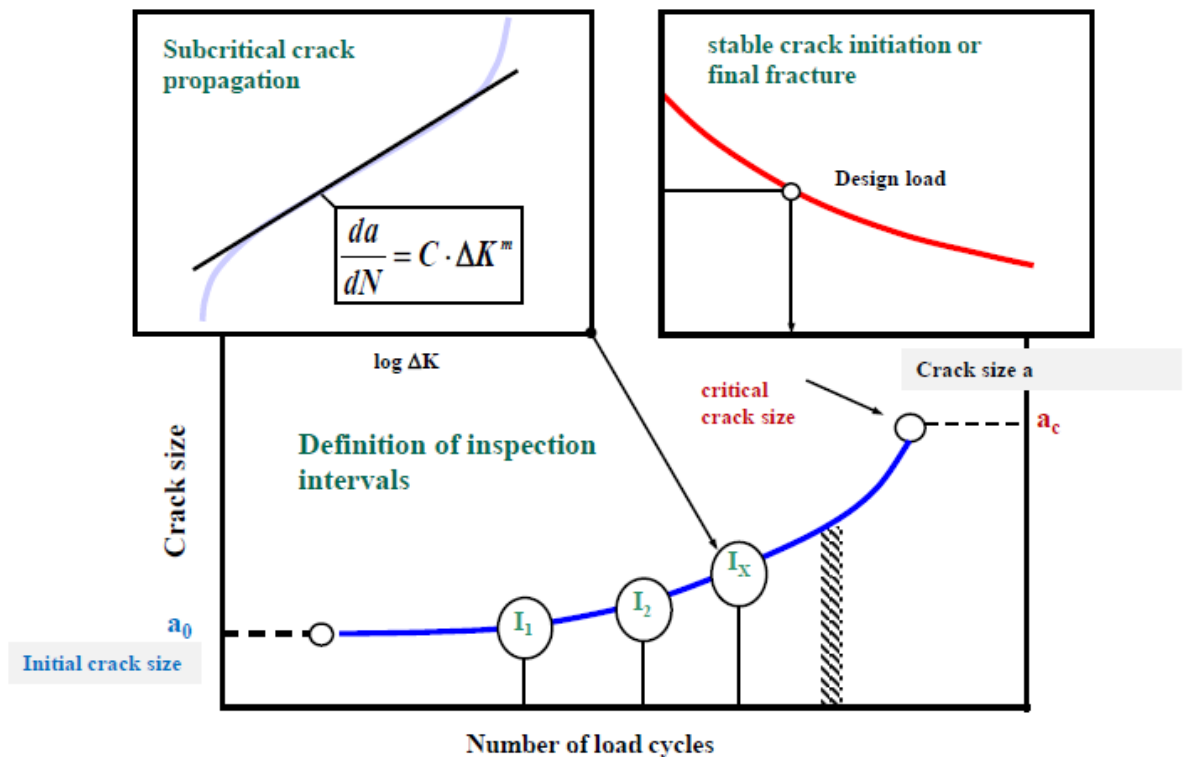


Рисунок 2.1 – Графік залежності довжини тріщини від кількості циклів навантаження [9]

Діаграми втомного руйнування описуються такими основними параметрами:

- швидкість росту тріщини da/dN , це кількість збільшення довжини тріщини da за один цикл навантаження dN , залежить від інтенсивності напружень K та умов навантаження R ;
- коефіцієнт інтенсивності напружень K , що відображає величину напружень навколо тріщини, що залежить від розмірів тріщини, геометрії конструкції та прикладеного навантаження:

$$K = \sigma \sqrt{\pi a}$$

де:

- σ — номінальне напруження;
- a — довжина тріщини.

Значення K поділяється на мінімальне ($K_{\min}$), максимальне ($K_{\max}$) та амплітуду $\Delta K = K_{\max} - K_{\min}$

- коефіцієнт асиметрії циклу R , це співвідношення між мінімальним і максимальним значеннями коефіцієнта інтенсивності напружень:

$$R = \frac{K_{\min}}{K_{\max}}$$

Відображає тип циклічного навантаження:

- $R = 0$: повністю розтягуюче навантаження;
- $R > 0$: частково розтягуюче;
- $R < 0$: чергування розтягування та стискання.

Діаграма втомного руйнування зазвичай будується у логарифмічному масштабі, де по горизонтальній осі відкладається ΔK , а по вертикальній осі — da/dN . Вона умовно поділяється на три області:

- низькоамплітудної (I), у цій області ΔK є недостатнім для росту тріщини та швидкість росту тріщини da/dN практично дорівнює нулю;
- середньоамплітудної (II), тут спостерігається стабільна залежність;
- високоамплітудної (III), при досягненні критичного значення ΔK_c тріщина починає рости неконтрольовано, що призводить до руйнування.

На рис. 2.2 показано залежність між навантаженням та коефіцієнтом інтенсивності циклічних напружень.

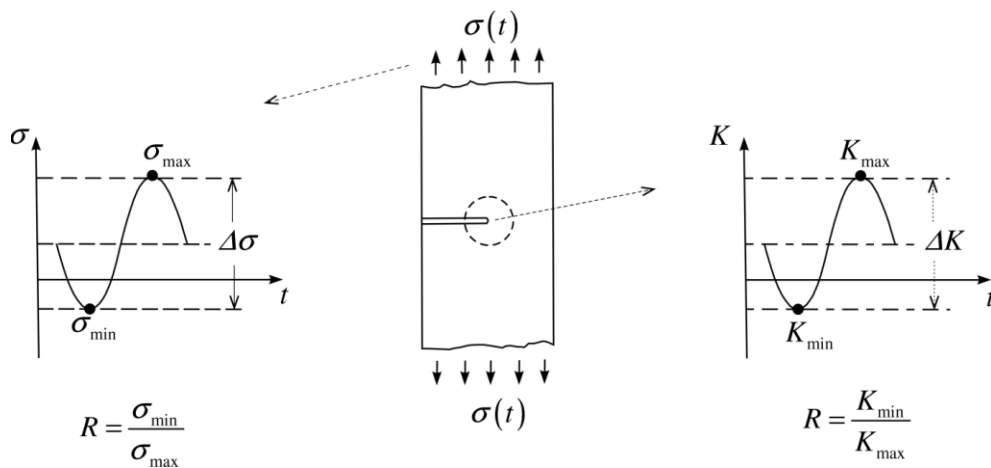


Рисунок 2.2 – Зв'язок між навантаженням і коефіцієнтом інтенсивності напружень [10]

Алюмінієві сплави, зокрема сплав Д16Т, є широко використовуваними матеріалами в авіаційній, автомобільній та інших галузях завдяки їхній високій міцності та легкості. Однак ці сплави мають обмежену втомну довговічність, що вимагає детального аналізу процесів їх руйнування.

Особливості втомного руйнування алюмінієвих сплавів [12-19]:

- чутливість до коефіцієнта R , тобто, зміна R значно впливає на швидкість росту тріщини da/dN , при високих значеннях R матеріал демонструє вищу швидкість росту тріщини, що вказує на зниження втомної довговічності.
- вплив мікроструктури, а саме, наявність зерен, мікропор і включень у структурі алюмінієвих сплавів суттєво впливає на параметри C та m , що визначають залежність da/dN від ΔK ;
- температурний фактор, зокрема, алюмінієві сплави є чутливими до температури: при підвищених температурах знижується їхня втомна міцність.

У рамках задачі класифікації діаграми втомного руйнування дозволяють розподіляти режими навантаження за коефіцієнтом R на певні класи. Наприклад:

- Клас 1: $R=0.2$ (низьке навантаження).

- Клас 2: $R=0.4$ (середнє навантаження).
- Клас 3: $R=0.6$ (високе навантаження).

Моделі МН, такі як нейронні мережі, випадкові ліси та підсилені дерева, можуть використовувати параметри da/dN , ΔK та R для класифікації типів навантаження.

ДВР є важливим інструментом для аналізу втомних характеристик матеріалів, таких як алюмінієві сплави. Основні параметри цих діаграм, зокрема da/dN , ΔK та R , визначають поведінку матеріалів під циклічними навантаженнями та можуть бути використані для класифікації режимів навантаження. Використання таких діаграм у поєднанні з методами машинного навчання дозволяє автоматизувати процес оцінки стану матеріалу, що підвищує точність і швидкість аналізу у виробничих та дослідницьких задачах.

2.2 Постановка задачі класифікації типу навантаження за коефіцієнтом асиметрії R

Коефіцієнт асиметрії циклу R є важливим параметром, що характеризує тип циклічного навантаження на матеріал. Залежно від значення R , матеріал може перебувати в умовах:

- повного розтягування ($R = 0$),
- часткового розтягування ($R > 0$),
- чергування розтягування і стискання ($R < 0$).

Класифікація типу навантаження за R має критичне значення у задачах прогнозування втомного руйнування матеріалів, таких як алюмінієві сплави, адже різні режими навантаження по-різному впливають на швидкість росту тріщини da/dN та на втомну довговічність конструкції.

Загалом, взаємозв'язки між da/dN , ΔK та R є нелінійними, тому традиційні статистичні методи часто не можуть ефективно розв'язати цю задачу. Методи МН, зокрема нейронні мережі, випадкові ліси та підсилені дерева, здатні

враховувати ці складні залежності. Зокрема, дані ДВР можуть містити шум (похибки вимірювань, неоднорідність матеріалу), що вимагає використання моделей, стійких до таких збурень, наприклад, випадкових лісів. Тому, для задач втомного руйнування дані часто мають обмежену кількість спостережень, що потребує оптимального використання існуючих даних, зокрема через техніки балансування вибірок і перехресної перевірки.

Для розв'язання задачі класифікації типу навантаження використовуються сучасні методи машинного навчання. Основні етапи методології такі:

- попередня обробка даних, а саме нормалізація даних da/dN , ΔK та R для приведення до однакового масштабу, визначення кореляції між вхідними параметрами для вибору найбільш релевантних ознак, видалення шумів та пропусків у даних;
- розділення вибірки, тобто, розподіл даних на навчальну вибірку (70%) і тестову вибірку (30%) для перевірки точності моделі;
- вибір моделей МН, зокрема, нейронні мережі для виявлення складних нелінійних залежностей між параметрами, випадкові ліси для стійкої класифікації та інтерпретації важливості ознак, підсилені дерева для досягнення високої точності класифікації;
- навчання моделей, зокрема, налаштування гіперпараметрів для кожної моделі (наприклад, кількість шарів і нейронів у нейронних мережах, кількість дерев у випадкових лісах і підсилених деревах);
- оцінка якості моделей, а саме, використання метрик точності (Accuracy), повноти (Recall), точності класифікації (Precision) та F1-міри для оцінки результатів;
- інтерпретація результатів, тобто, аналіз важливості параметрів da/dN , ΔK та R для класифікації та вибір найефективнішої моделі для впровадження.

Очікувані результати такі:

- класифікація типу навантаження, де точність моделі для класифікації типу навантаження повинна бути вище 90%, що забезпечить ефективність у реальних виробничих умовах.
- важливість ознак, де визначення, які з параметрів da/dN , ΔK та R мають найбільший вплив на точність класифікації;
- розробка автоматизованої системи, результати класифікації повинні бути інтегровані у комп'ютеризовану систему керування технологічними процесами для автоматичного моніторингу стану матеріалів.

Отже, постановка задачі класифікації типу навантаження за коефіцієнтом асиметрії R базується на аналізі даних діаграм втомного руйнування. Застосування методів машинного навчання дозволяє автоматизувати процес класифікації, підвищити точність аналізу та забезпечити своєчасне прогнозування стану матеріалів у технологічних процесах. Ця задача є ключовою для забезпечення надійності конструкцій та зниження ризику втомного руйнування в критичних умовах експлуатації.

2.3 Опис вхідних даних: параметри da/dN , K та R

Параметри da/dN , ΔK та R є основними вхідними ознаками для задачі класифікації типу навантаження, які відображають фізичні характеристики процесу втомного руйнування матеріалів. Ці параметри отримуються на основі експериментальних досліджень діаграм втомного руйнування алюмінієвого сплаву і характеризують:

- швидкість росту тріщини da/dN є швидкістю, з якою матеріал руйнується під впливом циклічного навантаження;
- коефіцієнт інтенсивності напружень ΔK характеризує міру напружень у матеріалі поблизу тріщини;

- коефіцієнт асиметрії циклу R є характеристикою типу циклічного навантаження.

