

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження моделей штучного інтелекту оцінювання родючості ґрунту
для ефективного вирощування сільськогосподарських культур
(тема)

Виконав:
студент 2 курсу, групи СШМ-22-2
Доценко М.В.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)

Керівник проф. Смеляков К.С.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____
(підпис)

В.О. Філатов
(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)
Кафедра _____ Штучного інтелекту _____
(повна назва)
Рівень вищої освіти _____ другий (магістерський) _____
Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)
Тип програми _____ освітньо-наукова _____
(освітньо-професійна або освітньо-наукова)
Освітня програма _____ Системи штучного інтелекту _____
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
« _____ » _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові _____ Доценку Микиті Віталійовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Дослідження моделей штучного інтелекту оцінювання родючості ґрунту для ефективного вирощування сільськогосподарських культур _____

затверджена наказом університету від 1 квітня 2024 р. № 260Ст

2. Термін подання студентом роботи до екзаменаційної комісії 5 червня 2024 р.

3. Вихідні дані до роботи _____ Науково-технічні публікації, дані Інтернет-джерел та відомих наукових проектів які використовуються для аналізу ґрунтів, набори даних зображень з підписами, набори даних з атрибутами ґрунтів для хімічного аналізу ґрунтів _____

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі і постановка задачі дослідження


2) Огляд способів дослідження ґрунтів

3) Створення прототипів моделей для отримання результатів в роботі з ґрунтами

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	01.04.2024	виконано
2	Аналіз предметної галузі	01.04.2024 – 05.04.2024	
3	Постановка задачі	08.04.2024 – 12.04.2024	виконано
4	Порівняльний аналіз існуючих підходів щодо аналізу ґрунтів із застосуванням машинного навчання	15.04.2024 – 19.04.2024	виконано
5	Розробка підходів та способів досліджень ґрунтів	22.04.2024 – 03.05.2024	виконано
6	Створення експериментальних моделей для отримання можливих результатів	06.05.2024 – 17.05.2024	виконано
7	Аналіз отриманих результатів результатів	20.05.2024 – 22.05.2024	виконано
8	Пропозиції щодо реалізації прототипу	23.05.2024 – 24.05.2024	
9	Написання пояснювальної записки	01.05.2024 – 31.05.2024	виконано
10	Попередній захист	31.05.2024	виконано
11	Захист перед ЕК	05.06.2024	виконано

Дата видачі завдання 1 квітня 2024 р.

Студент 
(підпис)

Керівник роботи _____ проф. Смеляков К.С.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 64 с., 26 рис., 3 табл., 1 дод., 29 джерел.

МОДЕЛІ НЕЙРОННИХ МЕРЕЖ, НЕЙРОННІ МЕРЕЖІ, ОБРОБКА ДАНИХ, СІЛЬСЬКЕ ГОСПОДАРСТВО, ХІМІЧНИЙ АНАЛІЗ, ШТУЧНИЙ ІНТЕЛЕКТ.

Предмет дослідження – методи оцінки родючості ґрунтів за допомогою моделей штучного інтелекту, навчання цих моделей та оцінка якості отриманих результатів.

Об'єкт дослідження – процес оцінки ґрунтів за допомогою моделей на основі датасетів з даними про ґрунт та зображень поверхностей ґрунтів, особливості підходів та їх використання в польових умовах.

Мета роботи – вивчення та застосування моделей штучного інтелекту для хімічного та візуального аналізу ґрунтів. Отримання результатів моделі та аналіз їх відповідності до реальності.

Методи дослідження включають методи машинного та глибинного навчання для аналізу даних про ґрунти, аналіз науково-технічної літератури в галузі хімічного та поверхневого аналізу ґрунтів, а також порівняльний аналіз існуючих рішень, які вже представлені на ринку.

У цій роботі розглядається процес аналізу ґрунтів за допомогою моделей штучного інтелекту та порівняння різних методів аналізу ґрунтів для оцінки їх родючості, порівняння різних підходів та можливість їх використання у польових умовах.

ABSTRACT

Master's thesis contains: 64 pp., 26 fig., 3 tabl., 1 ann., 29 references.

AGRICULTURAL, ARTIFICIAL INTELLIGENCE, CHEMICAL ANALYSIS, DATA PROCESSING, NEURAL NETWORK MODELS, NEURAL NETWORKS.

The subject of the study is methods of soil fertility assessment using artificial intelligence models, training of these models, and evaluation of the quality of the results obtained.

The object of research is the process of soil assessment using models based on datasets with soil data and images of soil surfaces, the features of approaches and their use in the field.

Purpose – to study and apply artificial intelligence models for chemical and visual soil analysis. Obtaining model results and analyzing their correspondence to reality.

Research methods include machine learning and deep learning methods in relation to the problem under study, analysis of scientific and technical literature in the field of chemical analysis and surface analysis of soils, and comparative analysis of available solutions already on the market.

This paper discusses the process of soil analysis using artificial intelligence models and compares different methods of soil analysis to assess soil fertility, compares different approaches and the possibility of their use in the field.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ	8
1 Аналіз предметної галузі	10
1.1 Актуальність використання машинного навчання та штучного інтелекту в сільськогосподарській галузі.....	10
1.2 Опис впливу стану ґрунту на вирощування сільськогосподарських культур.....	12
1.3 Аналіз існуючих рішень в сфері оцінки ґрунту.....	13
1.4 Моделі штучного інтелекту та збір даних	20
1.5 Різновиди видів машинного навчання.....	26
2 Постановка задачі	28
2.1 Задача роботи.....	28
2.2 Збір та аналіз даних	28
3 Створення моделі для хімічного аналізу родючості ґрунтів.....	33
3.1 Кореляційний аналіз.....	33
3.2 Створення моделі машинного навчання для хімічного аналізу ґрунту	36
3.3 Опис датасету та моделі для проведення дослідження на основі даних, отриманих з дронів	45
3.4 Згорткові мережі.....	49
4 Сценарії використання моделей.....	55
4.1 Практичне використання моделей у створенні рішень	55
Висновки.....	58
Перелік джерел посилання	60
Додаток А Відомість кваліфікаційної роботи.....	64

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

Горизонт Bhs – горизонт ґрунту, який зазнав ілювації (накопичення матеріалів, вимитих з попередніх шарів) і містить значну кількість сесквіоксидів (оксидів заліза та алюмінію);

Горизонт E – мінеральний горизонт, в якому органічна речовина була вимита, залишивши концентрацію частинок піску і мулу. Він часто світліший за кольором, ніж нижчі горизонти;

Горизонт Oa – це поверхневий або органічний горизонт мінерального ґрунту, що складається переважно з органічного матеріалу на різних стадіях розкладання. Він багатий на органічні речовини та поживні речовини;

ШІ – штучний інтелект;

Ca, Mg, K – основні катіони кальцію (Ca), магнію (Mg) і калію (K), які є необхідними поживними речовинами для росту рослин;

Cd-горизонт – горизонт у ґрунтовому профілі, який представляє собою невивітрювану материнську породу або орний шар. Він має низький рівень утримання катіонів і робить незначний внесок у родючість ґрунту;

GAN – Generative Adversarial Network – генеративно-змагальні мережі;

4IR – The Fourth Industrial Revolution – четверта промислова революція.

ВСТУП

Родючість ґрунту – це здатність ґрунту підтримувати ріст і розвиток рослин. Ця характеристика залежить від наявності в ґрунті необхідних рослинам поживних речовин, вологи та повітря, а також від структури ґрунту та наявності мікроорганізмів. Родючість ґрунту може бути підвищена за допомогою застосування органічних та мінеральних добрив, компосту, а також методів, які підвищують вміст органічного матеріалу в ґрунті, наприклад, збереження решток рослинного походження (поживних решток), обробка ґрунту з використанням зеленої маси тощо [1]. За визначенням родючість ґрунтів це сукупність факторів таких як природні та штучні. До природних можемо віднести – тип ґрунту, вологість, кліматичні умови та наповненість поживними речовинами, до штучних – забезпечення ротації культури яка висаджується, система поливу, використання добрив та своєчасність виявлення проблем які пов'язані з природними чинниками такими як хвороби і усунення цих проблем. Основними критеріями оцінювання самого ґрунту є: хімічний аналіз (лужність ґрунтів, поглинальні властивості, буферність ґрунтів) та вимірювання рН (оцінка кислотності ґрунтів, яка впливає на розчинність поживних речовин, якщо рівень рН занадто низький, то знижується кількість поглинутих макро- та мікроелементів, тобто добрив).

Штучний інтелект може застосовуватись для аналізу даних про стан ґрунту на місці, тобто отримуючи дані з датчиків вологості або реагуючи на інші зміни погоди такі як підвищення сонячного світла. Також можна автоматизувати лабораторії по дослідженню мікро- та макро-елементів. Розуміючи як культура росте та отримуючи постійні данні про стан ґрунту та зовнішні чинники можна вчасно зрозуміти як покращити стан довколишнього середовища для культури або як вплинути на адаптацію до цього стану. Традиційно якість ґрунту та здоров'я врожаю визначалися на основі людських спостережень та суджень. Але цей метод не є ані точним,

ані своєчасним. Натомість тепер можна використовувати дрони (БПЛА) для зйомки аерофотознімків і навчати моделі комп'ютерного зору використовувати їх для інтелектуального моніторингу стану посівів і ґрунту. ШІ візуального зондування може аналізувати та інтерпретувати ці дані, щоб: відстежувати стан посівів, робити точні прогнози врожайності, виявляти нестачу поживних речовин набагато швидше, ніж людина, ШІ-моделі можуть інформувати фермерів про конкретні проблемні зони, щоб вони могли вжити негайних заходів [2]. Головним чинником для інтеграції ШІ в агрономію є якраз своєчасна реакція на різні ситуації, які не контролюються людиною, наприклад, критична вологість для насіння пшениці, жита і ячменю становить 14,5–15,5%, для кукурудзи – 13–14%, проса – 12–13%, соняшнику – 7–9%, ці та інші чинники впливають на врожайність і не завжди можуть бути вчасно визначені [3].

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Актуальність використання машинного навчання та штучного інтелекту в сільськогосподарській галузі

Штучний інтелект став цікавий широкому загалу не так давно і одразу ж люди почали впроваджувати його в усі галузі де можна оптимізувати витрати або збільшити виробництво. Найшвидше ШІ впроваджують у ІТ галузі, наукових дослідженнях та у галузі фін-теху, ці дані підтверджує графік з дослідження Artificial Intelligence Index Report 2024 (рисунок 1.1).

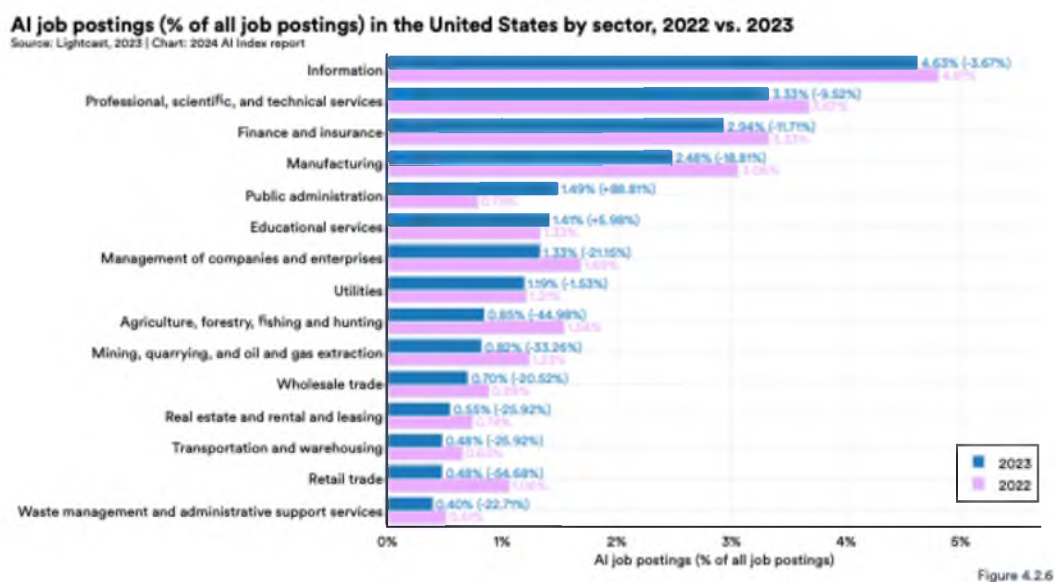


Рисунок 1.1 – Графік створених вакансій для AI спеціалістів за галузями у США за 2022-2023 [4]

Сільськогосподарська галузь не є останньою, але необхідність в спеціалістах, що мають навички роботи з ШІ, невисока. Запровадження AI у галузі сільського господарства допоможе боротися з багатьма викликами які з'являються зараз – зростання споживання, мінімізація впливу на

довкілля та збільшення продуктивності при скороченні ресурсів (рисунок 1.2).



Рисунок 1.2 – Стан запровадження технології в агросекторі [5]

Зараз ШІ використовується для аналізу великих обсягів даних з різних джерел, таких як супутникові знімки, дрони та сенсори у ґрунті. Ці засоби допомагають зі своєчасною реакцією на зміну стану ґрунту, водного балансу або стану самої рослини, що допомагає оптимізувати кількість добрив що використовуються, води так зменшує витрати на обробку полів.

Підхід прецизійного землеробства включає в себе використання ШІ, після того як він набув популярності, це дозволяє оптимізувати витрати, та максимізувати врожай і зменшити екологічний вплив галузі на природу шляхом більш точного аналізу ґрунту, що дозволяє зменшити застосування штучних чинників родючості, тобто добрив. Маючи запас даних отриманих з інших полів можна отримати модель ШІ, яка зможе прогнозувати та давати рекомендації для подальших дій фермеру.

Є багато кейсів застосування для ШІ у галузі сільського господарства, такі як автоматичний полив або прогнозування необхідності застосування добрив, що допоможе зекономити ресурси.

Ці переваги та покращення можливі тільки якщо буде подолано деякі виклики, що пов'язані з високою вартістю технологій, необхідністю

навчання фермерів, якщо програми треба буде обслуговувати самим або створення ринку для спеціалістів, що будуть працювати з фермерами, тому що молодих фермерів хто цікавиться новими технологіями небагато, а з фермерами старого гарду це буде важко.

1.2 Опис впливу стану ґрунту на вирощування сільськогосподарських культур

Ґрунт має різні хімічні властивості, котрі впливають на його родючість та які культури можна використовувати для отримання максимального результату. Почнемо з розгляду горизонтів ґрунтів. Кислотність ґрунту є ультра кислою в горизонтах Oa, E і Bhs при вимірюванні в сольових розчинах, і є дуже кислою або дуже сильно кислою в інших горизонтах і у всіх горизонтах при вимірюванні у воді (Soil Survey Staff, 1993). Висока кислотність цього ґрунту обмежує тип рослинності, яка може розвиватися на цих ґрунтах, обмежується видами, які лише незначною мірою чутливі до низького рівня рН.