Загалом, da/dN відображає поведінку матеріалу під різними типами навантаження R і може мати суттєві відмінності залежно від режиму роботи:

- при малих значеннях R ($R=0.2$) швидкість росту тріщини є нижчою;
- при високих значеннях R ($R=0.6$) da/dN збільшується.

Зокрема, ΔK має прямий вплив на швидкість росту тріщини da/dN і визначає режим навантаження. Наприклад:

- за низьких значень ΔK тріщина практично не зростає (режим порогового навантаження);
- за високих значень ΔK спостерігається нестабільний ріст тріщини.

А також, R безпосередньо впливає на швидкість росту тріщини da/dN :

- за низьких значень R тріщина розвивається повільніше через зниження напружень;
- за високих значень R швидкість росту тріщини збільшується через більший вплив розтягуючих напружень;

Для класифікації типу навантаження R виступає **цільовим параметром**, який необхідно передбачити за вхідними ознаками da/dN і ΔK .

Параметри da/dN , ΔK та R є фундаментальними характеристиками діаграм втомного руйнування алюмінієвих сплавів. Вхідні дані da/dN , ΔK та R дозволяють описати поведінку матеріалу під циклічними навантаженнями, тоді як R є цільовим параметром для класифікації. Нелінійний взаємозв'язок між цими параметрами робить задачі класифікації складними, але застосування методів МН, таких як нейронні мережі, випадкові ліси та підсилені дерева, дає змогу ефективно вирішувати ці задачі.

2.4 Попередня обробка даних: нормалізація, масштабування, очистка

Попередня обробка даних є важливим етапом у розв'язанні задач класифікації, оскільки якість вхідних даних суттєво впливає на точність і стабільність роботи алгоритмів МН. Зокрема, у задачі класифікації типу навантаження за коефіцієнтом асиметрії R , необхідно забезпечити коректну обробку таких параметрів, як da/dN , ΔK та R та інших, які можуть мати різні масштаби, рівень шуму і пропуски. Основні етапи попередньої обробки включають:

1. Нормалізацію даних;
2. Масштабування даних;
3. Очищення даних (очистка від шуму та пропусків).

Нормалізація — це перетворення даних у такий спосіб, щоб їх значення лежали в заданому діапазоні (зазвичай $[0, 1]$). Це важливо для роботи моделей машинного навчання, чутливих до масштабів вхідних даних, зокрема нейронних мереж. Формула нормалізації (Min-Max Scaling):

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

де:

- x_{norm} — нормалізоване значення,
- x — початкове значення,
- $x_{\min}$, $x_{\max}$ — мінімальне і максимальне значення ознаки відповідно.

Масштабування - це перетворення даних таким чином, щоб вони мали середнє значення 0 і стандартне відхилення 1, зокрема, це особливо корисно для алгоритмів, які використовують метрики відстані (наприклад, у випадкових лісах або підсилених деревах). Формула масштабування (Z-Score Scaling):

$$x_{\text{scaled}} = \frac{x - \mu}{\sigma}$$

де:

- x_{scaled} — масштабоване значення,
- μ — середнє значення ознаки,
- σ — стандартне відхилення ознаки.

Різниця між нормалізацією і масштабуванням, є те, що нормалізація перетворює дані у фіксований діапазон $[0, 1]$, а масштабування усуває вплив масштабу даних через перетворення до стандартного розподілу (з нульовим середнім і одиничним стандартним відхиленням).

Очищення даних необхідне для видалення шуму, пропусків і аномальних значень, які можуть негативно вплинути на результати моделі. Основними методами очищення є:

- робота з пропущеними значеннями, тобто, видалення спостережень із пропусками. Застосовується, якщо кількість пропусків у даних є незначною (менше 5%);
- заповнення пропусків. Зокрема, середнє значення (mean) або медіана (median) використовується для числових даних. Логічне інтерполювання для тимчасових рядів (наприклад, дані про ΔK).

Нормалізація та масштабування забезпечують однакову важливість усіх вхідних у процесі навчання. А також, покращують збіжність моделей, особливо для нейронних мереж, де великі значення можуть впливати на градієнт.

Очищення даних стосується видалення шуму і викидів, підвищує точність моделей машинного навчання. Робота з пропусками дозволяє зберегти цінну інформацію, уникаючи втрати спостережень.

Зокрема, попередня обробка дозволяє моделі краще узагальнювати закономірності у даних і забезпечує високу точність класифікації типу навантаження R .

Попередня обробка даних є необхідним етапом для забезпечення коректної роботи алгоритмів машинного навчання. Нормалізація та масштабування усувають проблему різних діапазонів значень ознак, очищення від шуму та пропусків зменшує ймовірність погіршення точності моделі. Завдяки якісній обробці даних задачі класифікації типу навантаження за коефіцієнтом асиметрії R можуть бути вирішені ефективніше, що забезпечує стабільність і надійність прогнозування.

2.5 Вибір метрик для оцінки якості класифікації

Оцінка якості роботи моделей машинного навчання є критично важливим етапом у процесі побудови класифікаційної моделі. Правильний вибір метрик для оцінювання дозволяє визначити, наскільки ефективно модель виконує поставлену задачу, а також виявити її сильні та слабкі сторони. У задачі класифікації типу навантаження за коефіцієнтом R використовуються декілька метрик для комплексної оцінки результатів:

- точність класифікації (Accuracy):

$$\text{Accuracy} = \frac{\text{Кількість правильних передбачень}}{\text{Загальна кількість спостережень}}$$

Ця метрика показує частку правильно класифікованих спостережень від загальної кількості. Використовується як базовий показник ефективності моделі. Підходить для задач із збалансованими класами та не є інформативною у випадках, коли один клас значно переважає інший (незбалансовані дані);

- Повнота (Recall):

$$\text{Recall} = \frac{\text{Кількість правильних передбачень класу}}{\text{Кількість реальних спостережень класу}}$$

Ця метрика визначає здатність моделі знаходити всі об'єкти певного класу. Відображає, скільки елементів певного класу було правильно виявлено з усіх реальних елементів цього класу та є ключовою метрикою для задач, де важливо уникнути пропуску важливих класів (наприклад, при класифікації критичних типів навантаження);

- точність передбачення (Precision):

$$\text{Precision} = \frac{\text{Кількість правильних передбачень класу}}{\text{Кількість передбачень класу}}$$

Показує, наскільки точними є передбачення моделі. Відображає частку коректно класифікованих об'єктів серед усіх об'єктів, які модель віднесла до цього класу. Важлива у випадках, де помилкова класифікація небажана (наприклад, помилкове визначення високого навантаження як низького);

- F1-міра:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-міра є середньою гармонійною між точністю передбачення (Precision) і повнотою (Recall). Вона забезпечує баланс між двома показниками, особливо коли вони знаходяться в конфлікті. А також корисна у випадках, коли дані є незбалансованими та показує загальну ефективність моделі.

Додатково, матриця неточностей дозволяє оцінити роботу моделі для кожного класу окремо. Вона містить наступні елементи (рис.2.3):

- TP (True Positives): правильно класифіковані спостереження даного класу.
- FP (False Positives): кількість спостережень інших класів, помилково віднесених до даного класу.
- FN (False Negatives): кількість спостережень даного класу, помилково віднесених до інших класів.

- TN (True Negatives): кількість спостережень, які правильно не належать до даного класу.

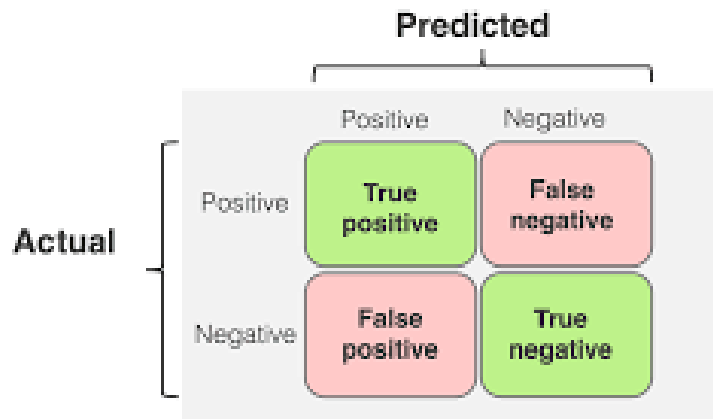


Рисунок 2.3 – Confusion Matrix для класифікації

А також, ROC-крива - це графік, який показує співвідношення між True Positive Rate (Recall) і False Positive Rate для різних порогових значень. Тоді як AUC є значенням площі під ROC-кривою. Чим ближче AUC до 1, тим краще модель виконує класифікацію.

Загалом, для оцінки якості класифікації типу навантаження за коефіцієнтом асиметрії R необхідно використовувати комплексний підхід із застосуванням кількох метрик. Точність (Accuracy) надає загальну оцінку роботи моделі, а повнота (Recall) і точність передбачення (Precision) дозволяють детально проаналізувати якість роботи для кожного класу. F1-міра забезпечує баланс між точністю і повнотою, особливо для незбалансованих даних, а матриця неточностей надає глибокий аналіз помилок. Застосування цих метрик разом із візуалізацією через ROC-криві допомагає забезпечити повноцінну оцінку моделі та її придатність до практичного використання.

РОЗДІЛ 3. КОНСТРУКТОРСЬКА ЧАСТИНА

3.1 Розробка архітектури комп'ютеризованої системи керування з інтеграцією методів МН

Метрики оцінки якості класифікації дозволяють забезпечити комплексний аналіз ефективності моделей, виділяючи їх сильні та слабкі сторони. Для побудови комп'ютеризованої системи керування з інтеграцією методів МН важливо враховувати всі етапи обробки даних, навчання моделей і передачі результатів у систему прийняття рішень. Запропонована архітектура забезпечує автоматизацію класифікації, масштабованість і точність, необхідну для роботи в умовах реальних технологічних процесів. Основні вимоги до архітектури (рис.3.1):

- автоматизація процесів, тобто, система повинна автоматично збирати дані, обробляти їх, класифікувати типи навантажень і генерувати рекомендації;
- гнучкість і масштабованість, а саме, система повинна підтримувати інтеграцію нових алгоритмів МН і обробляти різні типи даних;
- реальний час, тобто, система має забезпечувати аналіз даних у реальному часі для своєчасного реагування на зміни в технологічних процесах;
- інтерпретованість, зокрема, результати класифікації повинні бути зрозумілими для користувачів, з можливістю відстеження ключових факторів, що впливають на рішення моделі.

Компонентами архітектури є:

- модуль збору даних, що отримує дані з датчиків, систем моніторингу чи баз даних. Зокрема, дані включають параметри da/dN , ΔK та R ;
- модуль попередньої обробки даних, що виконує нормалізацію, масштабування, очищення від шумів і пропусків та перетворює дані у формат, зручний для роботи моделей МН;

- модуль навчання та класифікації, що інтегрує алгоритми МН, зокрема, нейронні мережі, випадкові ліси, підсилені дерева та забезпечує навчання моделей на навчальній вибірці та класифікацію вхідних даних;
- модуль оцінки якості моделей, що використовує метрики (Accuracy, Precision, Recall, F1-mіру) для оцінки ефективності моделей та зберігає результати тестування для подальшого аналізу.
- модуль візуалізації показує результати класифікації, ключові метрики, матриці неточностей, графіки ROC-кривих та генерує звіти для аналізу роботи системи.
- модуль інтеграції забезпечує зв'язок між класифікаційною системою і загальною системою керування технологічними процесами та передає результати класифікації в модуль прийняття рішень;
- модуль зберігання даних, що архівує навчальні, тестові і поточні дані для подальшого аналізу і навчання моделей.



Рисунок 3.1 – Схема архітектури системи

3.2 Вибір програмного забезпечення для інтеграції методів класифікації

Ефективна інтеграція методів класифікації у комп'ютеризовану систему керування технологічними процесами залежить від правильно обраного програмного забезпечення (ПЗ). Воно повинно забезпечувати зручність реалізації алгоритмів машинного навчання, підтримку необхідних форматів даних, високу продуктивність і можливість візуалізації результатів. У цьому підрозділі розглянуто критерії вибору ПЗ та рекомендовані інструменти для реалізації задач класифікації типу навантаження.

STATISTICA є програмним забезпеченням для статистичного аналізу, моделювання даних та реалізації алгоритмів машинного навчання. Воно має широкий спектр інструментів, що дозволяє створювати, налаштовувати та оцінювати моделі класифікації для задач у різних галузях, включаючи технічні процеси.

Для задач класифікації типу навантаження за коефіцієнтом асиметрії R , використання STATISTICA є доцільним через її інтеграцію з інженерними даними, зручність у роботі та підтримку алгоритмів машинного навчання, таких як нейронні мережі, випадкові ліси та підсилені дерева. Зокрема, STATISTICA дозволяє створювати класифікаційні моделі на основі:

- нейронних мереж;
- випадкових лісів;
- підсилених дерев;
- інших статистичних і математичних методів.

Інструменти побудови моделей реалізовані через графічний інтерфейс, який спрощує процес розробки навіть для користувачів без досвіду програмування. Зокрема, підтримка великого обсягу даних дозволяє працювати з експериментальними даними діаграм втомного руйнування da/dN , ΔK та R . Додатково, STATISTICA може працювати з базами даних (SQL, CSV) і легко інтегрується в існуючі системи керування. А також, платформа підтримує

автоматизацію побудови, оцінки та валідації моделей класифікації. Тому результати класифікації можуть бути представлені у вигляді графіків, матриць неточностей, діаграм важливості ознак тощо.

Етапами інтеграції методів класифікації в STATISTICA є:

- імпорт даних, зокрема, STATISTICA дозволяє завантажувати дані з різних джерел:
 - локальних файлів (CSV, Excel);
 - баз даних (SQL);
 - інтеграція з іншими системами моніторингу;
 - попередня обробка даних, STATISTICA має інструменти для:
 - нормалізації та масштабування даних;
 - видалення пропусків і шумів;
 - аналізу кореляцій між ознаками;
 - розробка моделі класифікації, використання графічного інтерфейсу для налаштування моделей:
 - нейронні мережі, а саме, створення багатошарових персептронів із налаштуванням кількості шарів і нейронів;
 - випадкові ліси, налаштування кількості дерев, глибини дерев;
 - Підсилені дерева, визначення кількості ітерацій та швидкості навчання;
 - навчання моделі, STATISTICA виконує автоматичне навчання моделей на навчальних даних із можливістю налаштування гіперпараметрів.
 - оцінка якості моделі, інструменти для розрахунку метрик:
 - точність (Accuracy), повнота (Recall), F1-міра;
 - побудова матриць неточностей та ROC-кривих;
- а також, порівняння результатів різних моделей для вибору оптимального підходу;
- візуалізація результатів, STATISTICA дозволяє створювати:

- графіки залежностей між вхідними ознаками та цільовим параметром;
 - діаграми важливості ознак (наприклад, вплив ΔK на класи R).
 - інтерактивні графіки ROC-кривих;
- розгортання моделі, а саме, експорт готової моделі для інтеграції в існуючі системи керування (SCADA, ERP) та автоматизація процесів класифікації через сценарії.

Перевагами STATISTICA є:

- зручність для користувачів без програмування, тобто, інтуїтивний графічний інтерфейс дозволяє створювати класифікаційні моделі без знання мов програмування;
- підтримка сучасних методів класифікації, зокрема, STATISTICA має вбудовані алгоритми для реалізації нейронних мереж, випадкових лісів і підсилених дерев;
- швидкий аналіз, тобто, інструменти автоматичного аналізу дозволяють швидко оцінювати ефективність моделей;
- візуалізація даних і результатів, а саме, потужні засоби візуалізації дозволяють зрозуміти роботу моделей і результати класифікації;
- інтеграція з іншими системами, тобто, легка інтеграція з базами даних, а також експорт моделей для використання у зовнішніх додатках;
- автоматизація, зокрема, можливість налаштування сценаріїв для автоматичної обробки даних і класифікації.

Недоліками STATISTICA є:

- висока вартість ліцензії, тобто, STATISTICA є комерційним продуктом, що може бути недоступним для невеликих організацій або індивідуальних користувачів;
- обмежена гнучкість, у порівнянні з Python, STATISTICA має обмежену можливість кастомізації моделей.

- проблеми з масштабованістю, тобто, для дуже великих наборів даних продуктивність може бути нижчою, ніж у спеціалізованих мов програмування (наприклад, Python).

Загалом, STATISTICA є потужним і зручним інструментом для реалізації задач класифікації, зокрема класифікації типу навантаження за коефіцієнтом R . Завдяки підтримці сучасних алгоритмів машинного навчання, простоті використання та інструментам візуалізації, STATISTICA дозволяє швидко і якісно реалізувати класифікаційні моделі. Однак її обмеження, такі як висока вартість і менша гнучкість у порівнянні з Python, можуть впливати на вибір залежно від масштабу проєкту та доступних ресурсів.

3.3 Забезпечення надійності та продуктивності системи

Комп'ютеризована система керування з інтеграцією методів машинного навчання повинна відповідати високим вимогам до надійності та продуктивності. Це забезпечує її стабільну роботу в умовах реального часу та безперебійне виконання завдань класифікації і керування технологічними процесами.

Надійністю системи є її здатність працювати без відмов у визначених умовах експлуатації. Для підвищення надійності слід враховувати такі аспекти:

- резервування компонентів, а саме, апаратне резервування, використання дублюючих модулів (серверів, сховищ даних, мережевого обладнання) для забезпечення стійкості до збоїв та програмне резервування, дублювання критичних функцій системи для автоматичного перемикавання на резервний компонент у разі збою;
- механізми відновлення після збоїв, тобто, резервне копіювання відповідає за регулярне створення резервних копій баз даних і моделей класифікації, тоді як автоматичне відновлення за можливість відновлення роботи системи з останнього стабільного стану;

- моніторинг системи, тобто, впровадження модулів моніторингу для відстеження стану апаратних і програмних компонентів системи та виявлення потенційних проблем до моменту виникнення критичних збоїв.
- управління помилками, тобто, визначення і обробка виняткових ситуацій у програмному забезпеченні;
- забезпечення безпеки, зокрема, використання авторизації та шифрування для захисту конфіденційних даних та регулярне оновлення системи для усунення вразливостей.

Продуктивність системи визначає її здатність виконувати задачі класифікації та аналізу даних у визначений час, навіть при збільшенні обсягу обчислень або кількості даних. Для підтримання надійності та продуктивності системи необхідно регулярно оцінювати її ефективність за такими показниками як пропускна здатність, тобто, кількість запитів на класифікацію, які система може обробити за одиницю часу, затримка обробки, тобто, середній час від отримання даних до отримання результату класифікації, час безвідмовної роботи, тобто, середній час, протягом якого система функціонує без збоїв та рівень помилок, тобто, частка запитів, які завершилися помилкою, через технічні або програмні збої.

Загалом, забезпечення надійності та продуктивності комп'ютеризованої системи керування є ключовою вимогою для її ефективного функціонування в умовах промислових технологічних процесів. Використання сучасних технічних і програмних рішень, таких як STATISTICA, Python або MATLAB, у поєднанні з потужним апаратним забезпеченням дозволяє створити стабільну, продуктивну та ефективну систему, яка відповідає потребам класифікації та управління в реальному часі.

РОЗДІЛ 4. НАУКОВО-ДОСЛІДНА ЧАСТИНА

4.1 Побудова моделей МН для класифікації

Побудова моделей МН для класифікації є ключовим етапом у створенні системи автоматичного розпізнавання типу навантаження за коефіцієнтом асиметрії R . Цей процес включає вибір алгоритмів, попередню обробку даних, навчання моделей, оптимізацію їх параметрів і оцінку якості класифікації [18].

Під час навчання набір даних було розділено на дві вибірки: навчальну та тестову. База даних містила 225 елементів, з яких 70% були випадковим чином відібрані з усіх експериментальних даних при різних температурах, а 30% були залишені для оцінки якості прогнозування [19]. Зокрема, 70% для навчального набору забезпечує достатню кількість даних для того, щоб модель ефективно навчилася виявляти закономірності, особливо у випадках, коли доступні дані обмежені, тоді як 30% для тестового набору дозволяє надійно оцінити продуктивність моделі без ризику перенавчання, яке може виникнути через недостатню кількість тестових даних. Таким чином, співвідношення 70/30 є компромісом між ефективним навчанням моделі та надійною оцінкою її точності.

На рисунках 4.1 показано матриці помилок. У рамках задачі класифікації діаграми втомного руйнування дозволяють розподіляти режими навантаження за коефіцієнтом R на певні класи. Тут Клас 0: $R=0.2$ (низьке навантаження), клас 1: $R=0.4$ (середнє навантаження), клас 2: $R=0.6$ (високе навантаження). Тоді як вхідними параметрами були da/dN , ΔK та R .

Training Set				
TARGET \ OUTPUT	Class0	Class1	Class2	SUM
Class0	28 41.79%	0 0.00%	0 0.00%	28 100.00% 0.00%
Class1	0 0.00%	20 29.85%	0 0.00%	20 100.00% 0.00%
Class2	0 0.00%	0 0.00%	19 28.36%	19 100.00% 0.00%
SUM	28 100.00% 0.00%	20 100.00% 0.00%	19 100.00% 0.00%	67 / 67 100.00% 0.00%

a)

Training Set				
TARGET \ OUTPUT	Class0	Class1	Class2	SUM
Class0	23 33.82%	1 1.47%	0 0.00%	24 95.83% 4.17%
Class1	0 0.00%	25 36.76%	0 0.00%	25 100.00% 0.00%
Class2	0 0.00%	0 0.00%	19 27.94%	19 100.00% 0.00%
SUM	23 100.00% 0.00%	26 96.15% 3.85%	19 100.00% 0.00%	67 / 68 98.53% 1.47%

б)

Training Set				
TARGET \ OUTPUT	Class0	Class1	Class2	SUM
Class0	20 32.26%	1 1.61%	0 0.00%	21 95.24% 4.76%
Class1	0 0.00%	18 29.03%	0 0.00%	18 100.00% 0.00%
Class2	0 0.00%	0 0.00%	23 37.10%	23 100.00% 0.00%
SUM	20 100.00% 0.00%	19 94.74% 5.26%	23 100.00% 0.00%	61 / 62 98.39% 1.61%

в)

Рисунок 4.1 – Матриці помилок, отримані методом а) нейронних мереж, б) випадкових лісів та в) підсилених дерев

Метрики класифікації були отримані для кожного застосованого методу. Було визначено такі параметри, як точність, пригадування, оцінка F1 та підтримка. Рис. 4.2 містить метрики класифікації (точність, пригадування, оцінка F1 та підтримка) для кожного класу.