Вміст основних катіонів у ґрунті для використання рослинами є найвищим у горизонті Oa. Втрата органічного горизонту призведе до дефіциту основних поживних речовин у ґрунті. Як SES (Soil Evaluation Score), так і ESES (Easily Soluble Electrolytes of Soil) відображають високу кислотність, виявлену при аналізі за допомогою будь-якого методу оцінки кислотності, показуючи, що основним катіоном, який присутній у ґрунті, є алюміній. Ті хто обробляє ці ґрунти повинні докладати всіх зусиль, щоб утримати якомога більше Oa, щоб забезпечити рослини поживними речовинами і рослинний покрив.

Мінеральні горизонти демонструють низький рівень утримання поживних катіонів Ca, Mg і K. Оскільки горизонт Cd має найнижчий рівень утримання катіонів, від подальшого вивітрювання орного шару можна очікувати незначного внеску в родючість ґрунту [6]. З цього аналізу бачимо,

що через хімічний аналіз ґрунту можна отримати розуміння того що є культури котрі приживаються в одному ґрунті, але в іншому можуть мати необхідність у додаткових добривах. Через різні типи ґрунту одна й та ж культура може давати різні результати в момент збору врожаю, тому що рослини можуть не до отримати певні хімічні елементи необхідні для росту та плодоносіння. Постійна ротація культур може допомогти у збереженні певного набору хімічних елементів у ґрунті. Хімічні властивості ґрунту, такі як його кислотність, мають суттєвий вплив на його родючість та придатність для вирощування рослин. Ультра кислі горизонти ґрунту обмежують тип рослинності, що може на них рости, переважно вказуючи на види, які менше чутливі до низького рівня рН. Втрата органічного горизонту може призвести до дефіциту поживних речовин у ґрунті, тому важливо докладати зусиль для його збереження.

Фермерам, що обробляють ці ґрунти з високою кислотністю, рекомендується зосередитися на збереженні органічного горизонту для забезпечення рослин поживними речовинами та рослинним покривом. Мінеральні горизонти ґрунту можуть мати низький рівень утримання поживних катіонів, що може обмежувати внесок у родючість ґрунту внаслідок подальшого вивітрювання орного шару.

Ці висновки підкреслюють важливість збалансованого управління хімічними властивостями ґрунту для забезпечення високих врожаїв та стійкості екосистем.

1.3 Аналіз існуючих рішень в сфері оцінки ґрунту

МакКініон та Леммон у 1985 році були піонерами у впровадженні штучного інтелекту в аграрну сферу, розробивши GOSSYM, комп'ютерну модель для імітації процесу вирощування бавовни. Вони використали експертну систему, щоб удосконалити виробництво, враховуючи такі чинники, як іригація, добрива, контроль над бур'янами, кліматичні умови та

інші агротехнічні заходи [7], [8]. GOSSYM – це динамічна модель балансу маси, яка імітує процеси кругообігу вуглецю, азоту і води в кореневій зоні рослин і ґрунту протягом усього життєвого циклу бавовнику. Вона прогнозує ріст культури (з детальною хімією рослин, морфогенезом і фенологією) і реакцію ґрунту на екологічні стреси, в першу чергу, пов'язані з теплом, водою, вуглецем і поживними речовинами (рисунок 1.3). Ці стреси визначаються кліматичними змінними, такими як сонячна радіація, температура, кількість опадів і вітер, а також властивостями ґрунту і культурними практиками, включаючи зрошення і внесення добрив [9].

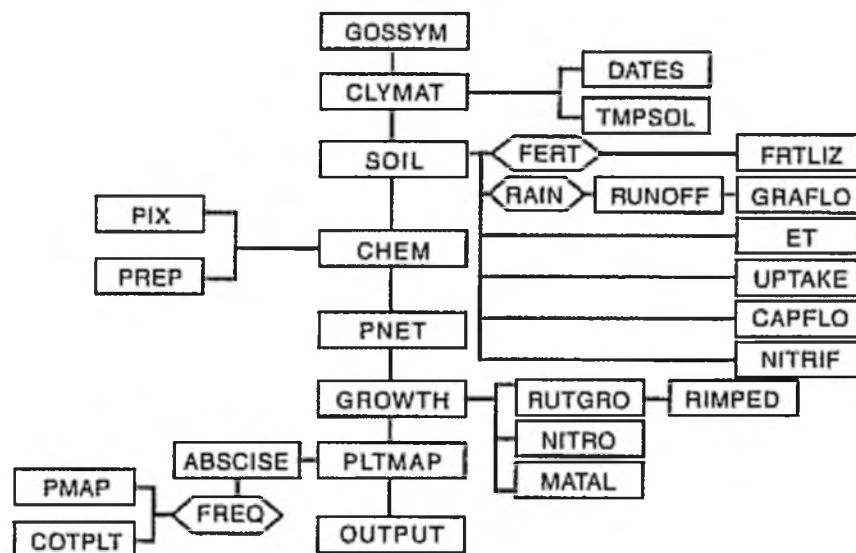


Рисунок 1.3 – Схема експертної системи GOSSYM

Ця модель показує, що ще в 80-х роках створювались перші системи для оптимізації та збільшення урожаю за допомогою аналізу даних. Використовуючи наявні ресурси, тобто комп'ютери та мову програмування Fortran було створено програму для ефективної обробки бавовни.

З розвитком технологій в агрономії теж відбувались певні зміни, які допомагали з підвищенням ефективності вирощування та збільшенням врожайності. Наприклад, стартапи які виникли під час 4IR (англ. The Fourth Industrial Revolution, також англ. Industry 4.0, нім. Industrie 4.0, укр.

Промисловість 4.0 – поняття, що означає розвиток і злиття автоматизованого виробництва, обміну даних і виробничих технологій в єдину саморегульовану систему, з як найменшим або взагалі відсутнім втручанням людини у виробничий процес) [10] (рисунок 1.4).

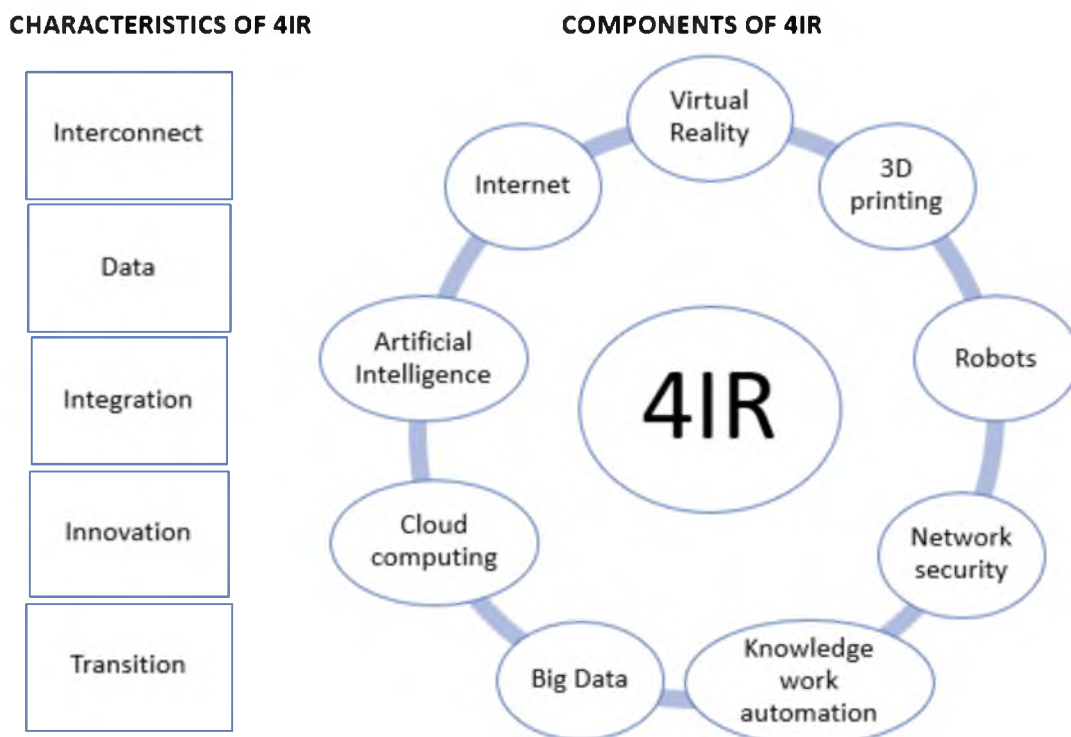


Рисунок 1.4 – Характеристики та Компоненти 4IR [11]

Один із таких стартапів є Farmerline, що базується в Гані, – це SMS-сервіс, який надає малим фермерам актуальну інформацію та поради що стосуються вирощування сільськогосподарських культур. Ця компанія має власну платформу Merqdata, яка дозволяє фермерам та покупцям продукції отримувати дані про якість продукції, передбачати об'єми врожаю та отримувати сертифікацію товару. Основним мінусом цього сервісу є те, що дані збираються за допомогою самих фермерів, але маючи великі обсяги даних вже можна навчати моделі, які будуть ефективнішими ніж GOSSYM. Нажаль до їх даних не можна отримати доступ навіть через запит, або мати

можливість отримати більше інформації ніж пропонує сайт стартапу. Ще одним стартапом є M-Farm, який також використовує дані для надання інформації про погоду та ціни на сільськогосподарські культури. M-Farm аналізує десятирічні дані з Системи раннього попередження голоду (Famine Early Warning System) набір даних агрокліматології [12]. Основними плюсами обох систем є те що вони працюють с невеликими фермерством у країнах за обмеженими ресурсами та кліматом який не зовсім підходить для сільськогосподарської діяльності. Для підтвердження того, що ґрунти Кенії та інших країн центральної Африки погано підходять для сільськогосподарської діяльності було створено порівняльну таблицю 1.1 з Україною.

Таблиця 1.1 – порівняння стану сільськогосподарських галузей України та Кенії

Україна	Кенія
Основний тип ґрунтів – це родючі чорноземи, які мають високий вміст органічних речовин та поживних елементів, що дозволяє вирощувати велике різноманіття сільськогосподарських культур	Має велике різноманіття ґрунтів, яке включає в себе червоні та сірі вологі тропічні ґрунти, які мають менший вміст органічних речовин та є менш родючими. Деякі регіони страждають від ерозії та засолення.
Клімат переважно помірний, що впливає на сезонність робіт та вибір культур. В зимовий період необхідний спеціальний підхід до оброки ґрунту і це диктує які культури можуть бути використані (озима пшениця).	Знаходиться у тропічному поясі з чітко вираженим сухим та вологим сезонами. Це позитивно впливає на сільське господарство з точки зору того що можна проводити землеробство напротязі всього року, але потребує додаткової адаптації до засушливих сезонів.

Продовження таблиці 1.1

Використовують наявні технології, штучний полив, іригацію, добрива.	Залежать від природних опадів для поливу.
Має розвинуту аграрну інфраструктуру та інвестує в аграрний комплекс, що сприяє розвитку аграрних підприємств (нажаль не зовсім актуальний пункт наразі, через великі збитки завдані війною).	Бороться з високим рівнем бідності серед сільського населення, що впливає на доступність ресурсів для фермерів та обмежує їх в можливості модернізації або розширення списку культур для вирощування.
Переважають нейтральний або слаболужний рН, особливо чорноземи. Нейтральний рН сприяє доступності більшості поживних речовин для рослин.	Переважають кисли або дуже кислий рН. Кислий рН може обмежувати доступність деяких поживних речовин і вимагати додаткової обробки перед посівом.
Багаті на калій, фосфор, кальцій та магній	Мають дефіцит основних поживних речовин, особливо фосфору та нітрогену, що вимагає регулярного застосування добрив.

Можливо, Кенія не найкраща країна для порівняння з Україною, але стартапи про які згадувалося раніше працюють саме там. Беручи до уваги порівняння з таблиці, можна отримати розуміння того, що стартапи та технології перш за все потрібні слаборозвинутим країнам, які мають значні мінуси для ведення сільськогосподарської діяльності у зв'язку з місцем положення. Використовую хімічний аналіз ґрунтів на таких територіях, можна отримати розуміння які культури потрібно розвивати та яким чином нівелювати наявні мінуси. Проблема, яку важко помітити – це відсутність різноманітності культур, що погано впливає на ґрунти та виснажує їх. Тож маючи такі проблеми, багато стартапів будуть приділяти увагу саме

територіям з проблемами, щоб зробити їх прибутковими за допомогою вирощування підходящих культур.

Наступний приклад це – InCeres, які розробили додаток для оцінки якості ґрунту та його родючості. Для оцінки використовують дані про хімічний склад, погоду, види культур та знімки з супутників. За словами Леонардо Менегатті, провідного наукового співробітника InCeres, традиційний спосіб хімічного аналізу коштуватиме фермеру 200 реалів(бразильський реал) за акр протягом 10 років, тоді як новий підхід коштуватиме лише 40 реалів за той самий період, заощаджуючи 80%. Технології вже зараз дають відповідь на питання, як зробити сільськогосподарський бізнес прибутковим. У міру розвитку програми вона буде вивчати та передбачати якість ґрунту в майбутньому [13] (рисунок 1.5).



Рисунок 1.5 – Клієнтська платформа InCeres

Цей стартап також оптимізує та допомагає агросектору у країні де є проблеми з ґрунтами. Тут вже можна бачити, що проводиться хімічний аналіз та все це відображено у зрозумілому інтерфейсі та дає змогу отримати інсайти щодо того які культури будуть показувати найкращі показники врожаю виходячи зі статистики.

Найкращий приклад системи для допомоги фермерам – це EOSDA Crop Monitoring. EOSDA Crop Monitoring – це платформа для дистанційного

моніторингу стану ґрунтів і сільськогосподарських культур за допомогою супутникових знімків. Вона дозволяє аналізувати вологість, хімічний склад ґрунту, рівень рН та інші важливі показники. Платформа використовує індекси NDVI і NDMI для оцінки стану посівів і прогнозування врожайності, надаючи фермерам можливість приймати обґрунтовані рішення щодо управління полями. Це допомагає оптимізувати внесення добрив, зрошення та покращити загальну продуктивність сільськогосподарських культур. Аналіз якості ґрунту дозволяє зрозуміти стан поля, але для отримання повної картини важливо використовувати кілька джерел даних. У цьому контексті корисною є функція зонування на платформі EOSDA Crop Monitoring. З її допомогою можна створювати карти продуктивності, які показують динаміку стану полів за допомогою вегетаційних індексів, таких як NDVI. Це дозволяє точніше визначити зони для відбору проб і застосовувати агрохімікати залежно від стану культур та характеристик кожної зони (рисунок 1.6).