Class Name	Precision	1-Precision	Recall	1-Recall	F1 score
Class0	1.0000	0.0000	1.0000	0.0000	1.0000
Class1	1.0000	0.0000	1.0000	0.0000	1.0000
Class2	1.0000	0.0000	1.0000	0.0000	1.0000
Accuracy	1.0000				
Misclassification Rate	0.0000				
Macro-F1	1.0000				
Weighted-F1	1.0000				

a)

Class Name	Precision	1-Precision	Recall	1-Recall	F1 score
Class0	0.9583	0.0417	1.0000	0.0000	0.9787
Class1	1.0000	0.0000	0.9615	0.0385	0.9804
Class2	1.0000	0.0000	1.0000	0.0000	1.0000
Accuracy	0.9853				
Misclassification Rate	0.0147				
Macro-F1	0.9864				
Weighted-F1	0.9853				

b)

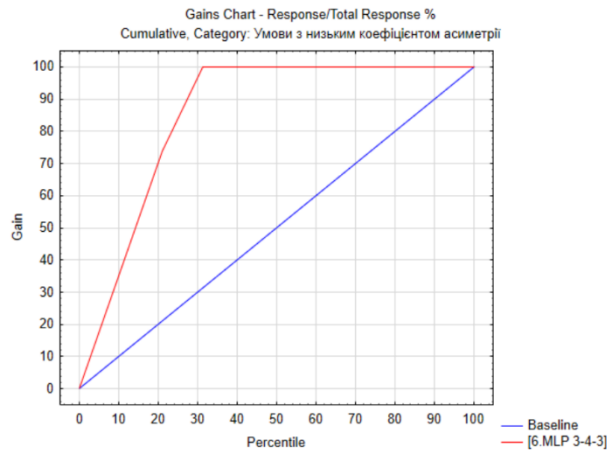
Class Name	Precision	1-Precision	Recall	1-Recall	F1 score
Class0	0.9524	0.0476	1.0000	0.0000	0.9756
Class1	1.0000	0.0000	0.9474	0.0526	0.9730
Class2	1.0000	0.0000	1.0000	0.0000	1.0000
Accuracy	0.9839				
Misclassification Rate	0.0161				
Macro-F1	0.9829				
Weighted-F1	0.9838				

в)

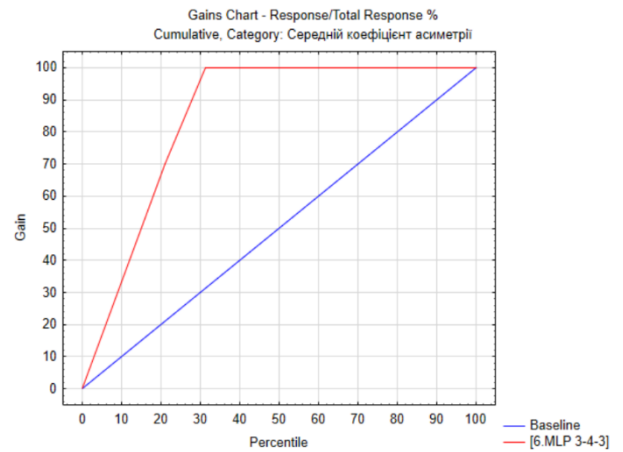
Рисунок 4.2 – Метрики класифікації, отримані методом а) нейронних мереж, б) випадкових лісів та в) підсилених дерев

На рисунку 4.3 представлено графік приросту (Gains Chart), котрий застосовують для оцінки якості моделі МН, зокрема нейронних мереж. Тут сині лінії (Baseline), лінія базового рівня відображає ідеальний випадок, коли кожен відсоток спостережень має однакову ймовірність належати до класифікованого класу та нахил цієї лінії вказує на рівномірний розподіл приросту відсотка вхідних даних. Червоні лінії, зокрема, лінія приросту моделі (у даному випадку багатошаровий персептрон із конфігурацією 3-4-3) демонструє фактичну ефективність класифікації моделі, швидкий підйом червоної лінії в перших 10–20% спостережень вказує на те, що модель дуже добре справляється з класифікацією класу "Високий коефіцієнт асиметрії" для цих даних, а після досягнення 100% приросту лінія залишається постійною, що свідчить про вичерпання релевантних спостережень.

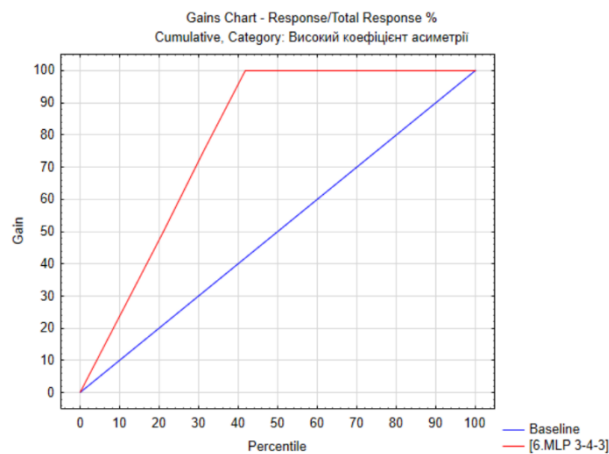
Отже, червона лінія значно вище за базову (синю), що свідчить про високу ефективність моделі класифікації, гарний початковий приріст, зокрема, у перших 20% даних модель класифікує майже 100% релевантних об'єктів.



а)



б)



в)

Рисунок 4.3 – Графік приросту, отриманий методом нейронних мереж

На рисунку представлено графік важливості змінних *Importance Plot* для залежної змінної *U*, який показує внесок кожного з параметрів у якість прогнозування моделі.

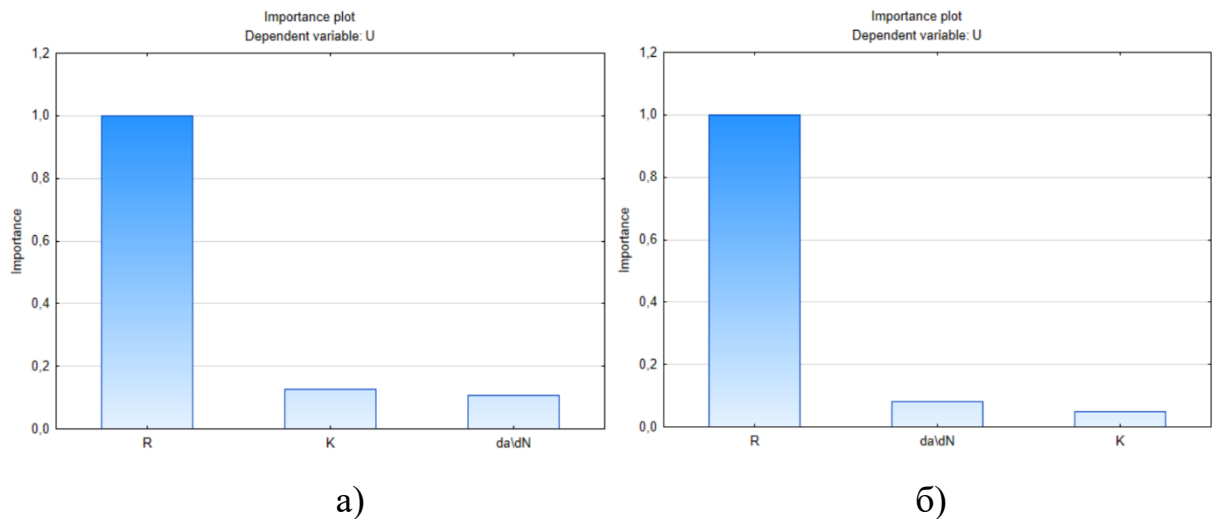


Рисунок 4.4 – Графік важливості змінних, отриманий методом а) випадкових лісів та б) підсилених дерев

Для класифікації R методом НМ було створено модель MLP 3-4-3, з функцією активації - тангенціальною та Softmax функцією для вихідного шару. У ансамблевих методах використовувалось 200 дерев.

Побудова моделей машинного навчання для класифікації включає кілька ключових етапів: підготовка даних, вибір алгоритмів, навчання, оптимізація та оцінка. Застосування нейронних мереж, випадкових лісів і підсилених дерев дозволяє досягти високої точності класифікації типу навантаження за коефіцієнтом асиметрії R . Інтеграція моделей у комп'ютеризовану систему керування забезпечує автоматизацію процесів класифікації та підвищення ефективності технологічних процесів.

4.2 Аналіз впливу параметрів моделей на якість класифікації

Аналіз впливу параметрів моделей машинного навчання є важливим етапом у процесі оптимізації алгоритмів. Різні параметри моделей, такі як кількість нейронів у шарі, кількість дерев у випадкових лісах або швидкість навчання в підсилених деревах, безпосередньо впливають на точність, узагальнення та стійкість моделі.

Нейронні мережі характеризуються великою кількістю параметрів, які потребують налаштування. Основні параметри, що впливають на якість класифікації:

- кількість шарів і нейронів у шарах:
 - збільшення кількості нейронів і шарів може покращити здатність моделі до знаходження складних залежностей;
 - однак надмірна кількість нейронів призводить до перенавчання, тобто, модель показує високу точність на навчальній вибірці, але низьку на тестовій;
- функції активації:
 - використання нелінійних функцій активації (наприклад, ReLU, Sigmoid, Tanh) дозволяє моделі враховувати складні залежності; Наприклад, ReLU працює краще для глибоких мереж, оскільки вона швидко збігається, тоді як Sigmoid підходить для завдань, де потрібно отримувати ймовірності, але може спричинити проблему затухаючого градієнта;
- швидкість навчання:
 - висока швидкість навчання може призводити до нестабільної збіжності;
 - занадто низька швидкість до повільного навчання або застрягання в локальному мінімумі;
- регуляризація:
 - використання L1, L2-регуляризації або Dropout дозволяє зменшити перенавчання.

Метод випадкових лісів залежить від наступних параметрів:

- кількість дерев (n_estimators):
 - збільшення кількості дерев покращує точність класифікації, оскільки зменшується вплив шуму в окремих деревах;

- однак надмірна кількість дерев збільшує час навчання та використання пам'яті;
- максимальна глибина дерева (`max_depth`):
 - глибші дерева краще моделюють залежності, але збільшують ризик перенавчання;
 - обмеження глибини допомагає покращити узагальнення моделі;
- розмір вузла (`min_samples_split`, `min_samples_leaf`):
 - зменшення мінімального розміру вузла дозволяє створювати більш точні дерева, але це також може спричинити перенавчання;
 - великий мінімальний розмір вузлів зменшує ризик перенавчання, але може пропустити важливі залежності;
- кількість ознак (`max_features`):
 - оптимальний вибір кількості ознак дозволяє покращити точність і зменшити час навчання.

Підсилені дерева, як-от XGBoost, LightGBM, залежні від налаштування гіперпараметрів:

- швидкість навчання (`learning_rate`):
 - низька швидкість навчання покращує здатність моделі до узагальнення, але вимагає збільшення кількості ітерацій;
- кількість ітерацій (`n_estimators`):
 - більша кількість ітерацій дозволяє моделі вивчити складніші залежності, але збільшує час навчання;
- максимальна глибина дерева (`max_depth`):
 - як і у випадкових лісах, занадто глибокі дерева можуть призвести до перенавчання;
- розмір вузла:
 - аналогічно випадковим лісам, цей параметр регулює, коли модель зупиняється в розгалуженні дерева;
- рівень регуляризації (L1, L2):

- регуляризація зменшує ризик перенавчання і покращує стійкість моделі.

Відомо, що занадто складні моделі із великою кількістю параметрів мають високу точність на навчальній вибірці, але низьку на тестовій, тоді як занадто прості моделі (наприклад, низька глибина дерева) не здатні адекватно відобразити складні залежності у даних. Зокрема, моделі по-різному реагують на вхідні дані. Наприклад, нейронні мережі можуть бути більш стійкими до шуму, ніж підсилені дерева.

Загалом, правильний вибір параметрів суттєво впливає на якість класифікації, а також, для кожної моделі важливо знайти баланс між складністю моделі та її здатністю до узагальнення. Тому рекомендується використовувати автоматизовані методи оптимізації, як-от Grid Search або Bayesian Optimization, щоб знайти оптимальні параметри.

4.3 Оцінка узгодженості результатів класифікації з експериментальними даними ДВР

Оцінка узгодженості результатів класифікації є важливим етапом валідації моделей машинного навчання, особливо для задач, пов'язаних із прогнозуванням характеристик матеріалів, таких як втомне руйнування. Мета цього етапу полягає в порівнянні передбачених результатів класифікації з реальними експериментальними даними, отриманими для алюмінієвих сплавів.

Оцінка узгодженості ґрунтується на:

- аналізі помилок класифікації, тобто, оцінюється кількість правильних і неправильних передбачень для кожного класу навантажень (низький, середній, високий коефіцієнт асиметрії R). А також визначаються такі метрики, як точність (accuracy), повнота (recall), точність прогнозу (precision) та F1-міра для кожного класу;

- графічному порівнянні, візуалізація діаграм відповідності (Gain Chart), які дозволяють оцінити, наскільки модель перевищує базовий рівень у класифікації;
- порівнянні з експериментальними діаграмами, тобто, результати класифікації (наприклад, тип навантаження) порівнюються з реальними даними діаграм втомного руйнування, побудованими за параметрами da/dN , ΔK та R , а також, використовуються експериментальні точки з високою точністю вимірювань для перевірки моделі.