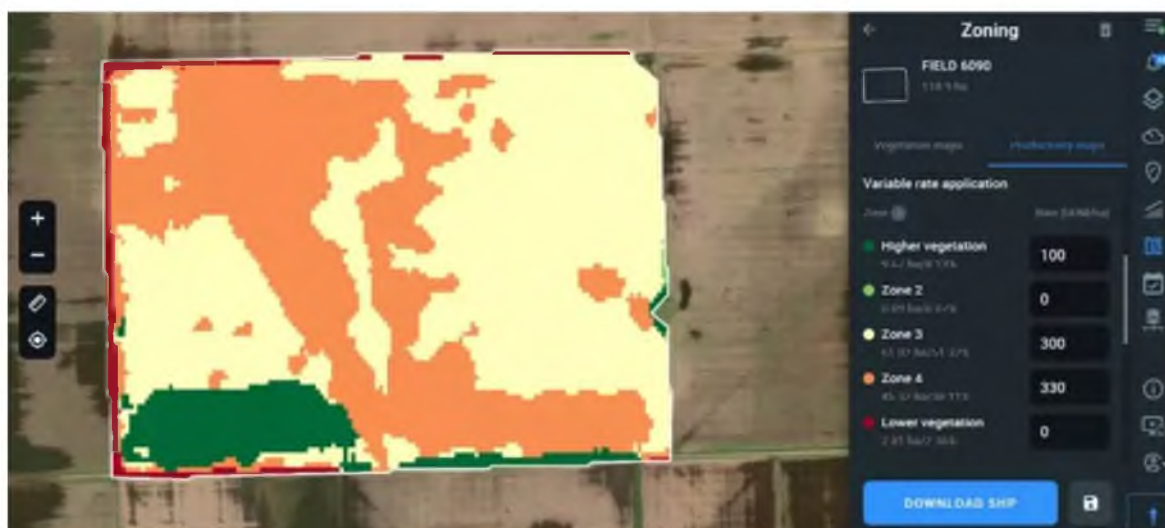


Рисунок 1.6 – Карта продуктивності кукурудзяного поля та рекомендації щодо диференційованого внесення добрив

На прикладі кукурудзяного поля, поділеного на п'ять зон за індексом NDVI, рекомендується проводити відбір зразків ґрунту в зонах з низькою

інтенсивністю вегетації, особливо у «Зоні 3» та «Зоні 4». Зона з менш інтенсивною вегетацією, позначена червоним кольором, є краєм поля, де нічого не росте, а «Зона 2» замала для відбору проб. Рекомендації щодо внесення добрив надаються для кожної зони залежно від інтенсивності вегетації. Наразі ще є дослідження, які допомагають отримати дані про вологість ґрунтів за допомогою відеокамер. Цей аналіз робиться за допомогою розподілу ґрунтів за кольором, щоб потім отримати дані про вірогідну кількість вологи.

1.4 Моделі штучного інтелекту та збір даних

У агросекторі є низка викликів для вирішення яких можна застосувати технології для підвищення продуктивності та сталого використання ресурсів. Зміни клімату, що можуть зміщувати сезонність і фермерам треба буде адаптувати графік до нових реалій; зростання населення, що призводить до криз та потреби оптимізувати шляхи доставки продовольства та збільшення обсягів виробництва, що у свою чергу має вплив на клімат та необхідність збільшення врожайності без шкоди для довкілля, бо не тільки клімат змінюється, а й структура ґрунту, якість води та кількість поживних речовин у ґрунтах. Для цього добре підходять технології штучного інтелекту, тому що маючи достатньо даних для аналізу, технології можуть допомогти фермерам приймати обґрунтовані рішення, оптимізувати кількість добрив та своєчасно реагувати на зміни у середовищі де знаходиться культура.

Моделі ШІ, зокрема машинне навчання та глибоке навчання, використовуються для ідентифікації зразків у великих даних, що включають зображення з супутників, які необхідні для поверхневого аналізу аби зрозуміти де знаходиться досліджувана ділянка та застосувати до знімків дані з датчиків, що збирають відомості про кількість поживних речовин та кислотність ґрунтів. Також використовуються історичні дані про родючість

та хімічні показники, наприклад, як вже було згадано про Системи раннього попередження голоду (Famine Early Warning System), які мають зібрані дані про те які були умови для врожаю та як це вплинуло на нього.

За допомогою датасетів з таких систем можна створити моделі, що будуть прогнозувати оптимальні умови для вирощування врожаю, аналізувати родючість ґрунту та передбачати врожайність, а також це свого роду може бути рекомендаційним механізмом для фермера, тому що раніше згадані стартапи мають щось подібне, але через недостатню кількість відкритих даних не можна сказати, чи використовують ці стартапи свої власні моделі, чи покладаються на експертну думку людини.

Ефективність таких моделей може бути іноді не краще за експертну думку людини, тому що залежить від кількості даних на яких вони були натреновані. Збір точних репрезентативних даних є критичним для створення надійних моделей. Це включає детальний збір інформації про стан ґрунтів, кліматичні умови, водні ресурси та інші параметри.

Дані повинні бути актуалізовані та масштабовані таким чином, щоб вони могли бути застосовані для реальних умов, це значить що не можна проводити експеримент у вакуумі і треба брати до уваги не тільки хімічний аналіз, а й інші чинники. Бажано мати показники для різних ґрунтів за різних сезонів, бо вони можуть потребувати різних заходів для покращення родючості, от як наприклад вже описані червоні ґрунти можуть потребувати вапнування аби знизити рН перед посівом, тому для створення моделі може бути необхідно багато різних параметрів, які залежать від середовища та культури яка сіється.

Використання ШІ в агрономії значною мірою залежить від здатності моделей аналізувати великі масиви даних і визначати корисні зразки для прийняття рішень. Розглянемо три основних типи моделей ШІ, які застосовуються для таких завдань:

- нейронні мережі (Neural Networks) – це моделі, що наслідують структуру та функціонування людського мозку. Вони складаються з вузлів,

або «нейронів», які з'єднані між собою і можуть обробляти вхідні дані для отримання рішення. Такі мережі особливо ефективні для розпізнавання зразків та обробки зображень, що підходить для аналізу супутникових знімків та аерофотозйомки ґрунті (рисунок 1.7). Їх можна використовувати для ідентифікації типів ґрунтів, стану рослинності та виявлення хвороби. Наприклад Доктор Алі Аль-Наджі з Університету Південної Австралії та професор Яван Чал співпрацювали над системою, яка аналізує показники ґрунту, такі як вологість за допомогою цифрової камери. Це говорить про те, що створивши датасет належної якості, можна отримати модель, яка зможе пропонувати необхідне зволоження для певної культури виходячи тільки зі знімків за допомогою дрона;

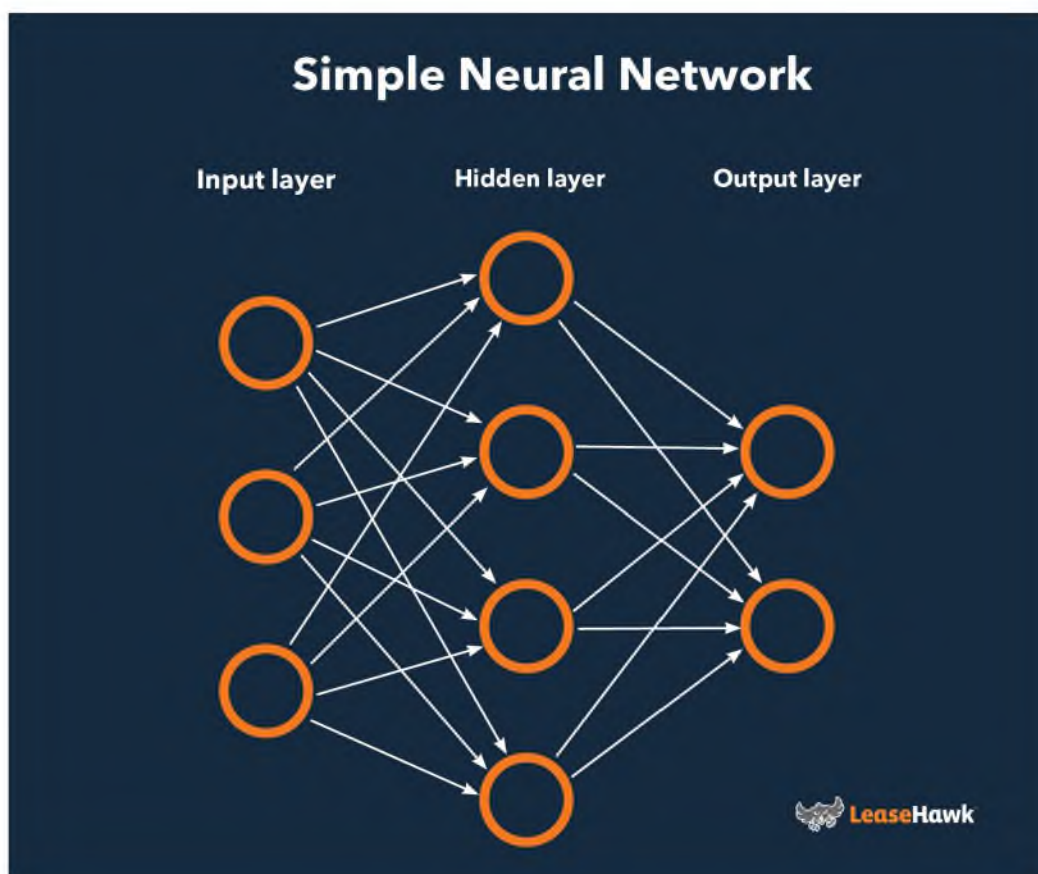


Рисунок 1.7 – Проста нейронна мережа [14]

– машинне навчання (Machine Learning) – це галузь штучного інтелекту, що дозволяє системам вчитися без чітко поставленої задачі. ML моделі тренуються на великих наборах даних і використовують статистичний аналіз для створення прогнозів або прийняття рішень. Можна застосовувати для прогнозування врожайності, аналізу впливу кліматичних умов на процес зростання культур та оптимізувати кількість необхідних добрив або водних ресурсів (рисунок 1.8);

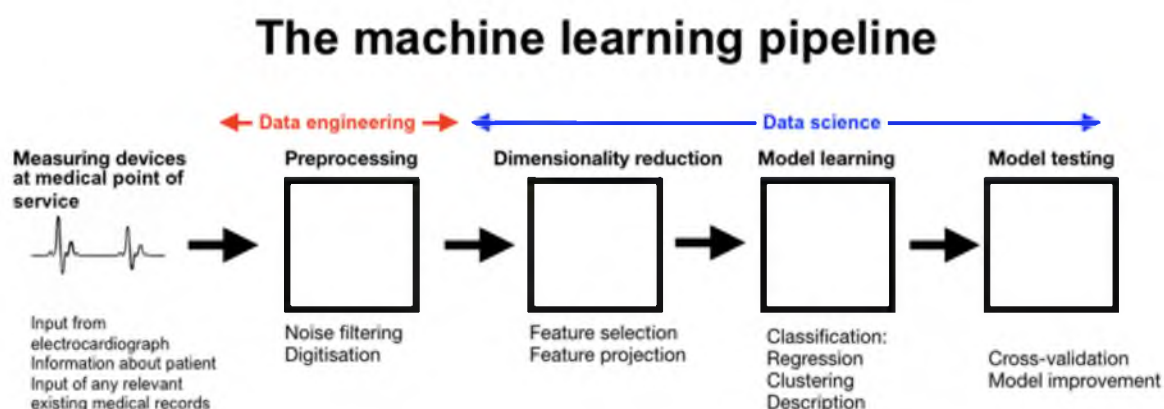


Рисунок 1.8 – Схематична діаграма процесу машинного навчання [15]

– генеративно-змагальні мережі (Generative Adversarial Networks, GANs) – це клас алгоритмів ШІ, що складається з двох мереж: генеративної, що намагається створювати дані, максимально схожі на реальні, та дискримінативної, що намагається відрізнити справжні дані від сфабрикованих генератором. Ці дві мережі змагаються одна з одною, що дозволяє підвищити точність обох. За допомогою ГЗМ можна генерувати дані, яких ще немає у системі, тобто симулювати середовища та отримувати дані для них, щоб мати краще представлення про середовища про які мало відомо (рисунок 1.9).

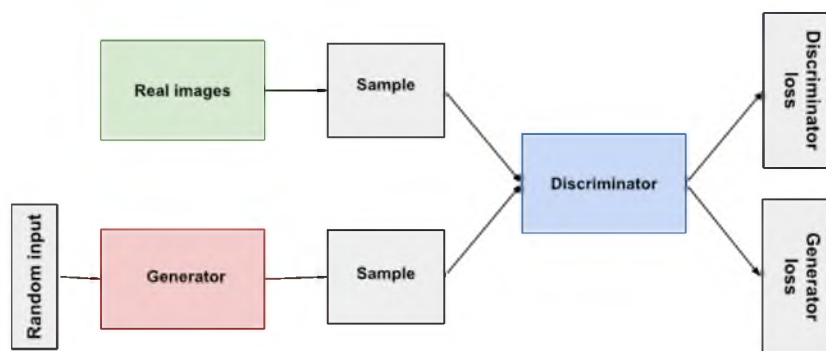


Рисунок 1.9 – Діаграма як влаштована GAN [16]

Необхідний засіб потрібно обирати виходячи з кількості даних та цілі для якої потрібно навчити модель, тому що кожен вид моделі може допомогти сільськогосподарській галузі по-своєму.

Для того щоб будь-яка з раніше зазначених моделей ШІ могла працювати, потрібно зібрати дані. Збір даних є фундаментальним кроком у процесі використання будь-якої моделі незалежно від галузі використання, бо точність даних впливає на якість прийнятих рішень. Основні джерела для збору інформації для оцінки ґрунту:

- супутникові знімки та знімки з дронів – з них можна отримати інформацію про розмір та стан полів, а також про атрибути ґрунту такі як зволоженість, як це було зазначено у прикладі для нейронних мереж раніше та стадії розвитку рослини для отримання більш якісних рішень, бо рослина може потребувати різного догляду на різних стадіях росту. Для збору такої інформації використовуються дрони з камерою великої роздільності, або камери що знаходяться поруч з полями, які можуть збирати інформацію навіть вночі;

- сенсори у ґрунті потрібні для вимірювання різних параметрів таких як кислотність, зволоженість, рівень рН та поживних речовин. Ці сенсори розміщуються на різних глибинах в різних точках поля для забезпечення комплексного збору даних, які можуть бути передані у

реальному часі на сервер, або у локальний хаб для перегляду актуальної інформації та реагування на неї;

– хімічний аналіз ґрунтів дозволяє визначити вміст поживних речовин та забруднювачів, що заважають розвитку рослини. Ці дані допомагають, наприклад, зрозуміти скільки добрив використати. Для збору даних застосовують проби з різних глибин та місць на полі, потім роблять аналізи у лабораторіях, де використовують методи газової хроматографії, мас-спектрометрії та інші;

– дані про погодні умови такі як температура, вологість, кількість опадів та сонячне випромінювання, впливають на ріст рослини та стан ґрунту. Різні ґрунти мають свої проблеми, так і наприклад як різні середовища можуть мати проблеми через недостатню вологість, бо штучний полив не завжди можливий (рисунок 1.10). Для збору інформації про кліматичні умови використовуються кліматичні станції та онлайн-сервіси, які мають як актуальні дані так і історичні про клімат.