Для оцінки узгодженості застосовується кілька моделей машинного навчання, зокрема, нейронні мережі, випадкові ліси, підсилені дерева. Аналіз включає порівняння матриць неточностей (confusion matrix) для кожної моделі та оцінку середніх значень метрик класифікації:

Для узгодження результатів класифікації з експериментальними даними застосовується порівняльний аналіз:

- експериментальні діаграми втомного руйнування, тобто, для кожного коефіцієнта асиметрії R побудовані експериментальні залежності da/dN від ΔK , а також класифікація типу навантаження проводиться на основі цих залежностей;
- порівняння прогнозованих і реальних класів, зокрема, зіставлення реальних класів (експериментально отриманих) із результатами моделі.

Загалом, якщо експериментальні та прогнозовані значення збігаються з високою точністю, модель вважається надійною. У випадках значних розбіжностей необхідно повторно оцінити модель або провести додатковий збір експериментальних даних.

РОЗДІЛ 5. СПЕЦІАЛЬНА ЧАСТИНА

5.1 Практична реалізація класифікації типу навантаження

Для класифікації типу навантаження у промислових системах можна застосовувати нейронні мережі, що дозволить підвищити точність і ефективність процесу. Розглянемо детально етапи практичної реалізації такої класифікації (рис.5.1-5.9).

Спочатку обираємо розділ *Data Mining* (рис. 5.1, 5.2). Для класифікації вибираємо тип аналізу "*Classification*" (рис. 5.3).

	1 da\dn	2 K	3 R	4 U	5 Var5
1	0,3579	-9,27	0,2	Умови з низьким коефіцієнтом асиметрії	
2	0,361	-10,451	0,2	Умови з низьким коефіцієнтом асиметрії	
3	0,3673	-9,7496	0,2	Умови з низьким коефіцієнтом асиметрії	
4	0,3754	-9,3354	0,2	Умови з низьким коефіцієнтом асиметрії	
5	0,3945	-9,0057	0,2	Умови з низьким коефіцієнтом асиметрії	
6	0,3972	-9,0209	0,2	Умови з низьким коефіцієнтом асиметрії	
7	0,4242	-8,9355	0,2	Умови з низьким коефіцієнтом асиметрії	
8	0,4346	-8,983	0,2	Умови з низьким коефіцієнтом асиметрії	
9	0,4396	-8,6635	0,2	Умови з низьким коефіцієнтом асиметрії	
10	0,493	-8,4962	0,2	Умови з низьким коефіцієнтом асиметрії	
11	0,4953	-8,6556	0,2	Умови з низьким коефіцієнтом асиметрії	
12	0,5201	-8,6216	0,2	Умови з низьким коефіцієнтом асиметрії	
13	0,5222	-8,4789	0,2	Умови з низьким коефіцієнтом асиметрії	
14	0,5534	-8,4828	0,2	Умови з низьким коефіцієнтом асиметрії	
15	0,5933	-8,3799	0,2	Умови з низьким коефіцієнтом асиметрії	
16	0,5942	-8,251	0,2	Умови з низьким коефіцієнтом асиметрії	
17	0,5959	-8,4123	0,2	Умови з низьким коефіцієнтом асиметрії	
18	0,5977	-8,5171	0,2	Умови з низьким коефіцієнтом асиметрії	
19	0,5986	-8,5229	0,2	Умови з низьким коефіцієнтом асиметрії	
20	0,6012	-8,301	0,2	Умови з низьким коефіцієнтом асиметрії	
21	0,6038	-8,4123	0,2	Умови з низьким коефіцієнтом асиметрії	

Рисунок 5.1 – Дані для класифікації

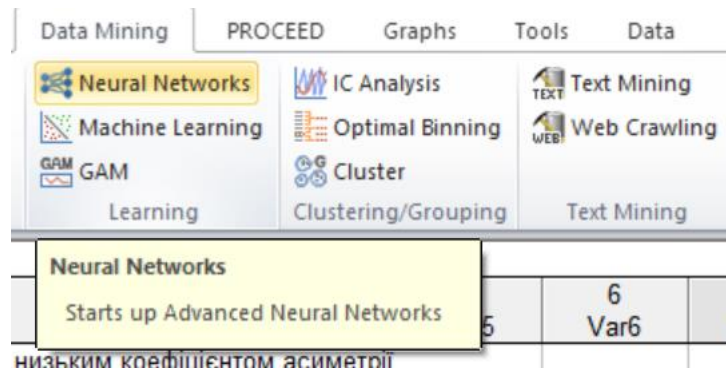


Рисунок 5.2 – Меню *Data Mining*

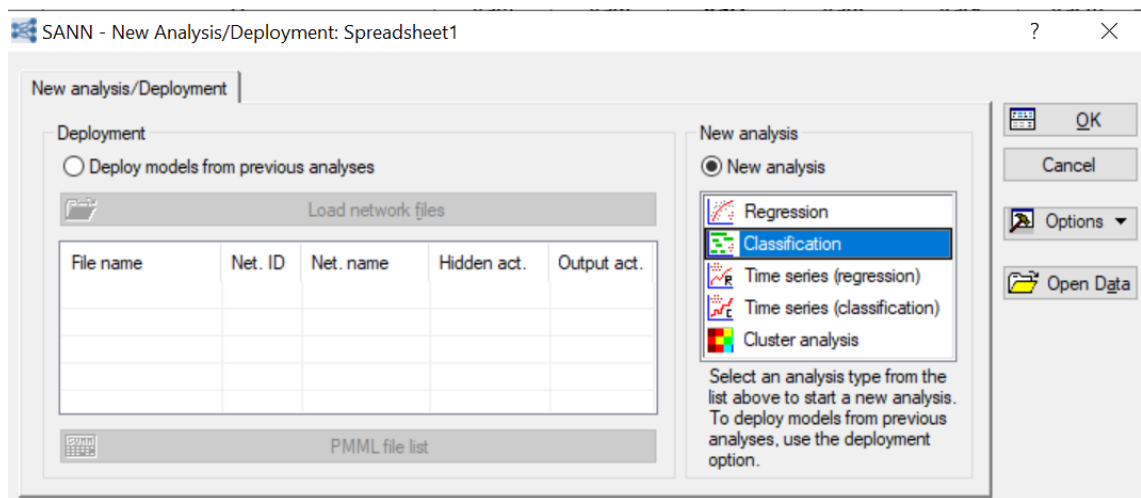


Рисунок 5.3 – *Classification* як тип аналізу даних

Обираючи *Variables* (рис. 5.4), вказуємо вихідну змінну (перший стовпець), а тоді вибираємо вхідні змінні (другий стовбець). Переходимо до *Quick* (рис. 5.5).

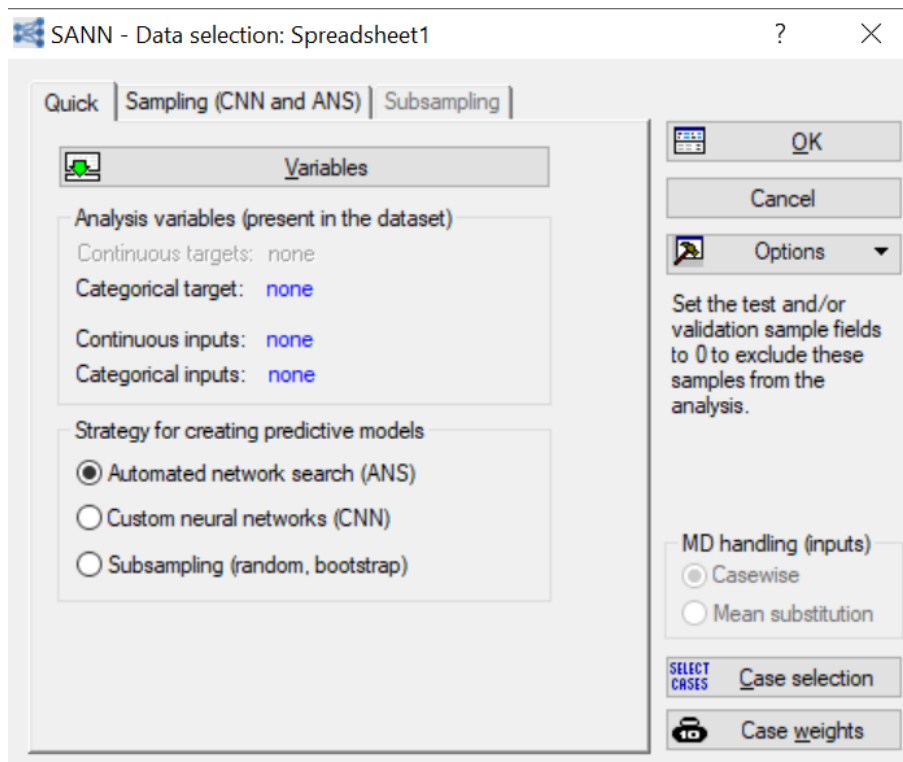


Рисунок 5.4 – *Variables* для аналізу даних

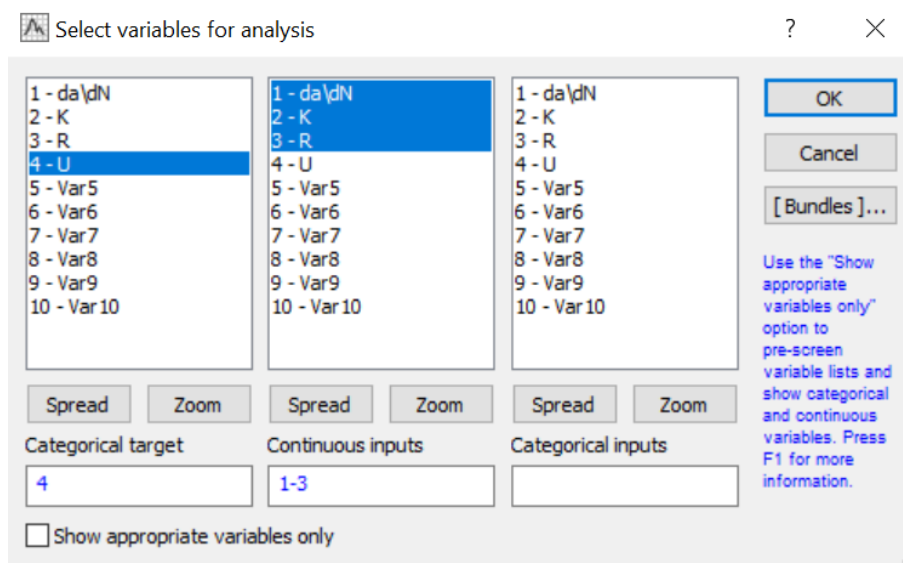


Рисунок 5.5 – Для перевірки відповідних змінних *Quick*

У меню *Sampling* розділяємо дані на навчальну та тестову вибірки (рис. 5.6).

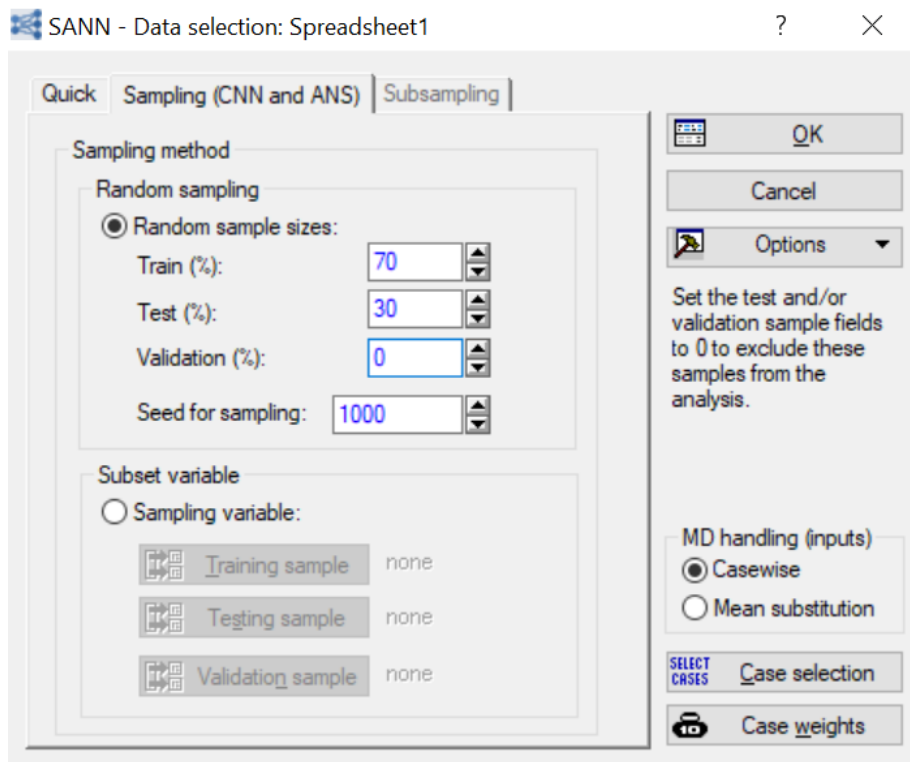


Рисунок 5.6 – Для розподілу на вибірки меню *Sampling*

Дані для побудови нейронної мережі показані на рис. 5.7.

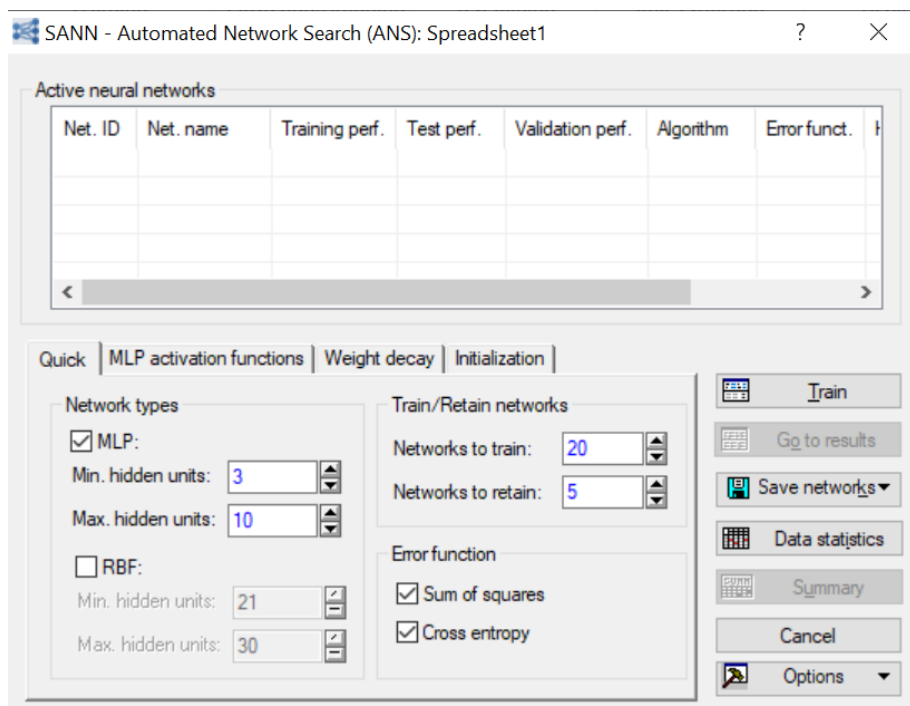


Рисунок 5.7 – Побудова мережі

Процес навчання нейронних мереж показаний на рис. 5.8.

Neural network training in progress...

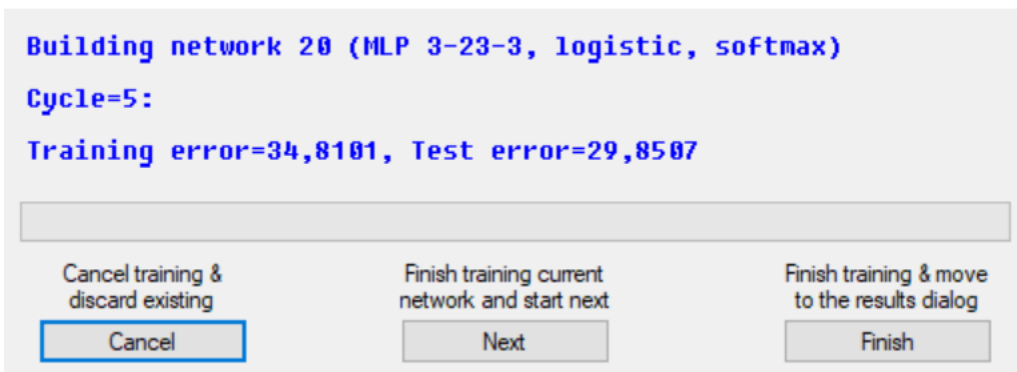


Рисунок 5.8 – Навчання нейронних мереж

Результати класифікації, отримані за допомогою нейронних мереж, зображені на рис. 5.9.

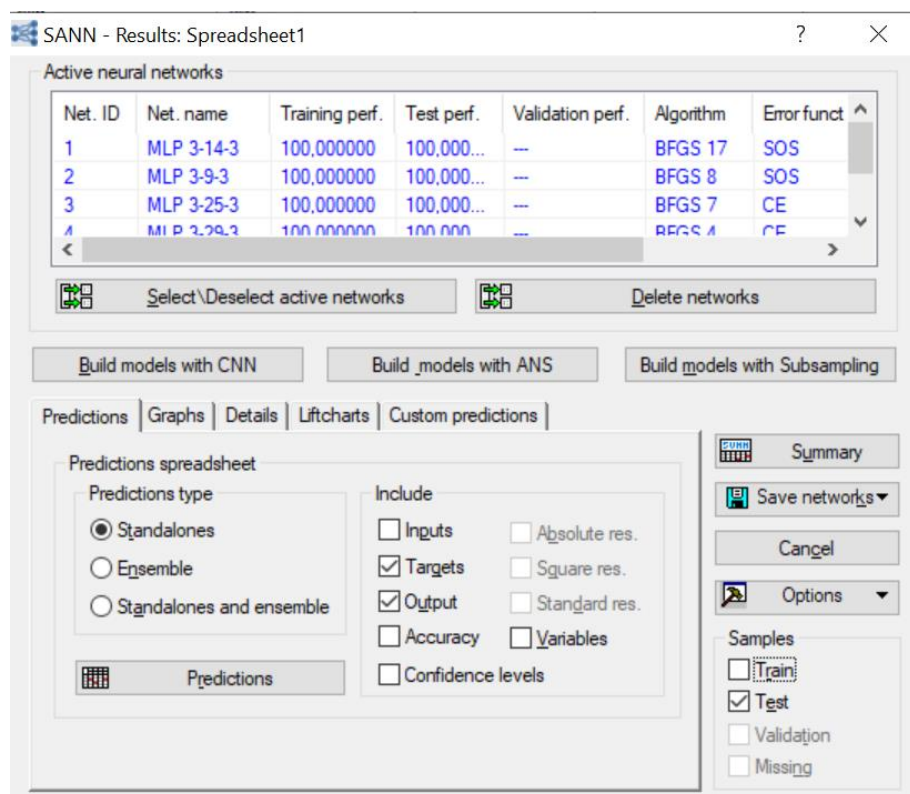


Рисунок 5.9 – Результати класифікації НМ

Класифікація типу навантаження НМ є потужним інструментом для автоматизації і покращення технологічних процесів у промислових системах. Використання нейронних мереж дозволяє досягати високої точності класифікації, зменшувати витрати на діагностику та підвищувати ефективність керування. Для успішної реалізації важливо забезпечити якісні дані, належне апаратне забезпечення і достатню кваліфікацію персоналу.

5.2 Перевірка надійності та стійкості системи у різних режимах роботи

Надійність та стійкість системи машинного навчання для класифікації типу навантаження є ключовими параметрами її успішного функціонування в реальних умовах. Перевірка цих параметрів здійснюється шляхом моделювання різних сценаріїв роботи системи та оцінки її поведінки в умовах, що відрізняються від стандартних.

Система перевіряється на типових наборах даних, які використовувались для навчання та валідації:

- стабільність роботи, оцінюється, чи працює система без збоїв на стандартному тестовому наборі даних;
- час обробки, перевіряється час класифікації одного зразка та ефективність системи при обробці великих обсягів даних;
- відтворюваність результатів, виконується декілька запусків з однаковими даними для перевірки, чи система забезпечує однакові результати.

Реальні дані часто містять похибки або шум (наприклад, некоректні значення або відсутні дані). Перевірка системи включає введення шуму в вхідні параметри і оцінка, як це впливає на результати, випадкове видалення або заміна частини даних та оцінка стійкості моделі до таких змін за допомогою метрик точності та відсотка помилок.

Загалом, система тестується на даних, які суттєво відрізняються від тренувального набору (наприклад, нові типи навантаження, що не були включені

до навчання моделі), аналіз результатів класифікації на нестандартних даних та розробка механізмів обробки винятків, тобто, система має сигналізувати про ситуації, коли вхідні дані виходять за рамки навчального діапазону.

РОЗДІЛ 6. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

6.1 Характеристика шкідливих факторів виробничого середовища

Під час роботи на виробництві на людину можуть впливати один або низка небезпечних та шкідливих виробничих факторів. Безпека того чи іншого технологічного процесу може бути визначена за їх кількістю і за ступенем небезпеки кожного з них зокрема. Безпека праці на виробництві визначається ступенем безпеки окремих технологічних процесів.

Небезпечні й шкідливі виробничі фактори стандартом ГОСТ 12.0.003-74 поділяються на фізичні, хімічні, біологічні й психофізіологічні. Останні за характером впливу на людину підрозділяються на фізичні й нервово-психічні перевантаження, а інші – на конкретні небезпечні й шкідливі виробничі фактори. В процесі роботи на підприємстві на працівника можуть впливати такі небезпечні й шкідливі виробничі фактори:

- машини, що рухаються, автотранспорт і механізми;
- рухомі незахищені елементи механізмів, машин і виробничого обладнання;
- падаючі вироби техніки, інструмент і матеріали під час роботи;
- ударна хвиля (вибух посудини, що працює під тиском пари рідини);
- струмені газів і рідин, що стікають із посудин і трубопроводів під тиском;
- підвищене ковзання (через зледеніння, зволоження й замаслювання поверхонь, по яких переміщується робочий персонал);
- підвищені запыленість й загазованість повітря;
- підвищена чи знижена температура поверхонь техніки, обладнання й матеріалів;
- підвищена чи знижена температура, вологість і рухомість повітря;
- підвищений рівень шуму, вібрації, ультра та інфразвуку;
- підвищена напруга в електричному ланцюзі, замикання якого може відбутися через тіло людини;
- підвищений рівень статичної електрики;

- гострі кромки, задирки й шорсткість на поверхнях обладнання й інструментів;
- відсутність чи нестача природного світла;
- недостатня освітленість робочої зони;
- знижена контрастність об'єктів в порівнянні з фоном;
- пряма блискість (прожекторне освітлення територій виробництв, світло фар автотранспорту) і відбита блискість (від розливої води й інших рідин на поверхні територій виробництв);
- підвищена пульсація світлового потоку;
- підвищений рівень ультрафіолетової й інфрачервоної радіації;
- хімічні речовини (токсичні, подразнюючі, сенсibiliзуючі, канцерогенні, мутагенні, що впливають на репродуктивну функцію людина);
- хімічні речовини, що проникають в організм через органи дихання, шлунково-кишковий тракт, шкірні покриви і слизові оболонки;
- патогенні мікроорганізми (бактерії, віруси, гриби, найпростіші) і продукти їхньої життєдіяльності;
- перевантаження (статичні й динамічні) і нервово-психічні чинники (емоційні перевантаження, перенапруга аналізаторів, розумова перенапруга, монотонність праці).

Рівні небезпечних і шкідливих виробничих факторів не повинні перевищувати граничнодопустимих значень, встановлених у санітарних нормах, правилах і нормативно-технічній документації.