Рисунок 1.10 – Приклад сенсору який застосовується для отримання даних про кількість азоту, фосфору та інших речовин у ґрунті [17]

Ці дані генеруються та аналізуються за допомогою моделей ШІ і допомагають у прийнятті рішень та реакції на зміни середовища.

Використання історичних даних допомагає точніше рекомендувати та прогнозувати, що дозволяє адаптуватися до змін умов зі зменшенням витрат на нівелювання проблем.

1.5 Різновиди видів машинного навчання

Машинне навчання є ключовим підходом до розробки систем ШІ. Є декілька основних категорій, що використовуються для навчання моделей.

Контрольоване навчання (Supervised Learning) – у цьому підході система навчається на основі прикладів, які містять вхідні дані та відповідні їм мітки (outputs). Завдання полягає в тому, щоб навчити модель передбачати мітки для нових вхідних даних. Контрольоване навчання широко використовується для класифікації та регресії. Приклади включають розпізнавання зображень та прогнозування цін на акції. У випадку з сільськогосподарською галуззю цю категорію алгоритмів можна використовувати для того щоб оцінювати результати та корегувати їх використовуючи знання людини.

Неконтрольоване навчання (Unsupervised Learning) – у цьому випадку моделі навчаються на вхідних даних без міток. Завдання полягає у виявленні структур або шаблонів у даних. Неконтрольоване навчання використовується для кластеризації, виявлення аномалій та зниження розмірності даних. Прикладом може бути виявлення шаблонів на фото для категоризації об'єктів на фото, для подальшого отримання карт, які будуть характеризувати поля за різними ознаками і використовуватись для накладання результатів хімічного аналізу ґрунтів.

Навчання з підкріпленням (Reinforcement Learning) – цей підхід базується на взаємодії агента з середовищем, де агент отримує винагороди або покарання за свої дії. Мета полягає в максимізації сукупної винагороди шляхом навчання найкращих стратегій. Навчання з підкріпленням використовується для задач, де рішення приймаються послідовно, таких як

гра в шахи або керування автономними транспортними засобами. Такий підхід можна застосувати для оптимізації маршруту обльоту дроном полів, тому що як із хімічним аналізом тут теж треба формулювати шаблон маршруту який був би оптимальним для отримання даних про стан культур та ґрунту.

Напівконтрольоване навчання (Semi-Supervised Learning) – це комбінація контрольованого та неконтрольованого навчання, де використовується невеликий набір міток і великий набір немічених даних. Цей підхід дозволяє значно підвищити точність моделі при обмежених ресурсах для анотації даних.

Навчання з переносом (Transfer Learning) – у цьому підході модель, навчена на одній задачі, використовується як основа для навчання на іншій, пов'язаній задачі. Це дозволяє зекономити обчислювальні ресурси та час на навчання нових моделей. Навчання з переносом широко використовується в обробці зображень та природної мови. Для кожного поля, або ферми можна створювати власні локальні моделі, які б були навчені з початкових моделей, які були сформовані з доступних даних або зібраних раніше. Це зробило б моделі більш релевантними до місця використання.

Форми штучного інтелекту і різновиди видів навчання є основою для розвитку інноваційних технологій у різних галузях, включаючи сільське господарство. Розуміння цих концепцій дозволяє ефективно застосовувати ШІ для вирішення складних задач, таких як прогнозування врожайності, оптимізація використання ресурсів та моніторинг стану ґрунтів. Як можна зрозуміти з вище описаного, то для кожного з видів навчання можна знайти шлях застосування, тому що сільськогосподарська галузь має багато граней, які можна покращити, але потрібно пам'ятати, що це лише оптимізація та допомога для покращення результатів. Необхідність виключати людський фактор є, але без потрібних знань, лабораторних аналізів та фін-тюнінгу моделей не буде потрібного результату.

2 ПОСТАНОВКА ЗАДАЧІ

2.1 Задача роботи

Проаналізувавши предметну галузь було отримано висновки про актуальність застосування моделей ШІ у агросекторі та як можна отримати дані щоб застосувати їх для покращення оброки сільськогосподарських культур. Всі представлені аналоги дають лише поверхнево зрозуміти як проводяться дослідження для аналізу ґрунту та його поверхності на родючість та придатність для висадження певних культур.

2.2 Збір та аналіз даних

Збір даних для аналізу ґрунтів вимагає інтеграції інформації з різноманітних джерел. Основним обраним джерелом є датасет, що містить хімічні показники 880 зразків ґрунту. Ці дані мають вимірювання таких елементів як азот (N), фосфор (P), калій (K), рН (кислотність), електропровідність (EC), органічний вуглець (OC), а також мікроелементи, як сірка (S), цинк (Zn), залізо (Fe), мідь (Cu), марганець (Mn) і бор (B).

Для того щоб створити такий датасет необхідно провести багато дослідів у лабораторії. Процес збору проб ґрунту включає кілька етапів, такі як визначення місць відбору проб, тому що одне поле може мати різну структуру ґрунту на поверхності, або на глибі яка є досяжною для культури, що буде висаджена, тому потрібно підійти до вивчення поля серйозно аби отримати репрезентативні дані, наступний етап – це визначити глибину відбору, зазвичай ця глибина не переважає 30 см, але може бути різною, потрібно зважати на тип ґрунту, останній етап – визначення методик відбору проб, для цього використовуються різні прилади які дозволяють взяти проби з потрібної глибини та у потрібному виді.

Для кожного представленого елемента у датасеті є свої методи хімічного аналізу. Для азоту (N) використовується метод К'ельдаля, який вимірює вміст загального азоту в ґрунті шляхом перетворення азоту на амоній, який потім титрується (це лабораторний метод кількісного хімічного аналізу, при якому до аналізованого розчину поступово додають стандартний розчин титранта до досягнення точки еквівалентності для визначення концентрації досліджуваної речовини) або за допомогою використання Іонно-селективних електродів, які використовують для визначення кількості нітратів. Для визначення кількості Фосфору (P) використовують або Молібдатний метод (використання амоній молібдату для утворення реакції з якої потім робляться вимірювання) або Бретонський метод.

Отримання даних про Калій (K) досягається за допомогою полум'яної фотометрії (спалюють зразок та вимірюють випромінення під час процесу) або Іонної хроматографії (аналіз кількості катіонів). Дані про Сірку (S) отримують за допомогою Турбідиметричного методу (вимірюється кількість осаду за допомогою спектрофотометра) або Інфрачервоної спектроскопії (визначають різні форми сірки за допомогою ІЧ-спектроскопії). Такі елементи як Цинк (Zn), залізо (Fe), мідь (Cu), марганець (Mn) і бор (B) визначають шляхом Атомно-абсорбційної спектрофотометрії (аналіз концентрації металів у ґрунті шляхом вимірювання поглинання світла атомами в зразку) або Індуктивно зв'язана плазмова мас-спектрометрія (метод для точного визначення слідових кількостей мікроелементів). Органічний вуглець вимірюють за допомогою Методу Валькі-Блека (окислення проби ґрунту та вимірювання кількості вуглецю, що утворюється) або за допомогою сухого згоряння для якого спалюють зразок та аналізують кількість утворених газів. Також вимірюють властивості такі як Електропровідність за допомогою Кондуктометрії (вимірювання електропровідності розчину для визначення загальної кількості розчинних солей). Ці вимірювання можливі за

допомогою електрохімічних сенсорів, приклади яких вже наводились у попередньому розділі або за допомогою оптичних сенсорів (наприклад камери), які аналізують кількість відбитого світла. Тобто для отримання репрезентативних даних потрібно провести велику кількість дослідів, аби мати можливість дослідити ґрунт. Отримані дані треба досліджувати у спеціалізованих лабораторіях, які можуть виконувати описані вище методи дослідження ґрунтів на вміст хімічних елементів.

Наступним етапом потрібно визначити сітку за якою будуть відбиратися зразки ґрунту. Зазвичай, використовують сітку площею від 0,5 до 2 га, але це залежить від того наскільки точні результати потрібні. Проби потрібно відбирати за сіткою, щоб отримати репрезентативні дані для просторового аналізу, ці дані потім можна використовувати для побудови карт продуктивності, що відображаються родючість ґрунту на певній поверхності. Потрібно враховувати різноманітність ґрунтів, тому що чим більш різноманітний ґрунт на полі, тим більше проб буде необхідно зробити, якщо ж поле загалом не має варібельності, то можна робити мінімум проб і бути певним, що дані в середньому показують ситуацію на ньому.

Для того аби працювати з даними потрібно вміти їх інтерпретувати та мати приблизні дані про те який показник того чи іншого хімічного елементу є нормою. В залежності від того який ґрунт досліджується різні хімічні елементи можуть мати різні норми, наприклад, внесення азоту варіюються залежно від типу ґрунту. Глинисті ґрунти добре утримують поживні речовини, тому азотні добрива можна вносити в менших кількостях (0,6-0,9 кг азоту на 100 м²). Піщані ґрунти погано утримують поживні речовини, тому потрібно частіше вносити азот в менших кількостях (0,8-1 кг азоту на 100 м²). Суглинкові ґрунти мають середню здатність утримувати поживні речовини, тому підходять як стабільні, так і швидкорозчинні добрива (0,7-0,9 кг азоту на 100 м²). Кислі ґрунти потребують добрив, які також нейтралізують кислотність, наприклад,

кальцієву селітру (0,6-0,8 кг азоту на 100 м²). Існують загальні норми для хімічних елементів яких потрібно дотримуватись (таблиця 2.1).

Таблиця 2.1 – Оптимальна концентрація хімічних елементів у ґрунті.

Елемент	Норма, мг/кг (ppm)
Фосфор (P)	25 – 35
Сірка (S)	7 – 15
Цинк (Zn)	1 – 3
Залізо (Fe)	10 – 20
Марганець (Mn)	8 – 11
Мідь (Cu)	0.8 – 1
Калій (K)	165 – 220
Кальцій (Ca)	1400 або вище
Магній (Mg)	100 або вище
Натрій (Na)	80 – 120

Для забезпечення відповідності та репрезентативності даних необхідно перевірити кожен запис на наявність відсутніх значень. Це може бути зроблено за допомогою статистичних методів і візуального аналізу, наприклад, за допомогою графіків розподілу.

Перед аналізом, дані потребують обробки, що включає очищення, нормалізацію та можливо трансформацію. Більш детальний опис кроків:

- очищення даних – виявлення аномальних значень та заповнення відсутніх даних за допомогою статистичних методів, таких як медіана або середнє;
- нормалізація даних – приведення всіх числових значень до одного виду, що необхідно для коректного формування ваги;
- кодування категорій даних – використання методів Label Encoding та One-Hot Encoding для перетворення категорійних даних у

числовий формат, який потім може бути оброблений за допомогою алгоритмів використаних бібліотек.

Аналіз даних проводиться з метою визначення залежності між хімічними показниками та родючістю ґрунту. Для цього можна використати техніки статистичного аналізу та машинного навчання:

- кореляційний аналіз дозволить визначити статистичні зв'язки між різними хімічними показниками та родючістю. Це допомагає ідентифікувати найбільш значущі характеристики для подальшого моделювання;

- машинне навчання дозволить тренувати різні моделі (наприклад нейронні мережі, random forest, алгоритми підсилення) для визначення здатності цих моделей точно прогнозувати родючість ґрунтів на основі доступних хімічних показників.

Після тренування моделей проводиться їх валідація за допомогою крос-валідації або розділенням набору даних на тренувальну та тестову вибірки. Це дозволяє оцінити реальну здатність моделей прогнозувати родючість ґрунтів та виявити найбільш ефективні та точні підходи.

3 СТВОРЕННЯ МОДЕЛІ ДЛЯ ХІМІЧНОГО АНАЛІЗУ РОДЮЧОСТІ ГРУНТІВ

3.1 Кореляційний аналіз

Мета кореляційного аналізу в контексті цієї роботи – визначити, які хімічні показники ґрунту (наприклад, вміст азоту, фосфору, калію, рівень рН, органічний вуглець та інші) мають найбільший вплив на родючість ґрунту. Процедура аналізу:

- збір даних – використання датасету з хімічними показниками 880 зразків ґрунту;
- обробка даних – переконатися, що всі дані числові та заповнені;
- розрахунок кореляційної матриці – використання бібліотек, такі як `pandas` і `numpy` в Python, для розрахунку кореляційної матриці між усіма парами хімічних показників.

Види кореляцій:

- коефіцієнт кореляції Пірсона (r) є мірою лінійного зв'язку між двома змінними. Кореляційний аналіз зазвичай починається з графічного представлення зв'язку пар даних за допомогою діаграми розсіювання. Значення коефіцієнта кореляції варіюються від -1 до +1 [18];
- коефіцієнт рангової кореляції Спірмена (Spearman ρ) – це непараметрична кореляція вимірювань. Він використовується для визначення зв'язку, що існує між двома наборами даних [19].

Для візуалізації кореляцій створюється теплова карта, яка дозволяє швидко оцінити залежність між змінними. Використання бібліотеки `seaborn` в Python допомагає легко створити таку теплову карту (рисунок 3.1).

Цей аналіз допомагає ідентифікувати ключові показники, які мають найбільший вплив на родючість ґрунту (рисунок 3.2).

Теплова карта показує результати кореляційного аналізу між різними хімічними показниками ґрунту та їх вплив на родючість (Output). Кожне

значення в матриці представляє коефіцієнт кореляції між парою змінних, де 1.00 означає ідеальну позитивну кореляцію, а -1.00 – ідеальну негативну кореляцію, 0.00 – відсутність кореляції.