6.2 Забезпечення безпеки життєдіяльності при роботі з ПК

Під час роботи на комп'ютерах можуть діяти такі небезпечні та шкідливі фактори, як:

- фізичні;

– психофізіологічні.

Електробезпека при роботі.

Заходи щодо усунення небезпеки ураження електричним струмом зводяться до правильного розміщення устаткування та електричних кабелів. Інші заходи щодо забезпечення електробезпеки, збігаються з загальними заходами пожежо- та електробезпеки.

В якості профілактичних заходів для забезпечення пожежної безпеки слід використовувати скриту електромережу, надійні розетки з пожежобезпечних матеріалів, силові мережі живлення устаткування виконувати кабелями, розрахованими на підключення в 3-5 разів більшого навантаження, включати й виключати живлення обладнання за допомогою штатних вимикачів. Треба регулярно робити очистку внутрішніх частин комп'ютерів, іншого устаткування від пилу, розташовувати комп'ютери на окремих неспалюваних столах. Для запобігання іскріння необхідно рідше встромляти і виймати штепсельні вилки з розеток.

Освітлення.

Система освітлення повинна відповідати таким вимогам:

- освітленість на робочому місці повинна відповідати характеру зорової роботи, який визначається трьома параметрами: об'єктом розрізнення – найменшим розміром об'єкта, що розглядається на моніторі ПК; фоном, який характеризується коефіцієнтом відбиття; контрастом об'єкта і фону;

- необхідно забезпечити достатньо рівномірне розподілення яскравості на робочій поверхні монітора, а також в межах навколишнього простору;

- на робочій поверхні повинні бути відсутні різкі тіні;

- в полі зору не повинно бути відблискі (підвищеної яскравості поверхонь, які світяться та викликають осліплення);

- величина освітленості повинна бути постійною під час роботи;

– слід обирати оптимальну спрямованість світлового потоку і необхідний склад світла.

Вимоги до монітору.

Основним обладнанням робочого місця користувача комп'ютера є монітор, системний блок та клавіатура.

Робочі місця мають бути розташовані на відстані не менше 1,5 м від стіни з вікнами, від інших стін на відстані 1 м, між собою на відстані не менше 1,5 м. Відносно вікон робоче місце доцільно розташовувати таким чином, щоб природне світло падало на нього збоку, переважно зліва.

Робочі місця слід розташовувати так, щоб уникнути попадання в очі прямого світла. Джерела освітлення рекомендується розташовувати з обох боків екрану паралельно напрямку погляду. Для уникнення світлових відблисків екрану, клавіатури в напрямку очей користувача, від світильників загального освітлення або сонячних променів, необхідно використовувати антиполюскові сітки, спеціальні фільтри для екранів, захисні козирки, на вікнах – жалюзі.

Екран дисплея повинен бути розташованим перпендикулярно до напрямку погляду. Якщо він розташований під кутом, то стає причиною сутулості. Відстань від дисплея до очей повинна трохи перевищувати звичну відстань між книгою та очима. Перед екраном монітора, особливо старих типів, повинен бути спеціальний захисний екран. При його відсутності треба сидіти на відстані витягнутої руки від монітора.

Фільтри з металевої або нейлонової сітки використовувати не рекомендується, тому що сітка спотворює зображення через інтерференцію світла. Найкращу якість зображення забезпечують скляні поляризаційні фільтри. Вони усувають практично усі відблиски, роблять зображення чітким і контрастним.

При роботі з текстовою інформацією (в режимі введення даних та редагування тексту, читання з екрану) найбільш фізіологічним правильним є зображення чорних знаків на світловому (чорному) фоні.

Монітор повинен бути розташований на робочому місці так, щоб поверхня екрана знаходилася в центрі поля зору на відстані 400-700 мм від очей користувача. Рекомендується розміщувати елементи робочого місця так, щоб витримувалася однакова відстань очей від екрана, клавіатури, тексту.

Робоча поза.

Зручна робоча поза при роботі з комп'ютером забезпечується регулюванням висоти робочого столу, крісла та підставки для ніг. Раціональною робочою позою може вважатися таке положення, при якому ступні працівника розташовані горизонтально на підлозі або підставці для ніг, стегна зорієнтовані у горизонтальній площині, верхні частини рук – вертикальній. Кут ліктьового суглоба коливається в межах 70-90°, зап'ястя зігнуті під кутом не більше ніж 20°, нахил голови 15-20°.

Важливою є форма спинки крісла, яка повинна повторювати форму спини. Висота крісла повинна бути такою, щоб користувач не відчував тиску на куприк або стегна. Крісло бажано обладнати бильцями. Його потрібно встановити так, щоб не треба було тягтися до клавіатури. Періодично користувачу необхідно рухатися, вчасно змінювати положення тіла і робити перерви у роботі.

При напруженій роботі за комп'ютером щогодини необхідно робити перерву на 15 хвилин через кожну годину і треба займатися іншою справою. Декілька разів на годину бажано виконувати серію легких вправ для розслаблення.

Для нейтралізації зарядів статичної електрики в приміщенні, де виконується робота на комп'ютерах, в тому числі на лазерних та світлодіодних принтерах, рекомендується збільшувати вологість повітря за допомогою кімнатних зволожувачів. Не рекомендується носити одяг з синтетичних матеріалів.

Вимоги безпеки перед початком роботи:

- увімкнути систему кондиціонування в приміщенні;

- перевірити надійність встановлення апаратури на робочому столі. Повернути монітор так, щоб було зручно дивитися на екран – під прямим кутом (а не збоку) і трохи зверху вниз, при цьому екран має бути трохи нахиленим, нижній його край ближче до оператора;

- перевірити загальний стан апаратури, перевірити справність електропроводки, з'єднувальних шнурів, штепсельних вилок, розеток, заземлення захисного екрана;

- відрегулювати освітленість робочого місця;

- відрегулювати та зафіксувати висоту крісла, зручний для користувача нахил його спинки;

- приєднати до системного блоку необхідну апаратуру. Усі кабелі, що з'єднують системний блок з іншими пристроями, слід вставляти та виймати при вимкненому комп'ютері;

- ввімкнути апаратуру комп'ютера вимикачами на корпусах в послідовності: монітор, системний блок, принтер (якщо передбачається друкування);

- відрегулювати яскравість свічення монітора, мінімальний розмір світної точки, фокусування, контрастність. Не слід робити зображення надто яскравим, щоб не втомлювати очей.

Вимоги безпеки під час виконання роботи:

- необхідно стійко розташовувати клавіатуру на робочому столі, не опускати її хитання. Під час роботи на клавіатурі сидіти прямо, не напружуватися;

- для забезпечення несприятливого впливу на користувача пристроїв типу «миша» належить забезпечувати вільну велику поверхню столу для переміщення «миші» і зручного упору ліктьового суглоба;

- не дозволяються посторонні розмови, подразнюючі шуми;

- періодично при вимкненому комп'ютері прибирати ледь змоченою мильним розчином бавовняною ганчіркою пил з поверхонь апаратури. Екрани

ВДТ та захисний екран протирають ганчіркою, змоченою у спирті. Не дозволяється використовувати рідинні або аерозольні засоби очищення поверхонь комп'ютера.

Забороняється:

- класти будь-які предмети на апаратуру комп'ютера;
- закривати будь-чим вентиляційні отвори апаратури, що може призвести до її перегрівання і виходу з ладу.

ВИСНОВКИ

У магістерській роботі були отримані важливі результати, які підтверджують ефективність використання методів машинного навчання для вирішення задач класифікації типу навантаження за параметрами втомного руйнування. Оцінка якості роботи системи та її стійкості дозволила виявити сильні сторони запропонованих підходів та надати рекомендації для їх практичного впровадження.

Ефективністю методів машинного навчання є:

- нейронні мережі, зокрема, архітектура багатошарових персептронів (MLP) показала високі результати точності класифікації. А також, нейронні мережі демонструють здатність адаптуватися до складних нелінійних залежностей між параметрами da/dN , ΔK та R ;
- випадкові ліси, загалом, метод забезпечив високу стійкість до шуму у вхідних даних. Випадкові ліси показали швидшу обробку даних порівняно з нейронними мережами при збереженні високого рівня точності;
- підсилені дерева, зокрема, метод підсилення дерев дозволив досягти найкращої продуктивності на малих наборах даних, демонструючи відмінні результати у випадках, коли навчальна вибірка обмежена.

Проаналізувавши вхідні параметри, можна зробити висновки, що коефіцієнт асиметрії навантаження R виявився найбільш впливовим параметром для класифікації, що підтверджується графіками важливості ознак, а також, параметри da/dN , ΔK мають менший вплив на результати класифікації, але залишаються важливими для побудови моделі, особливо в комбінації з іншими даними.

Загалом, перевагами попередньої обробки даних є застосування нормалізації та масштабування суттєво підвищило точність класифікації, особливо для нейронних мереж, а також, попередня очистка даних (усунення

пропусків та аномалій) забезпечила стабільну роботу моделей навіть при введенні шуму у вхідні параметри.

Отже, найкращі результати були отримані за метриками F1-score та Ассурасу, які показали узгодженість класифікації моделей з експериментальними даними. А також, проведене тестування підтвердило, що система залишається надійною навіть за наявності шуму у вхідних даних чи обмежених обчислювальних ресурсів. Система демонструє високу стійкість у режимах роботи з великими наборами даних, а також у режимі реального часу. Загалом, результати класифікації узгоджуються з експериментальними даними діаграм втомного руйнування алюмінієвого сплаву та практичне тестування підтвердило, що запропонована система може бути ефективно використана для аналізу навантажень у реальних технологічних процесах.

Можна зробити висновки, що запропонована система класифікації типу навантаження, заснована на методах машинного навчання, є ефективним інструментом для аналізу втомного руйнування матеріалів. Її практичне застосування дозволяє підвищити надійність та безпеку технічних конструкцій, а також автоматизувати процес діагностики та прийняття рішень у комп'ютеризованих системах керування технологічними процесами.

ПЕРЕЛІК ДЖЕРЕЛ

1. McCrady S. G. Designing SCADA application software: A practical approach – Amsterdam, Singapore: Elsevier Science, 2013, 227 pages.
2. Frank D. Petruzella. Programmable Logic Controllers, Published by McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121. Copyright © 2017 by McGraw-Hill
3. Haykin, S., 2009 “Neural Networks and Learning Machines”, Third Edition, McMaster University, Hamilton, Ontario, Canada, 938 p.
4. Шаховська Н.Б., Камінський Р.М., Вовк О.Б. Системи штучного інтелекту: навч. посібник. Львів: Видавництво Львівської політехніки, 2018. 392 p.
5. Alpayndin E. Introduction to Machine Learning // The Knowledge Engineering Review. 2010. Vol. 25, № 3. P. 353–353.
6. Mitchell T.M. Machine learning. London: McGraw-Hill Science/Engineering/Math, 1997.
7. Wasserman Ph.D. Neural Computing: Theory and Practice. New York: Coriolis Group (Sd), 1989.
8. Smola A., Vishwanathan S.V.N. Introduction to Machine Learning. Cambridge University Press, 2010.
9. Kuna M. Finite Elements in Fracture Mechanics: Theory-Numerics-Applications. Springer, 2013. 446 p.
10. Варфоломеев І.В., Ясній.О.П. Моделювання руйнування імовірнісними методами елементів конструкцій з тріщинами // Фіз.-хім. механіка матеріалів. 2008. Vol. 44, № 1. P. 76–83.
11. Ясній О.П., Собчак.А.Р., Ясній В.П. Оцінювання ймовірності руйнування колектора пароперегрівника // Фіз.-хім. механіка матеріалів. 2014. Vol. 50, № 3. P. 63–68.
12. Paris P.C. The growth of fatigue cracks due to variations in load : Ph.D.Thesis. Lehigh University, 1962. 263 p.