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

df = pd.read_csv('dataset1.csv')
correlation_matrix = df.corr()
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
            fmt=".2f")
plt.title('Теплова карта кореляцій між хімічними показниками ґрунту')
plt.show()
```

Рисунок 3.1 – Лістинг коду для створення теплової карти кореляції

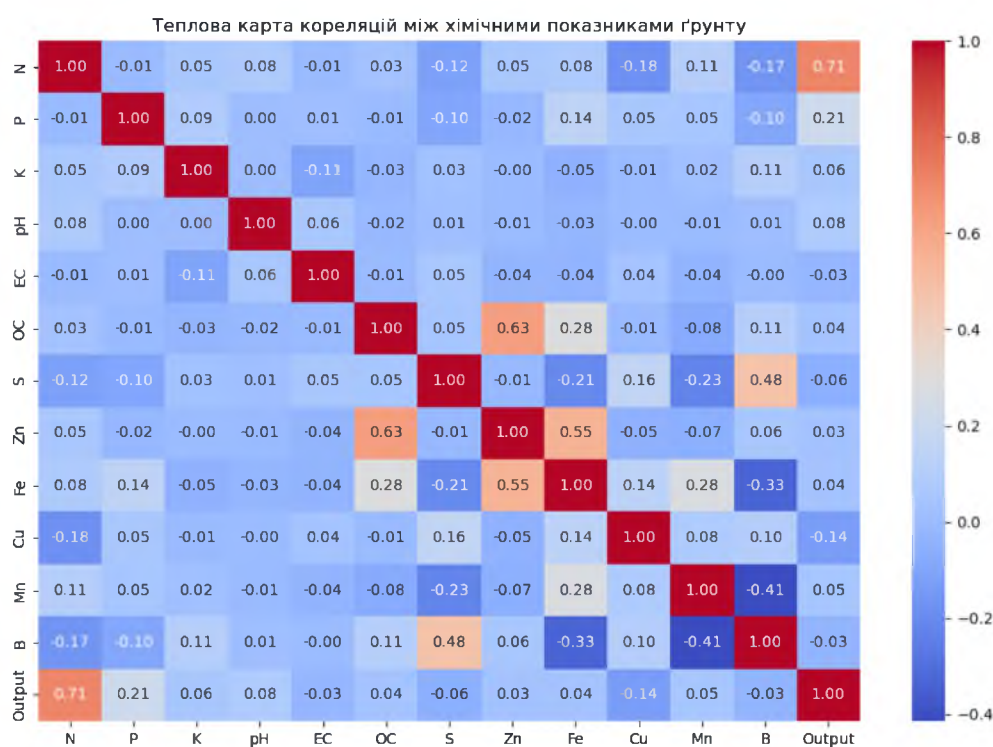


Рисунок 3.2 – Результати кореляційного аналізу датасету

Розглянемо кореляцію з родючістю (Output). N (азот) має найвищу позитивну кореляцію з родючістю (0.71), що підкреслює його значущість для родючості ґрунту. Високий вміст азоту асоціюється з більш родючими ґрунтами, що підтверджується підходом регенеративного землеробства, яке пропонує використання сівозміни для забезпечення постійного ґрунтового покриву, потенціал родючості ґрунту та рівня поживних речовин, якщо в сівозміну включені азотфіксуючі культури [20]. Наприклад, це перевіряли через вирощування бобів разом з іншими культурами, через невеликий відрізок часу перевіряли чи зав'язались боби (якщо розкопати плід, то там є невеликі клубочки), якщо так, то ці клубочки утримують азот та таким чином утримують його у ґрунті, тобто відбувається забезпечення азотом за допомогою природних процесів, без втручання людини, що дозволяє економити кошти на добривах.

Наступний за важливістю елемент за результатами аналізу – P (фосфор), який також має позитивну кореляцію з родючістю (0.21), але менш виражену, що вказує на його важливу роль, хоча і не таку критичну як азот. Фосфор (P) є важливим макроелементом, необхідним для росту і розвитку рослин [21]. Він відіграє життєво важливу роль у багатьох фізіологічних процесах, включаючи передачу енергії, поділ і реплікацію клітин, синтез ДНК і РНК, фотосинтез і дихання. [22]. Cu (мідь) і Mn (марганець) мають від'ємну кореляцію з родючістю (-0.18 та 0.41 відповідно), що може вказувати на токсичність високих концентрацій цих елементів для певних культур.

Як мікроелемент, оптимальна кількість Cu необхідна для забезпечення клітинних функцій, але в надмірній кількості вона спричиняє шкідливий вплив на первинну продукцію та виживання рослин. Zn (цинк) і Fe (залізо) мають високу позитивну кореляцію (0.63), що свідчить про взаємозалежні процеси поглинання в ґрунті. OC (органічний вуглець) має високу кореляцію з S (сіркою) (0.63), що може бути пов'язано з

біогеохімічними циклами у ґрунті, де органічний матеріал сприяє збереженню поживних речовин.

З цього дослідження можна зробити висновки, що Азот як ключовий показник родючості має бути в центрі уваги при розробці стратегій удобрення та культивації полів, а також його потрібно враховувати при сівозміні, бо, як з'ясувалося, що є культури які допомагають стримувати Азот у ґрунті. Потрібно глибше дослідити взаємодію між елементами які вже згадувалися – цинк та залізо, щоб розуміти їх вплив на засвоєння поживних речовин.

Мідь має негативний вплив, якщо має високу концентрацію у ґрунті, але вона необхідна для засвоєння поживних речовин та забезпечення клітинних функцій, тому для більш точного аналізу потрібно вивести безпечні рівні цих елементів у ґрунті для певних культур.

3.2 Створення моделі машинного навчання для хімічного аналізу ґрунту

Було вирішено створити власну модель для дослідження результатів хімічного аналізу. Ця модель буде визначати родючість ґрунту на основі елементного складу. Датасет включає 880 зразків ґрунту в яких виміряно різноманітні хімічні показники (таблиця 3.1).

Набір даних за допомогою якого проводилось дослідження на моделі, що визначає родючість ґрунту за допомогою хімічного аналізу має такі атрибути: N – відношення вмісту азоту (NH_4^+) у ґрунті, P – відношення вмісту фосфору (P) у ґрунті, K – відношення вмісту калію (K) у ґрунті, pH – кислотність ґрунту (pH), ec – електропровідність, oc - органічний вуглець, S – сірка (S), zn – цинк (Zn), fe – залізо (Fe), cu – мідь (Cu), Mn – марганець (Mn), B – бор (B). Кількість даних – 880 записів.

Таблиця 3.1 – Частина датасету, який був використаний для експерименту [23]

N	P	Kp	H	EC	OC	S	Zn	Fe	Cu	MN	B	Output
138	8.6	560	7.46	0.62	0.70	5.9	0.24	0.31	0.77	8.71	0.11	0
213	7.5	338	7.62	0.75	1.06	25.4	0.30	0.86	1.54	2.89	2.29	0
163	9.6	718	7.59	0.51	1.11	14.3	0.30	0.86	1.57	2.70	2.03	0
157	6.8	475	7.64	0.58	0.94	26.0	0.34	0.54	1.53	2.65	1.82	0
270	9.9	444	7.63	0.40	0.86	11.8	0.25	0.76	1.69	2.43	2.26	1

Класифікація родючості ґрунту в датасеті відбувається за трьома категоріями:

- 401 зразок позначено як неродючий ґрунт (вихід = 0);
- 440 зразків позначено як родючий ґрунт (вихід = 1);
- 39 зразків мають позначку вихід = 2, що є показником високої родючості ґрунту.

N, P, K: гістограми містять розподіл вмісту азоту (N), фосфору (P), та калію (K) в ґрунті. Вони важливі макроелементи, необхідні для росту рослин. Ці розподіли показують, що більшість зразків ґрунту мають середній вміст цих елементів, з деякими відхиленнями (рисунок 3.3).

pH: ця гістограма показує розподіл кислотності ґрунту. Більшість зразків мають pH близько нейтрального (7), що є сприятливим для більшості культур.

ЕС (електропровідність): розподіл електропровідності вказує на рівень солей в ґрунті. Низька електропровідність може свідчити про низький вміст мінералів.

ОС (органічний вуглець): гістограма для органічного вуглеця показує розподіл його вмісту в зразках ґрунту.

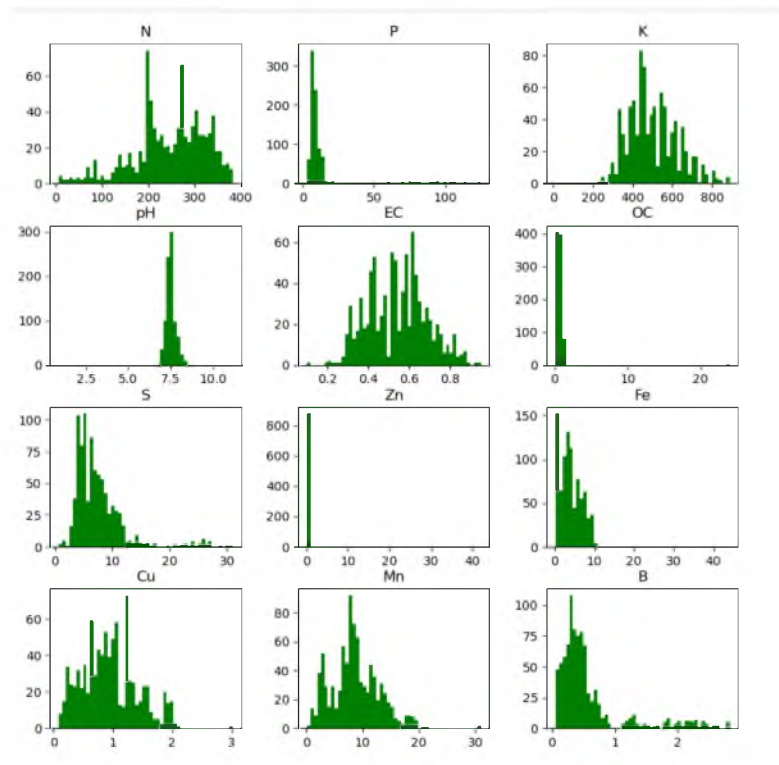


Рисунок 3.3 – Гістограми розподілу хімічних елементів та інших показників у ґрунті

S, Zn, Fe, Cu, Mn, B: Ці гістограми відображають розподіл мікроелементів сірки (S), цинку (Zn), заліза (Fe), міді (Cu), марганцю (Mn) та бору (B). Мікроелементи важливі для здоров'я рослин, хоч і потрібні в менших кількостях порівняно з макроелементами (рисунок 3.4).

За допомогою бібліотеки `sklearn` навчимо декілька моделей використовуючи такі класифікатори: `SVC`, `RandomForestClassifier`, `GaussianNB`, `KNeighbors` та `DecisionTreeClassifier` (рисунок 3.5).

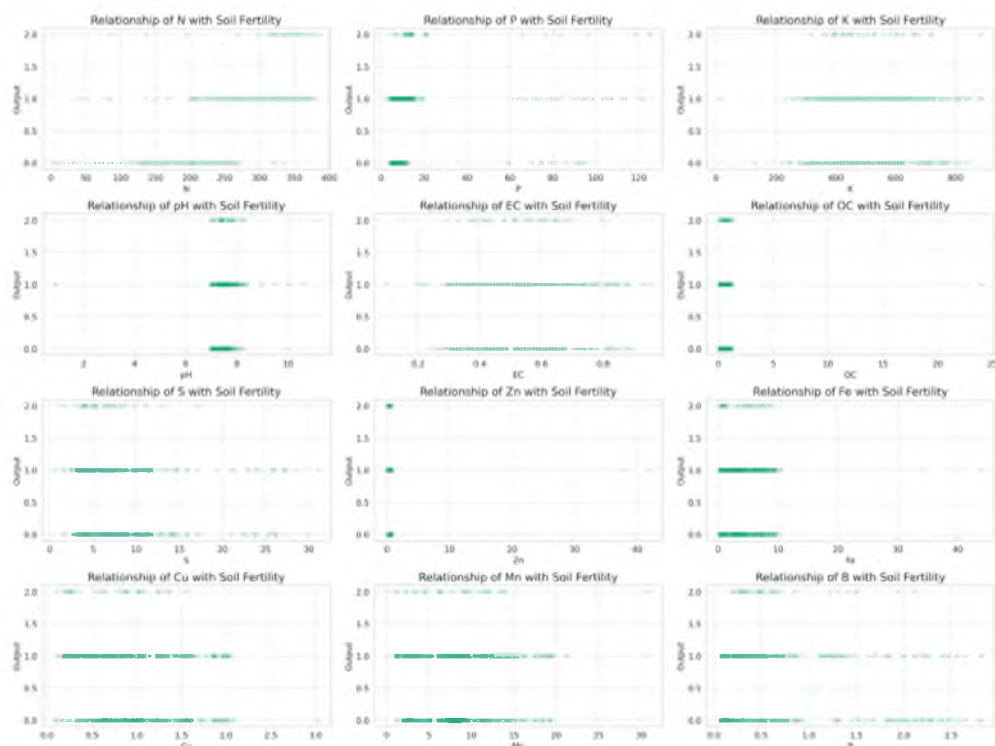