13. Pidaparti R.M.V., Palakal M.J. Neural network approach to fatigue-crack-growth predictions under aircraft spectrum loadings // *J Aircr.* 1995. Vol. 32, № 4. P. 825–831.
14. Mohanty J.R. et al. Application of artificial neural network for predicting fatigue crack propagation life of aluminum alloys // *Archives of Computational Materials Science and Surface Engineering.* 2009. Vol. 1, № 3. P. 133–138.
15. O. Yasniy, I. Didych, Yu. Lapusta: Prediction of fatigue crack growth diagrams by methods of machine learning under constant amplitude loading, *Acta Metallurgica Slovaca*, 26(1), 2020, 31 - 33.
16. Yasniy, O., Mytnyk, M., Maruschak, P., Mykytyshyn, A., & Didych, I., 2024 “Machine learning methods as applied to modelling thermal conductivity of epoxy-based composites with different fillers for aircraft,” *Aviation*, 28(2), pp. 64-71.
17. Ясній О.П. et al. Прогнозування діаграм втомного руйнування алюмінієвого сплаву Д16Т методами машинного навчання // *Фізико-хімічна механіка матеріалів.* 2018. Vol. 54, № 3. P. 43–48.
18. Tymoshchuk D, Yasniy O, Maruschak P, Iasnii V, Didych I. Loading Frequency Classification in Shape Memory Alloys: A Machine Learning Approach. *Computers.* 2024; 13(12):339. <https://doi.org/10.3390/computers13120339>
19. Ясній П., Пиндус Ю., Фостик В. Вплив асиметрії циклу навантаження на характеристики циклічної тріщиностійкості алюмінієвого сплаву Д16Т // *Вісник Тернопільськ. держ. техн. ун-ту.* 2007. Vol. 12, № 1. P. 7–12.

ДОДАТКИ

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний технічний університет імені Івана Пулюя (Україна)
Університет імені П'єра і Марії Кюрі (Франція)
Маріборський університет (Словенія)
Технічний університет у Кошице (Словаччина)
Вільнюський технічний університет ім. Гедимінаса (Литва)
Наукове товариство ім. Т.Шевченка

АКТУАЛЬНІ ЗАДАЧІ СУЧАСНИХ ТЕХНОЛОГІЙ

Збірник
тез доповідей

**ХІІІ Міжнародної науково-практичної
конференції молодих учених та студентів**
11-12 грудня 2024 року



УКРАЇНА
ТЕРНОПІЛЬ – 2024

Зміст

СЕКЦІЯ: ФІЗИКО-ТЕХНІЧНІ ОСНОВИ РОЗВИТКУ НОВИХ ТЕХНОЛОГІЙ

1. М.І. Яворська, к.т.н., доцент; Б.Я. Воляннюк; М.С. Попович; В.Г. Базан 8
РОЗРАХУНОК НАДІЙНОСТІ ПРИСТРОЮ ПРИЗНАЧЕНОГО ДЛЯ ВИМІРЮВАННЯ
АРТЕРІАЛЬНОГО ТИСКУ
2. Т.С. Дубиняк, к.т.н.; Н.Є. Юрик к.е.н.; Ю.І. Наконечний, ст.викл.; Ю.І. Юрик 10
МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА АНАЛІЗ ПРИСТРОЮ, ЩО ПРИЗНАЧЕНИЙ
ДЛЯ ВИМІРЮВАННЯ ВІДХИЛЕНЬ ФОРМИ ЦИЛІНДРИЧНИХ ПОВЕРХОНЬ
3. Т.С.Дубиняк, к.т.н., доцент; В.А.Невожай; В.О. Гесюк; Я.Я. Коляса 12
ІНФОРМАЦІЙНО ВИМІРЮВАЛЬНА СИСТЕМА ПРИСТРОЮ
УЛЬТРАЗВУКОВОГО СИГНАЛІЗАТОРА ЗАПАЛЮВАННЯ

СЕКЦІЯ: НОВІ МАТЕРІАЛИ, МІЦНІСТЬ І ДОВГОВІЧНІСТЬ ЕЛЕМЕНТІВ КОНСТРУКЦІЙ

1. А. Г. Микитишин, І. С. Дідич, Р. І. Яцишин 14
ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСИФІКАЦІЇ
ДАНИХ В КОМП'ЮТЕРИЗОВАНИХ СИСТЕМАХ КЕРУВАННЯ
2. І. С. Дідич, А. Г. Микитишин, Р. З. Золотий, П. А. Ясіновський 17
ІНТЕГРАЦІЯ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ ОПТИМІЗАЦІЇ
ПРОЦЕСІВ АВТОМАТИЗОВАНОГО КЕРУВАННЯ
3. В.Б. Ігнат'єва, В.О. Паньків, В.М. Тимкович, С.В. Синюк, Ю.В. Сокотов 20
ВЕЛИКОПРОЛІТНІ АРОЧНІ КОНСТРУКЦІЇ ПОКРИТТЯ
4. Г.М. Крамар, М.І. Гудь, Н.А. Яржемський 22
РОБОТА ПРОСТОРОВОЇ КОНСОЛЬНОЇ ФЕРМИ
5. Г.М. Крамар, Р.Я. Васирик, К.В. Клименко 23
ПРОСТОРОВІ ВЕЛИКОПРОЛІТНІ ПОКРИТТЯ
6. Д.Я. Баран, І.В. Габрик, С.В. Криницький, Г.Р. Гавришак 24
КОНСТРУКТИВНІ РІШЕННЯ ПРОСТОРОВИХ КАРКАСІВ ГРОМАДСЬКИХ
СПОРУД
7. Л.Г. Бодрова, О.В. Поліщук, Ю.Ю. Данилевич. 25
КОНСТРУКТИВНІ РІШЕННЯ ВЕЛИКОПРОЛІТНИХ КОНСОЛЕЙ
8. М.І. Гудь, О.М. Юркевич 27
БЕТОНИ НА ОСНОВІ ЕПОКСИДНИХ СМОЛ
9. М.І. Гудь, П.В. Гиря 28
ВЛАСТИВОСТІ ПОЛІМЕРБЕТОНІВ
10. М.І. Гудь, Р.А. Бойко, С.В. Кравчук. 29
КОНСТРУКТИВНІ РІШЕННЯ ПЕРЕКРИТТЯ ВИСОТНИХ ГРОМАДСЬКИХ
СПОРУД
11. Я.Л. Швед, В.Р. Поліщук, В.В. Мардинавка. 30
РЕШЧАСТІ СИСТЕМИ КРОКВЯНИХ ФЕРМ
12. А.Б.Гупка, О.І. Пришляк, М.І. Теплюк 31

**СЕКЦІЯ: НОВІ МАТЕРІАЛИ, МІЦНІСТЬ І ДОВГОВІЧНІСТЬ ЕЛЕМЕНТІВ
КОНСТРУКЦІЙ**

УДК 539.42, 004.032.26,

А. Г. Микитишин, канд. техн. наук, доцент, І. С. Дідич, доктор філософії, Р. І. Яцишин

Тернопільський національний технічний університет імені Івана Пулюя, Україна.

**ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ
КЛАСИФІКАЦІЇ ДАНИХ В КОМП'ЮТЕРИЗОВАНИХ СИСТЕМАХ КЕРУВАННЯ**

A. H. Mykytyshyn., Ph.D., Assoc. Prof., I. S. Didych, Ph.D., R. I. Yatsyshyn

**APPLICATION OF MACHINE LEARNING METHODS FOR DATA
CLASSIFICATION IN COMPUTERIZED CONTROL SYSTEMS**

Сучасні комп'ютеризовані системи керування технологічними процесами виконують критичні функції у промисловості, зокрема моніторинг, контроль і оптимізацію роботи обладнання. Зокрема, збільшення обсягів даних від сенсорів та інших пристроїв створює необхідність в інтелектуальній обробці та класифікації цих даних. Тому методи машинного навчання (ML), а саме, нейронні мережі, випадкові ліси, підсилені дерева, метод опорно-векторних машин та k -найближчих сусідів є перспективним інструментом для класифікації вхідних даних, виявлення аномалій і покращення автоматизованих рішень у таких системах [1].

Відомо, що методи машинного навчання дозволяють комп'ютерам аналізувати дані, виявляти закономірності та приймати рішення без явного програмування для кожного завдання. Ці методи дозволяють обробляти великі обсяги інформації, адаптуватися до змін умов і покращувати свої результати з часом [2].

Важливими етапами для класифікації даних є збір і обробка даних, побудова моделей та інтеграція їх в систему. На першому етапі, дані надходять із сенсорів і систем моніторингу, зокрема, важливою є попередня обробка даних, тобто, нормалізація, усунення шуму, обробка пропущених значень. Тоді як побудова моделей складається з вибору методу класифікації, котрий залежить від особливостей даних та вимог системи, а також навчання моделей на даних для забезпечення точності прогнозів. Далі виявлення аномалій чи класифікація подій ініціює дії системи (наприклад, аварійне вимкнення).

Нейронні мережі (НМ) є потужним методом для роботи зі складними залежностями. Це моделі, котрі складаються з шарів штучних нейронів, та обробляють дані, виявляючи приховані залежності [3]. Зокрема, застосування багатоварових НМ дозволяє класифікувати навіть великі та нелінійно розділені дані (рис.1).

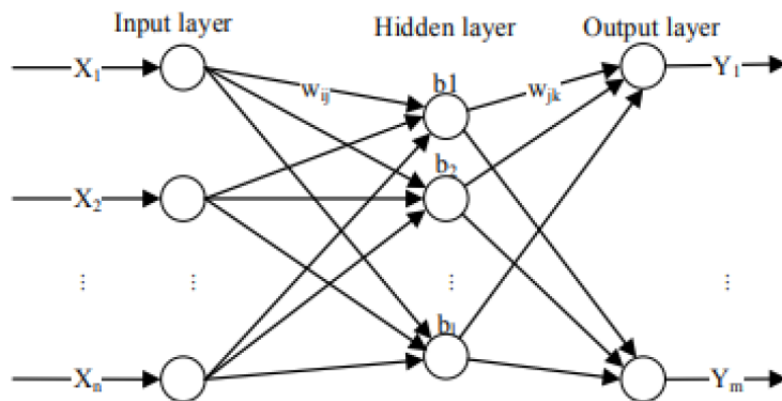


Рисунок 1. Архітектура трьохшарової НМ з прихованим та вихідним шаром нейронів [4]

Ефективним методом для обробки багатовимірних і багатофакторних даних є алгоритм випадкових лісів. Це ансамблевий метод, що створює безліч дерев рішень і об'єднує їхні результати. Зокрема, кожне дерево аналізує випадкову підмножину даних. Важливим є те, що моделі стійкі до шуму та перенавчання. Тоді як метод підсилених дерев послідовно будує дерева, кожне з яких виправляє помилки попереднього та дозволяє отримати високу точність класифікації.

Метод опорно-векторних машин застосовують для класифікації даних із чіткими межами між класами. Зокрема, алгоритм k -найближчих сусідів є простим і зрозумілим методом для класифікації нових даних за схожістю до навчальної вибірки та класифікує нову точку за класами найближчих до неї сусідів [5].

Перевагами застосування методів машинного навчання є швидка обробка даних у реальному часі, покращення точності контролю, зокрема, зменшення помилок у класифікації, підвищення ефективності виробничих процесів, тобто, автоматизовані рішення замість ручного аналізу та зниження витрат на обслуговування, а саме, раннє виявлення несправностей та їх запобігання.

Відомо, що у випадку роботи з матеріалами, які зазнають циклічних навантажень, особливо важливо проводити класифікацію даних, пов'язаних із втомним руйнуванням. Втомне руйнування - це поступове пошкодження матеріалів під дією повторюваних навантажень, що може призводити до катастрофічних наслідків у конструкціях. Для розв'язування таких задач, скористувались експериментальними даними для алюмінієвого сплаву Д16Т [6]. Як свідчать результати, наведені у працях [2, 7], методи машинного навчання дозволяють з високою точністю прогнозувати дані.

Методи машинного навчання, такі як нейронні мережі, випадкові ліси, підсилені дерева, опорно-векторні машини та метод k -найближчих сусідів, дозволяють ефективно вирішувати задачі класифікації даних у комп'ютеризованих системах керування технологічними процесами. Застосування цих методів сприяє підвищенню продуктивності, безпеки та надійності роботи систем у сучасній промисловості.