Рисунок 3.4 – Графік залежності між кількістю кожного хімічного елементу у ґрунті та його родючістю, яку позначено в датасеті. Кожна точка на графіку відповідає одному зразку з датасету. Вісь Y («Output») відображає рівень родючості ґрунту, де, «0» означає неродючий ґрунт, «1» – родючий, і можливо, «2» – висока родючість

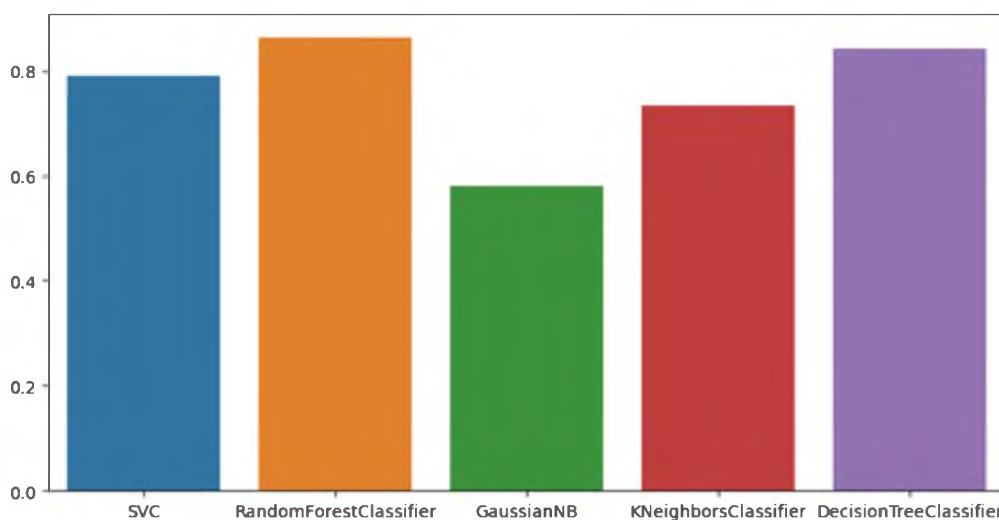


Рисунок 3.5 – Порівняльний графік точності кожного з класифікаторів

У рамках дослідження було розглянуто декілька популярних алгоритмів класифікації та оцінено їхні переваги та недоліки за кількома ключовими параметрами.

Support Vector Classifier (SVC) наведено на рисунку 3.6.

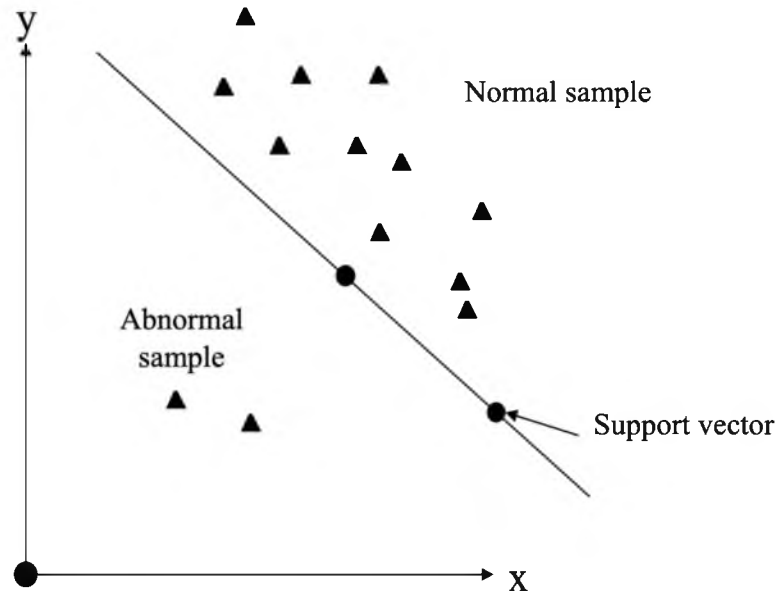


Рисунок 3.6 – Схематична діаграма SVC [24]

Переваги:

- визначає чіткий кордон розділення між класами, що ефективно для складних моделей;
- ефективний у високовимірних просторах, де кількість вимірів перевищує кількість зразків;
- економічно використовує пам'ять, що є важливим для обмежених обчислювальних ресурсів.

Недоліки:

- не підходить для аналізу великих наборів даних через високі обчислювальні витрати.

Random Forest Classifier наведено на рисунку 3.7.

Random Forest Classifier

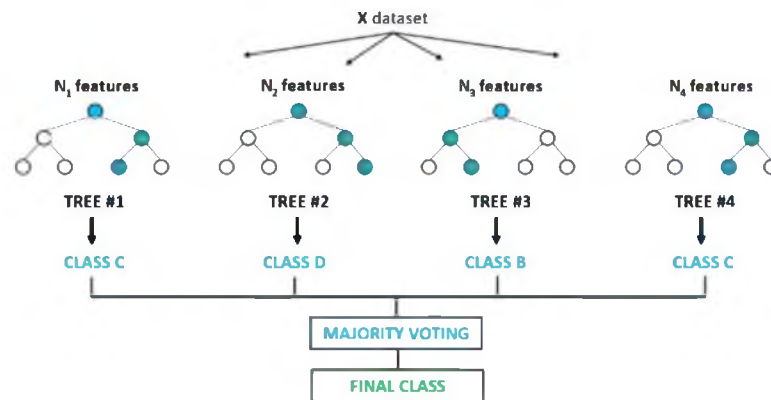


Рисунок 3.7 – Схематична діаграма Random Forest Classifier [25]

Переваги:

- швидкий в обчисленнях та здатний ефективно обробляти неврівноважені та відсутні дані;
- може бути ефективним для різноманітних даних без необхідності їх спеціальної підготовки.

Недоліки:

- можливе перенавчання на особливо шумних наборах даних;
- потребує значної кількості даних для навчання моделі.

Gaussian Naive Bayes (GaussianNB) наведено на рисунку 3.8.

Переваги:

- надзвичайно швидкий в роботі;
- показує високу ефективність при меншому об'ємі тренувальних даних, якщо дотримується припущення про незалежність ознак;
- особливо ефективний для категорійних вхідних змінних у порівнянні з числовими.

Недоліки:

- вимагає припущення про незалежність ознак, що може не завжди відповідати дійсності.

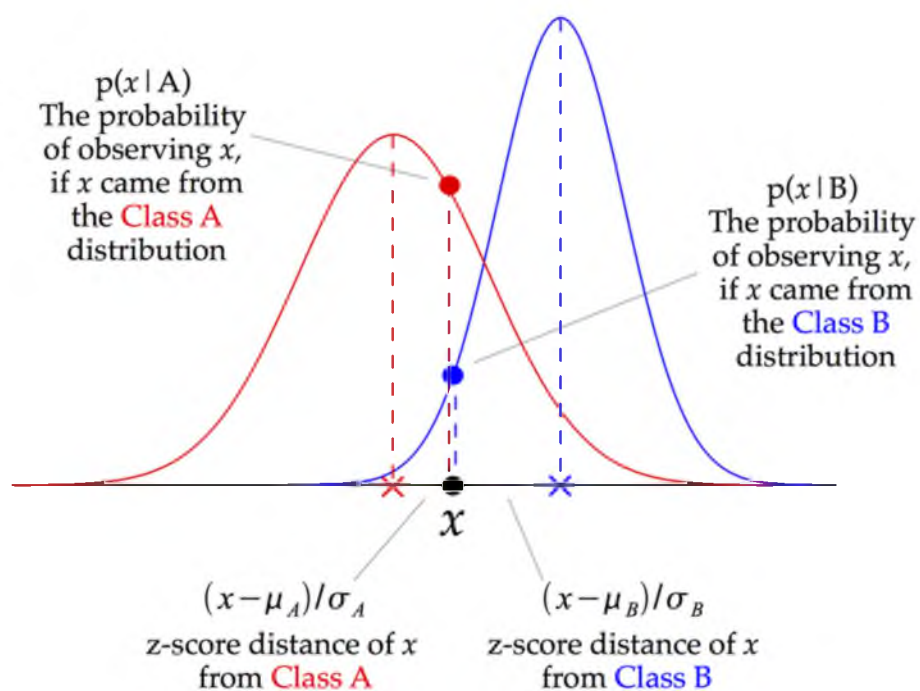


Рисунок 3.8 – Ілюстрація як працює Gaussian Naïve Bayes Classifier [26]

К-Nearest Neighbors (KNeighborsClassifier) наведено на рисунку 3.9.

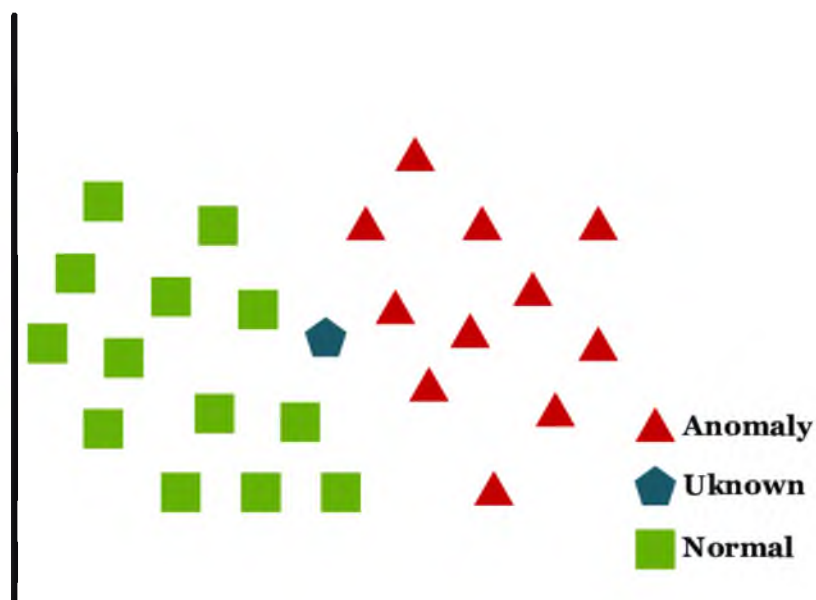


Рисунок 3.9 – Схематична ілюстрація K-Nearest Neighbors

Переваги:

- стійкий до шумних даних;
- ефективний, коли об'єм тренувальних даних є великим.

Недоліки:

- може бути вимогливим до обчислювальних ресурсів зі збільшенням об'єму даних.

Decision Tree Classifier наведено на рисунку 3.10.

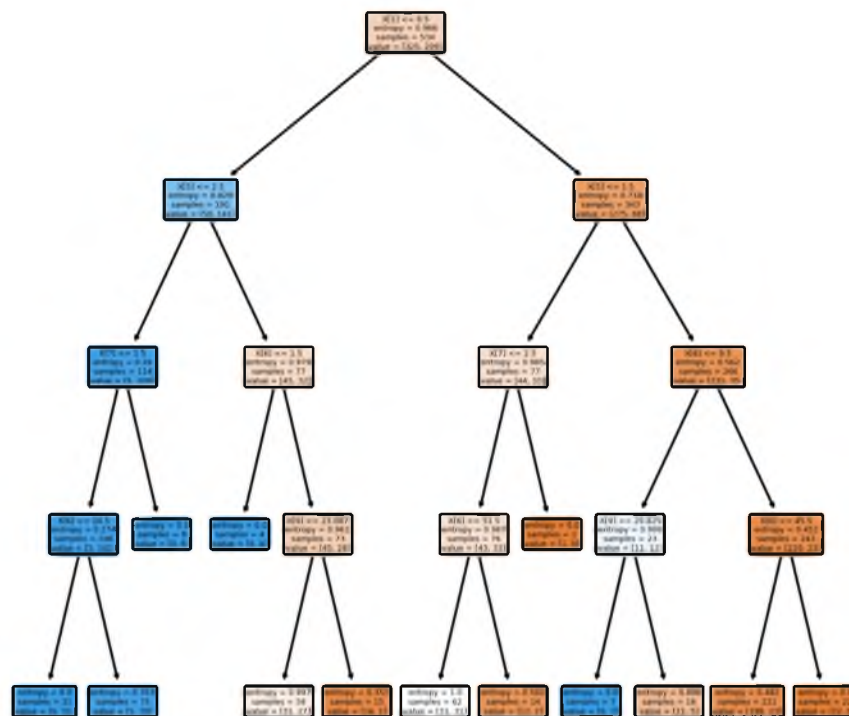


Рисунок 3.10 – Приклад Decision Tree

Переваги:

- не вимагає нормалізації або масштабування даних;
- інтуїтивно зрозумілий та легкий у інтерпретації.

Недоліки:

- високий час навчання моделі порівняно з іншими алгоритмами.

Виходячи з аналізу, найкращий вибір класифікатора для датасету зі зразками ґрунту залежить від декількох ключових факторів, таких як розмір датасету, кількість та тип ознак, а також наявність шуму та відсутніх даних.

Для датасету зі зразками ґрунту, який має розмірність 880 зразків та включає числові хімічні показники:

- SVC може бути ефективним, якщо у нас є високовимірний простір, але враховуючи, що SVC не підходить для великих датасетів через високі обчислювальні вимоги, цей метод може бути не найоптимальнішим;

- Random Forest Classifier є хорошим універсальним вибором, оскільки він може ефективно обробляти невеликі та середні датасети, а також він здатний впоратися з несбалансованими та відсутніми даними. Однак, можливість перенавчання слід враховувати;

- Gaussian Naive Bayes може бути швидким рішенням, але його ефективність залежить від того, чи дійсно ознаки незалежні одна від одної, що може бути не так у випадку хімічного складу ґрунту;

- K-Nearest Neighbors може не бути найкращим вибором, оскільки цей метод може бути вимогливим до обчислювальних ресурсів та часу на тренування зі збільшенням обсягу даних;

- Decision Tree Classifier не вимагає нормалізації даних, що є перевагою при роботі з хімічними показниками, але може вимагати більше часу для тренування моделі порівняно з іншими методами.

На підставі цього аналізу, Random Forest Classifier видається найкращим кандидатом для датасету зі зразками ґрунту, оскільки цей метод є стійким щодо різноманітності даних, здатний ефективно обробляти відсутні дані та добре масштабується до різних розмірів датасетів. Він також має вбудовані механізми для уникнення перенавчання та здатний впоратися з потенційним шумом у даних, що робить його відповідним для складних природних датасетів, як от зразки ґрунту.

Це лише один з великої кількості експериментів, які можливо провести для прогнозування родючості ґрунту. Також можливо провести

досліди зі знімками з дронів та класифікацією землі, яка придатна для посіву культур. Виходячи з того що оточує територію для посіву можна отримати дані про те які культури краще висадити, далі застосувати прогнозування для визначення хімічних властивостей ґрунту і на кінець створити модель, що буде прогнозувати родючість для певної культури, такі данні можна отримати дослідивши поля з вже засіяними культурами, та маючи результати їх родючості.

3.3 Опис датасету та моделі для проведення дослідження на основі даних, отриманих з дронів

Для отримання датасету з фотографій з дронів, потрібно розглянути оптимальні маршрути польоту визначити зони дослідження. Краще мати великий датасет з різними видами полів, тобто зробити фотографії які будуть показувати якомога більше можливих варіантів аби мати більш репрезентативні дані. Потрібно визначити досліджувані ділянки та розбити на ці ділянки на сектори аби полегшити управління даними. Для класифікації такого сектору можна використовувати такі ознаки як те що ділянка знаходиться близько до дороги, або має водойму, тому що ці ознаки можуть впливати на стан ґрунту та його склад (рисунки 3.11, 3.12).



Рисунок 3.11 – приклад фотографії поля з дрону DJI

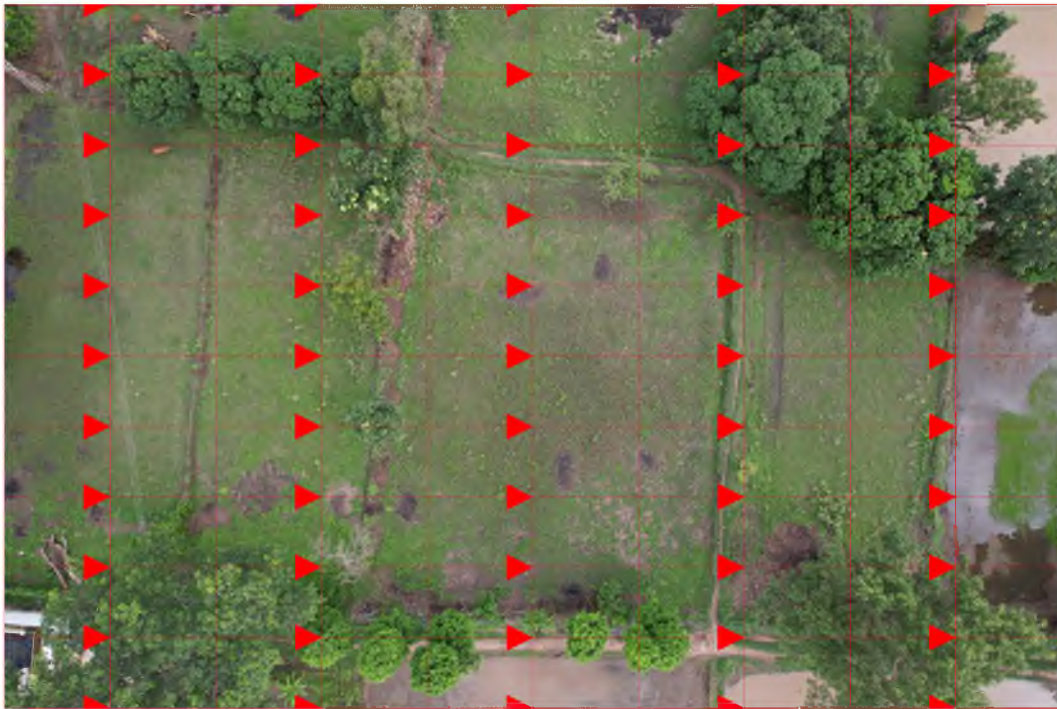


Рисунок 3.12 – Напрямок та сітка

Використовуючи дрон можна отримати фото, але потрібно додати сітку для того щоб потім зробити класифікацію поверхності (рисунок 3.13).

```

from PIL import Image, ImageDraw

image_path = 'DJI_0024-h60.jpeg'
image = Image.open(image_path)
draw = ImageDraw.Draw(image)

width, height = image.size
grid_spacing_x = width // 10
grid_spacing_y = height // 10

for y in range(0, height, grid_spacing_y):
    draw.line((0, y, width, y), fill='red', width=2)

for x in range(0, width, grid_spacing_x):
    draw.line((x, 0, x, height), fill='red', width=2)

arrow_length = grid_spacing_x // 4
for y in range(0, height, grid_spacing_y):
    for x in range(0, width, grid_spacing_x * 2):
        draw.line((x, y, x + grid_spacing_x, y), fill='red', width=2)
        draw.polygon([(x + grid_spacing_x, y), (x + grid_spacing_x - arrow_length, y - arrow_length /
/ 2), (x + grid_spacing_x - arrow_length, y + arrow_length // 2)], fill='red')

start_x = 0
start_y = 0
draw.ellipse((start_x - 5, start_y - 5, start_x + 5, start_y + 5), fill='blue')
output_path = 'image_with_flight_path.png'
image.save(output_path)
image.show()

```

Рисунок 3.13 – Додавання сітки на зображення

Використовуючи сітку певного розміру та масштабу можна більш точно проставити лейбли для того що є на фото. Це допомагає для визначення яка територія може бути засаджена та зрозуміти кількість вологи та можливість штучного поливу. Чим менше масштаб сітки, тим краще, тому що тоді дані про територію будуть точніші (рисунок 3.14).



Рисунок 3.14 – Зображення з доданими кольорами за типом поверхності

На зображенні додано лейбли за типом поверхності: зелений – насадження(vegetation), жовтий – поверхність придатна до посіву(ground), коричневий – дорога, або непридатна для насадження земля(road), синій – водойми (water). Подібні зображення можна побачити в інтерфейсі InCeres також є багато прикладів створення подібних зображень. Генерацію таких зображень можна оптимізувати створивши нейронну мережу, що буде за допомогою computer vision розпізнавати поверхність та накладати певний колір. Для автоматизації можна використовувати попередньо навчені моделі, які можуть розпізнавати об'єкти, наприклад U-Net, може бути

використана для детальної класифікації зображення. Використовуючи спеціальні інструменти такі як Labelbox або Supervisely, які дозволяють використати попередньо навчені моделі для автоматичної анотації зображень з можливістю ручного корегування.

За допомогою технік кластеризації та сегментації можна класифікувати пікселі на зображеннях, для цього використовуються алгоритми K-Means або DBSCAN. Ці алгоритми допомагають розділити зображення на різні сегменти на основі кольорових пікселів або текстурних ознак. Градієнтні методи та алгоритми порогового значення можуть бути використані для виявлення країв та сегментації зображень. Для того щоб побудувати систему автоматичної сегментації можна використати такий алгоритм:

- завантажити попередньо навчену модель використовуючи бібліотеку keras:

```
model = load_model('path_to_model.h5')
```

- завантажити зображення:

```
image = cv2.imread(image_path)
image_resized = cv2.resize(image, (256, 256))
image_resized = np.expand_dims(image_resized, axis=0)
/ 255.0
```

- додати прогнозування сегментації:

```
prediction = model.predict(image_resized)
predicted_mask = np.argmax(prediction[0], axis=-1)
```

- застосувати маски до оригінального зображення:

```
annotated_image = cv2.addWeighted(image, 0.5,
predicted_mask.astype(np.uint8) * 255, 0.5, 0)
```

- застосувати алгоритм до зображення:

```
image_path = 'path_to_image.jpg'
annotated_image = annotate_image(image_path, model)
```

Для того щоб отримати модель, яка зможе розставити лейбли, спочатку потрібно буде виділити класи які потрібно виділити – vegetation,

ground, road, water і далі навчити модель за допомогою пролейблованих зображень (рисунок 3.15).



Рисунок 3.15– Приклад пролейблованого зображення на якому присутня дорога та земля придатна до обробки

3.4 Згорткові мережі.

Згорткові нейронні мережі – це особливий вид багат шарових нейронних мереж. Як і майже всі інші нейронні мережі, вони навчаються за допомогою версії алгоритму зворотного поширення. Їх відмінність полягає в архітектурі [27] (рисунок 3.16).

Вони стали популярними завдяки властивості автоматично виліляти та навчатися характерним ознакам зображень, що робить їх особливо ефективними для задач комп'ютерного зору, таких як класифікація зображень, сегментація, виділення об'єктів та багато інших.

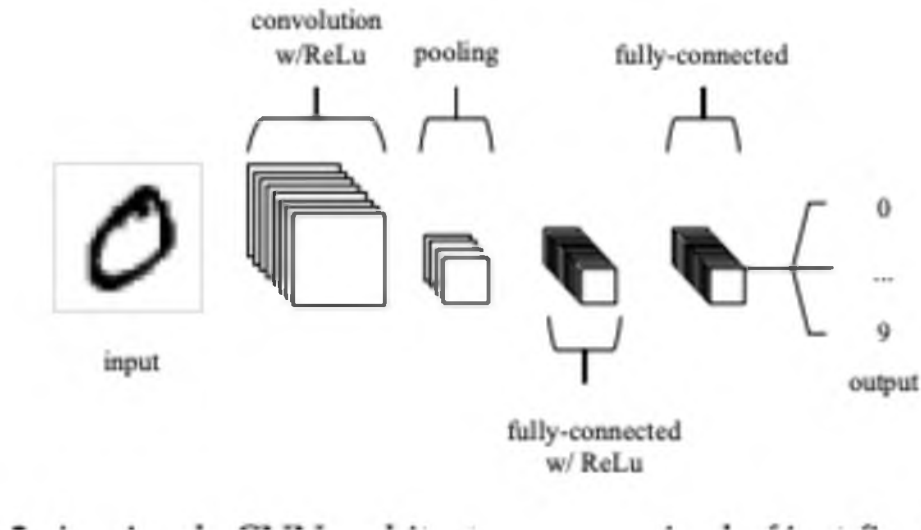


Рисунок 3.16 – Проста архітектура CNN, що складається з п'яти шарів [28]

Основна архітектура CNN складається з трьох шарів: згорткові шари (convolutional layers), шарів підвибірки (pooling layers) та шарів повнозв'язних (fully connected layers). Згорткові шари представляють собою шари які виконують операцію згортки, яка полягає у застосуванні фільтрів до вхідних даних. Кожен фільтр застосовується до всього зображення, виділяючи специфічні ознаки, такі як краї, текстури або кольори.

Результатом роботи є карта ознак (feature map), що містить інформацію про наявність конкретної ознаки в різних областях на зображенні, тобто за допомогою цього шару виділяється наприклад область водойми на фото, за допомогою інших фільтрів можна виділити різні області у яких є притаманний їм шаблон, або колір, або текстура. Шари підвибірки (pooling layers) застосовуються для зменшення карт ознак, що допомагає зменшити кількість параметрів та обчислювальну складність моделі. Цей шар потрібен для оптимізації роботи моделі. Повнозв'язні шари (fully-connected) це шари де кожний нейрон пов'язаний з усіма нейронами попереднього шару. Ці шари виконують функцію класифікації або регресії. Основним плюсом у використанні для виділення областей на

фотографіях з дронів є те що така мережа може автоматично вчитися виділяти важливі ознаки, тобто можна отримати декілька областей до яких потім можна додати лейбли, але ці області будуть виділені автоматично, що зменшить час на створення датасету. Зменшення параметрів добре впливає на роботу мережі і зменшує кількість необхідних ресурсів, а також дозволяє будувати великі мережі без ризику перенавчання.

Для створення таких карт за допомогою нейронних мереж можна використати згорткові нейронні мережі. Цей підхід дозволить обробляти зображення, визначати поверхності і класифікувати їх за тим чи придатні вони для насаджень, чи це вже існуючі насадження, дороги та водойми.

Спочатку потрібно зібрати дані, що відбувається за допомогою вже згаданих дронів, які можуть робити високоякісні зображення. Після цього потрібно додати лейбли для кожної піксельної області, для чого можна використати готові моделі, або створити самому. Аби збільшити кількість даних для навчання, можна застосувати техніки аугментації даних – розвертання на 90 градусів, масштабування, віддзеркалення. Далі треба розділити дані на датасети для тренування, валідації та тестування, щоб можна було оцінити результати моделі.

Для виконання цієї задачі потрібно обрати попередньо треновану згорткову нейронну мережу (наприклад U-Net, SegNet або DeepLab), що можуть працювати з задачами сегментації зображень. Код нейронної мережі, яка має архітектуру U-Net наведено у лістингу 3.1.

Лістинг 3.1 – Код нейронної мережі, яка має архітектуру U-Net

```
def unet(input_size=(256, 256, 3)):
    inputs = tf.keras.Input(input_size)

    conv1 = Conv2D(64, 3, activation='relu',
padding='same')(inputs)
    conv1 = Conv2D(64, 3, activation='relu',
padding='same')(conv1)
```

Продовження лістингу 3.1

```
pool1 = MaxPooling2D(pool_size=(2, 2))(conv1)

conv2 = Conv2D(128, 3, activation='relu',
padding='same')(pool1)
conv2 = Conv2D(128, 3, activation='relu',
padding='same')(conv2)
pool2 = MaxPooling2D(pool_size=(2, 2))(conv2)

conv3 = Conv2D(256, 3, activation='relu',
padding='same')(pool2)
conv3 = Conv2D(256, 3, activation='relu',
padding='same')(conv3)
pool3 = MaxPooling2D(pool_size=(2, 2))(conv3)

conv4 = Conv2D(512, 3, activation='relu',
padding='same')(pool3)
conv4 = Conv2D(512, 3, activation='relu',
padding='same')(conv4)
pool4 = MaxPooling2D(pool_size=(2, 2))(conv4)

conv5 = Conv2D(1024, 3, activation='relu',
padding='same')(pool4)
conv5 = Conv2D(1024, 3, activation='relu',
padding='same')(conv5)

up6 = UpSampling2D(size=(2, 2))(conv5)
up6 = concatenate([up6, conv4], axis=3)
conv6 = Conv2D(512, 3, activation='relu',
padding='same')(up6)
conv6 = Conv2D(512, 3, activation='relu',
padding='same')(conv6)

up7 = UpSampling2D(size=(2, 2))(conv6)
up7 = concatenate([up7, conv3], axis=3)
```

Продовження лістингу 3.1

```

conv7 = Conv2D(256, 3, activation='relu',
padding='same')(up7)
conv7 = Conv2D(256, 3, activation='relu',
padding='same')(conv7)

up8 = UpSampling2D(size=(2, 2))(conv7)
up8 = concatenate([up8, conv2], axis=3)
conv8 = Conv2D(128, 3, activation='relu',
padding='same')(up8)
conv8 = Conv2D(128, 3, activation='relu',
padding='same')(conv8)

up9 = UpSampling2D(size=(2, 2))(conv8)
up9 = concatenate([up9, conv1], axis=3)
conv9 = Conv2D(64, 3, activation='relu',
padding='same')(up9)
conv9 = Conv2D(64, 3, activation='relu',
padding='same')(conv9)

conv10 = Conv2D(4, 1, activation='softmax')(conv9)

model = Model(inputs=[inputs], outputs=[conv10])

model.compile(optimizer=Adam(lr=1e-4),
loss='categorical_crossentropy', metrics=['accuracy'])

```

Далі створемо та навчимо модель, код наведено у лістингу 3.2.

Лістинг 3.2 – Навчання моделі

```

history = model.fit(train_images, train_masks, epochs=50,
batch_size=32, validation_data=(val_images, val_masks))

```

Для отримання класифікованого зображення використаємо модель:

```

predictions = model.predict(new_images)

```

Продовження лістингу 3.2

```
for i, prediction in enumerate(predictions):  
    plt.figure(figsize=(10, 5))  
    plt.subplot(1, 2, 1)  
    plt.imshow(new_images[i])  
    plt.title('Original Image')  
    plt.subplot(1, 2, 2)  
    plt.imshow(np.argmax(prediction, axis=-1))  
    plt.title('Predicted Segmentation')  
    plt.show()
```

Використовуючи різні підходи можна отримати більш точні та наближені до реальних результати, що дасть змогу оцінити, чи потрібно взагалі використовувати нейронні мережі для таких задач. Створивши декілька моделей які працюють за різними підходами можна отримати повноцінну систему з догляду та прогнозування родючості ґрунтів.

4 СЦЕНАРІЇ ВИКОРИСТАННЯ МОДЕЛЕЙ

4.1 Практичне використання моделей у створенні рішень

Якщо системи будуть інтегровані на місцях, то вони покращать виявлення родючості ґрунту, прискорять цей процес та зроблять його дешевшим для самого фермера, як це зробили InCeres, їх данні доводять те що процес аналізу вдалося зробити дешевшим, що важливо для країн з економікою що розвивається. Системи що використовують дрони для отримання даних про поле, та що його оточує дозволять також оптимізувати те де засіяти певну культуру, тому що є стійкі до захворювань, а є ні, і ті що більш стійкіші можуть рости біля дороги, якщо така проходить біля поля.

Обмеження які можуть завадити впровадженню подібних технологій стосуються швидше регуляцій, які можуть бути застосовані до ШІ, а також ціна такого обладнання, тому що для невеликих господарств не потрібен постійний контроль над їх землею, достатньо проводити аудит для якого також можуть використовуватись подібні технології, але компаніями які надають послуги за необхідністю. Ферми що продають свою продукцію та повинні мати сертифікацію на продукцію навпаки можуть виграти від того що автоматизують процес верифікації якості.

Існує підхід відновлювального сільського господарства (Regenerative Agriculture), яке вже згадувалось. Його задачею є підвищення родючості верхнього шару ґрунту, системний підхід до ведення діяльності, швидка адаптація до змін навколишнього середовища. Головні принципи стосуються сівозміни, мінімізації технічного втручання у обробку землі та використання натуральних добрив. Основною проблемою і слабкістю цього підходу є те що він дає результат через роки. Якщо застосувати моделі для хімічного аналізу та поверхневого аналізу за допомогою дронів, можливо, вдасться мінімізувати час на адаптацію підходу. Також плюсом буде те, що зменшити механічне втручання буде можливим завдяки аналізу ґрунту,

наприклад можна буде зменшити кількість рослин шкідників, що дозволить уникнути необхідності полоти культуру, що і є механічним втручанням.

Використання ШІ зможе підтвердити ефективність цього підходу та вдосконалити його. Своєчасне реагування на зміни клімату та інші показники теж буде чинником для зменшення втручання у ріст культур. Використання мап стану культур дозволить локалізувати проблеми, там робити механічні втручання лише на певних територіях.

Однією з проблем використання таких моделей для оцінки полів може стати нерозуміння технології з боку фермерів, що завадить зібрати датасет. Без датасету, який має велику вибірку полів неможливо буде відтворити моделі з вище проведених експериментів з нейронними мережами (рисунок 4.1).

Ці моделі можна буде використовувати у екосистемі пристроїв, що збирають дані автономно, та не відправляють їх нікуди і використовують Lifelong Learnign підхід, це коли модель навчається на даних постійно без необхідності бути тренуваною заново.

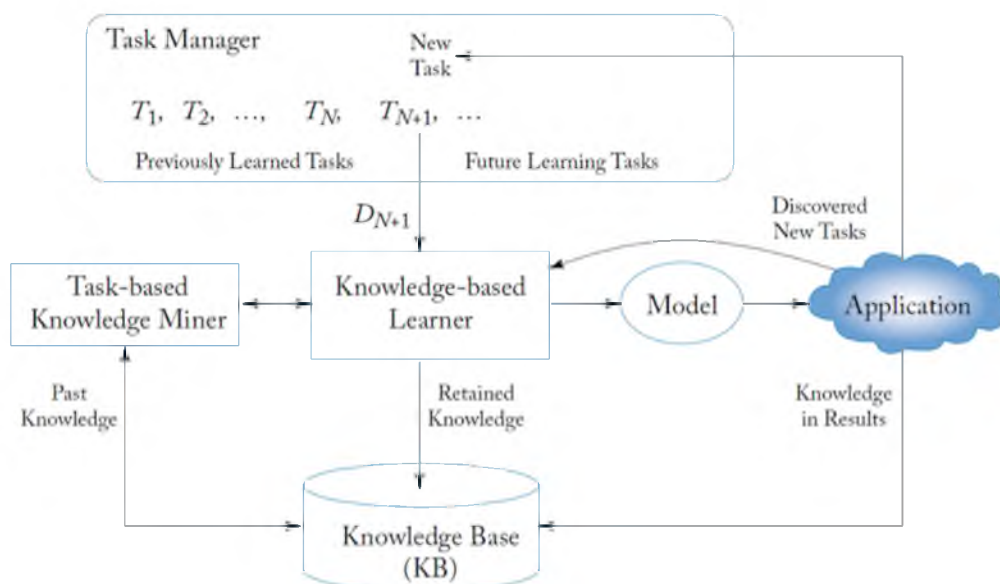


Рисунок 4.1 – Архітектура системи машинного навчання протягом усього життя [29]

Використовуючи підхід local-first (дані не відправляються на сервер, а залишаються на клієнті, також всі розрахунки відбуваються за рахунок використання ресурсів клієнта), можна отримати систему, що буде постійно обробляти локальні дані та навчати модель на даних які являються актуальними для певної території. Такі дані звичайно можуть бути корисні для створення початкової моделі, що буде обробляти датасет з різними видами поверхності, тому що це зробить модель більш гнучкою. Під екосистемою, мається на увазі система з різних типів датчиків, що можуть використовуватись для збору інформації та подальшої її обробки.

ВИСНОВКИ

У кваліфікаційній роботі було розглянуто та досліджено застосування моделей штучного інтелекту для оцінки родючості ґрунтів з метою підвищення ефективності вирощування сільськогосподарських культур.

Під час дослідження були виконані наступні задачі:

– розглянуто актуальність застосування ШІ в агросекторі: використання штучного інтелекту та машинного навчання в сільськогосподарській галузі є надзвичайно актуальним, оскільки ці технології дозволяють автоматизувати процеси аналізу ґрунтів, підвищувати врожайність та оптимізувати використання ресурсів, що важливо для економік, що розвиваються, та в умовах, які не є сприятливими для ведення сільськогосподарської діяльності;

– проаналізовано методи збору та аналізу даних: в роботі були розглянуті різні методи збору та аналізу даних для оцінки хімічних властивостей ґрунту. Основні елементи, що впливають на родючість ґрунту, були визначені за допомогою методів хімічного аналізу, таких як атомно-абсорбційна спектрофотометрія, іонна хроматографія та інші;

– проведено кореляційний аналіз: проведений кореляційний аналіз показав, що азот (N) є ключовим показником родючості ґрунту. Висока кореляція азоту з родючістю підкреслює важливість цього елемента для підтримання продуктивності ґрунтів. Інші елементи, такі як фосфор (P), калій (K) та органічний вуглець (OC), також мають значний вплив на родючість. Деякі елементи можуть впливати негативно у великій кількості (Fe, Cu), але у малих кількостях необхідні для засвоєння інших елементів;

– створено модель машинного навчання: було розроблено та навчено модель машинного навчання для прогнозування родючості ґрунту на основі хімічних показників. Модель дозволяє передбачати родючість, використовуючи дані про вміст ключових хімічних елементів у ґрунті;

– досліджено можливість використання дронів для збору даних: дрони використовуються для збору високоякісних зображень полів, що дозволяє аналізувати стан ґрунту та рослин на великих територіях. Це підвищує ефективність моніторингу стану ґрунтів та дає можливість створювати карти щоб потім на них навчати модель для оцінки ґрунту та наскільки він підходить для тої чи іншої культури;

– досліджено можливість автоматизації анотації даних та створення мап стану ґрунтів: застосування комп'ютерного зору та глибинного навчання для автоматизації анотації даних дозволяє значно прискорити процес підготовки даних для навчання моделей ШІ. Використання попередньо навчених моделей та методів кластеризації допомагає автоматизувати цей процес.

Отримані результати дають змогу оцінити чи є необхідність у використанні штучного інтелекту у агрономії. Використовуючи нові технології можна оптимізувати кількість ресурсів, допомогти країнам з поганими умовами для вирощування культур зрозуміти як можна змінити підхід, або на які культури потрібно переключитись, щоб мати кращі результати під час збору продукції.

Перспективи подальших досліджень у цій області можуть включати покращення моделей хімічного аналізу та збільшення параметрів для моделей. Можливість включення властивостей повітря, щоб отримувати дані про забруднення та використовувати їх для прогнозування хвороб рослин. Потрібно зосередитись на створенні методу аналізу та анотації знімків з дронів. Багато платформ працюють зі знімками з супутників, тож було б непогано розглянути і такі моделі, щоб зрозуміти їх ефективність. Створення прототипу екосистеми пристроїв для полів, аби мати повноцінну картину того що відбувається з полями, тобто створення системи оповіщення про зміни стану ґрунту та розробка механізму підказок, які б допомагали приймати рішення у разі необхідності аби максимізувати врожай при наявних ресурсах.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Про родючість ґрунту: види та способи покращити | блог – тетра агро. Купити мінеральні NPK добрива в Україні. Ціни Виробника | Тетра Агро. URL: https://tetra-agro.com.ua/news/rodyucist_gruntiv_yak_viznacity_ta_pokrashhiti (дата звернення: 26.05.2024).
2. AI in agriculture: 8 practical applications. V7 – Build Trustworthy AI at Scale | Automate Labeling. URL: <https://www.v7labs.com/blog/ai-in-agriculture#crop-and-soil-monitoring> (date of access: 26.05.2024).
3. Особливості збирання і зберігання зібраного врожаю з підвищеною вологістю зерна. Головна. URL: <https://www.cherk-consumer.gov.ua/novyny/339-osoblyvosti-zbyrannia-i-zberihannia-zibranoho-vrozhaiu-z-pidvyshchenoiu-volohistiu-zerna> (дата звернення: 26.05.2024).
4. Introduction to the AI. URL: https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf. – с. 228.
5. Artificial Intelligence for Agriculture Innovation. URL: https://www3.weforum.org/docs/WEF_Artificial_Intelligence_for_Agriculture_Innovation_2021.pdf.
6. Rourke RV. Chemical and physical properties of the Mapleton, Monson, Saddleback, and Sisk soil map units. Maine Agricultural and Forest Experiment Station Technical Bulletin 169. 2001. – с. 11.
7. Mckinion J M Lemmon HE. Expert systems for agriculture // *Computers & Electronics in Agriculture*. 1985. Vol. 1, No. 1. P. 31-40.
8. Gertsis AC Galanopoulou-Sendouca S Papathanasiou G Symeonakis A. USE OF GOSSYM-A COTTON GROWTH SIMULATION MODEL-TO MANAGE A LOW INPUT COTTON PRODUCTION SYSTEM IN GREECE. In First European Conference for Information Technology in Agriculture, 1997.

9. Gertsis AC Whisler FD. GOSSYM: A COTTON CROP SIMULATION MODEL AS A TOOL FOR THE FARMER // *Acta Horticulturae*. 1998. Vol. 476. P. 213-218. DOI: 10.17660/ActaHortic.1998.476.24.

10. Учасники проєктів Вікімедіа. Четверта промислова революція – Вікіпедія. Вікіпедія. URL: [https://uk.wikipedia.org/wiki/Четверта промислова революція](https://uk.wikipedia.org/wiki/Четверта_промислова_революція) (дата звернення: 26.05.2024).

11. Selamat A Rose Alinda Alias Syed Norris Hikmi Marlia Puteh Siti Hamisah Tapsir. Higher Education 4.0 : Current Status and Readiness in Meeting the Fourth Industrial.

12. Public Documents | The World Bank. URL: <https://thedocs.worldbank.org/en/doc/647641558460357412-0090022019/render/PASETPRESENTATIONPSVTTKENYAFINAL.pdf> (дата звернення: 26.05.2024).

13. Artificial Intelligence in Soil Quality Monitoring - Bannari Amman Institute of Technology. Bannari Amman Institute of Technology. URL: <https://www.bitsathy.ac.in/artificial-intelligence-in-soil-quality-monitoring/#:~:text=According%20to%20Leonardo,the%20program%20develops> (date of access: 26.05.2024).

14. Simple Neural Network. URL: https://assets-global.website-files.com/5a625b9d69b7f60001db7c53/5bd0fe26404e38d8d1463f1a_Neural-Network-v1-compressed.jpg.

15. George F Smoot George Debono Ivan. Artificial intelligence for EKG medical evaluation // *Qingdao Daxue Xuebao (Gongcheng Jishuban)/Journal of Qingdao University (Engineering and Technology Edition)*. 2018. URL: https://www.researchgate.net/publication/339427301_Artificial_intelligence_for_EKG_medical_evaluation.

16. Overview of GAN Structure | Machine Learning | Google for Developers. Google for Developers. URL: https://developers.google.com/machine-learning/gan/gan_structure (date of access: 26.05.2024).
17. Teralytic | A Soil Health Company. Teralytic | A Soil Health Company. URL: <https://teralytic.com/> (date of access: 26.05.2024).
18. Pearson's Correlation Coefficient. In: Kirch W. (eds) Encyclopedia of Public Health. Springer Dordrecht. 2008. DOI: 10.1007/978-1-4020-5614-7_2569.
19. Spearman Rank Correlation Coefficient. In: The Concise Encyclopedia of Statistics. Springer New York, NY. 2008. DOI: 10.1007/978-0-387-32833-1_379.
20. Teague WR Apfelbaum S Lal R Kreuter UP Rowntree J Davies CA Conser R Rasmussen M Hatfield J Wang T Wang F. The role of ruminants in reducing agriculture's carbon footprint in North America // *Journal of Soil and Water Conservation*. 2016. Vol. 71, No. 2. P. 156-164. DOI: 10.2489/jswc.71.2.156. ISSN 0022-4561.
21. Marschner P. Marschner's mineral nutrition of higher plants. Academic Press. 2012.
22. Raghothama KG. Phosphate acquisition // *Annual Review of Plant Physiology and Plant Molecular Biology*. 1999. Vol. 50, No. 1. P. 665-693.
23. Dataset information. URL: <https://drive.google.com/file/d/1s69LXadxz9bams9NS9FBP3tb5okr1u1H/view?usp=sharing>.
24. Huang G Chen J Liu L. One-Class SVM Model-Based Tunnel Personnel Safety Detection Technology // *Applied Sciences*. 2023. Vol. 13. P. 1734. DOI: 10.3390/app13031734.
25. Random Forest Classifier. URL: <https://www.freecodecamp.org/news/content/images/2020/08/how-random-forest-classifier-work.PNG> (дата звернення: 26.05.2024).

26. Raizada R Lee Y. Smoothness without Smoothing: Why Gaussian Naive Bayes Is Not Naive for Multi-Subject Searchlight Studies // *PloS One*. 2013. Vol. 8. P. e69566. DOI: 10.1371/journal.pone.0069566.

27. MNIST Demos on Yann LeCun's website. Wayback Machine. URL: <https://web.archive.org/web/20210224225707/http://yann.lecun.com/exdb/lenet/> (date of access: 26.05.2024).

28. An Introduction to Convolutional Neural Networks. URL: <https://arxiv.org/pdf/1511.08458> (дата звернення: 26.05.2024).

29. Chen Z Liu B. Lifelong Machine Learning. Morgan & Claypool Publishers. 2018.