




Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему Метод нейромережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів

Галузь знань 12 – Інформаційні технології
Шифр і назва галузі знань
Спеціальність 122 – Комп'ютерні науки
Шифр і назва спеціальності
Освітня програма Комп'ютерні науки
Назва освітньої програми

Виконав: студент групи КН-22-3  Костянтин СИДОРЧУК
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ
Керівник: д.т.н., професор каф. КН  Едуард МАНЗЮК
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ
Нормоконтроль: к.т.н., доц. каф. КН  Руслан БАГРІЙ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

До захисту допускаю:

Зав. кафедри КН, д.т.н., професор

18 06 2026 р.

 Олександр БАРМАК
Підпис Ім'я, ПРІЗВИЩЕ

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій

кафедра комп'ютерних наук

Освітній ступінь бакалавр

Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук



(підпис)

д.т.н., професор Олександр БАРМАК

« 22 » січня 2026 року

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА**

1. Тема кваліфікаційної роботи бакалавра: «Метод нейромережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів»

2. Завдання видано студенту Костянтину Сидорчуку
(Ім'я, прізвище)

3. Керівник роботи д.т.н., професор Едуард Манзюк
(посада, ім'я, прізвище)

4. Затверджено наказом університету від «20» січня 2026 р. № 4

5. Дата видачі завдання студенту: «22» січня 2026 р.

6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Метою кваліфікаційної роботи бакалавра є підвищення точності та оперативності оцінювання стану здоров'я і прогнозування стресових станів рослин шляхом розробки методу нейромережевої класифікації на основі MLP за числовими показниками біосенсорів, що дозволить замінити суб'єктивні ручні методи агрономічного контролю на об'єктивну автоматизовану систему, підвищити ефективність моніторингу агрокультур, зменшити втрати від несвоєчасного виявлення стресових станів та забезпечити стабільну оцінку фізіологічного стану рослин за даними біосенсорів.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	Вибір напрямку дослідження та узгодження теми кваліфікаційної роботи з керівником, складання календарного графіка виконання	січень 2026	Виконано
2	Ознайомлення з предметною областю, формулювання мети і задач дослідження, визначення об'єкта та предмета дослідження	лютий 2026	Виконано
3	Проектування методу розв'язання задачі, опис архітектурних рішень, розроблення математичних моделей та алгоритмів.	березень 2026	Виконано
4	Обґрунтування інструментарію розробки, програмна реалізація розробленого методу, проведення експериментального тестування та оцінювання ефективності.	квітень 2026	Виконано
5	Написання тексту кваліфікаційної роботи, урахування зауважень керівника, оформлення згідно з вимогами	травень 2026	Виконано
6	Розроблення презентаційних матеріалів та попередній захист кваліфікаційної роботи	травень 2026	Виконано
7	Отримання відгуку керівника, рецензії, перевірка тексту кваліфікаційної роботи на плагіат, нормоконтроль	червень 2026	Виконано
8	Підготовка до захисту та захист кваліфікаційної роботи	червень 2026	Виконано

Виконавець:

студент групи КН-22-3

Група виконавця

Костянтин Сидорчук
Підпис

Костянтин Сидорчук

Ім'я, ПРІЗВИЩЕ

Керівник:

д.т.н., професоркаф. КН

Науковий ступінь, посада

Едуард МАНЗЮК
Підпис

Едуард МАНЗЮК

Ім'я, ПРІЗВИЩЕ

Анотація

Тема кваліфікаційної роботи бакалавра: «Метод нейромережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів»

Виконавець кваліфікаційної роботи бакалавра: студент групи КН-22-3 Костянтин Сидорчук

Керівник кваліфікаційної роботи бакалавра: д.т.н., професор каф. КН Едуард Манзюк

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
53	26	5	42	2

Мета кваліфікаційної роботи бакалавра – підвищення точності та оперативності оцінювання стану здоров'я і прогнозування стресових станів рослин шляхом розробки методу нейромережевої класифікації на основі MLP за числовими показниками біосенсорів.

Для досягнення цієї мети та експериментального підтвердження дієвості запропонованого методу реалізовано програмно-моделюючий комплекс штучного інтелекту.

Практичне значення роботи полягає у створенні доступного комп'ютеризованого інструментарію для агрономів, який дозволяє оперативно оцінювати фізіологічний стан культур за вісьмома біосенсорними параметрами без залучення дорогого лабораторного устаткування. Запропонований підхід забезпечує автоматичне розпізнавання чотирьох рівнів життєдіяльності рослини від нормального стану до критичного стресу.

Ключові слова: нейромережева класифікація, біосенсорні дані, MLP, оцінка стану рослин, SMOTE, прецизійне землеробство.

Виконавець: студент групи КН-22-3

Група виконавця


Підпис

Костянтин СИДОРЧУК

Ім'я, ПРІЗВИЩЕ

Зміст

Перелік скорочень	3
Вступ.....	4
Розділ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій.....	5
1.1 Аналіз інформаційних моделей.....	5
1.2 Огляд теоретичних підходів до розв'язку подібних задач.....	7
1.3 Аналіз існуючих програмних засобів та наукових рішень.....	9
1.4 Мета, задачі та вимоги до реалізації методу	13
Розділ 2 Розробка нейромережевого методу оцінювання стану здоров'я рослин за даними біосенсорів.....	15
2.1 Формалізація задачі та інформаційна модель	15
2.2 Розробка нейромережевого методу.....	19
2.3 Алгоритм нейромережевої класифікації стану рослин за даними біосенсорів	25
2.4 Опис набору даних.....	28
2.5 Метрики оцінювання ефективності	31
2.6 Сценарій проведення експерименту	34
2.7 Висновки до розділу 2	35
Розділ 3 Експериментальні дослідження та аналіз результатів	37
3.1 Опис прикладного застосування	37
3.2 Дослідження стабільності навчання.....	40
3.3 Оцінка точності класифікації.....	42
3.4 Порівняльний аналіз з іншими методами.....	49
3.5 Висновки до розділу 3	52
Загальні висновки.....	53
Перелік посилань.....	54
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
Adam	Adaptive Moment Estimation
AUC	Area Under the Curve
CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
CTI	Crop Transferability Index
DSS	Decision Support System
ESG	Environmental, Social, Governance
FID	Fréchet Inception Distance
FPS	Frames Per Second
GAN	Generative Adversarial Network
GIS	Geographic Information System
Grad-CAM	Gradient-weighted Class Activation Mapping
GPU	Graphics Processing Unit
HLB	Huanglongbing (Citrus Greening)
kNN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MLP	Multilayer Perceptron
PLS-DA	Partial Least Squares-Discriminant Analysis
SSTAS	Spectral Signature Taxonomy & Analysis Software
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
ROC-AUC	Receiver Operating Characteristic – Area Under Curve
XAI	Explainable Artificial Intelligence

Вступ

Актуальність теми. У прецизійному землеробстві критично важливим є перехід до методів прямого моніторингу фізіологічного стану культур. Традиційні підходи не дозволяють виявляти порушення на ранніх стадіях. Рослини генерують сигнали, що реєструються біосенсорами та містять інформацію про водний чи термічний стрес. Проте ці дані є нелінійними та незбалансованими, що ускладнює їх аналіз стандартними алгоритмами. Застосування архітектури MLP із методами регуляризації та SMOTE-балансування дозволяє ефективно класифікувати стан рослин за 8 числовими ознаками (флуоресценція, вміст хлорофілу тощо).

Об'єкт дослідження – процеси автоматизованої діагностики стану рослин за даними біосенсорів.

Предмет дослідження – методи та алгоритми нейромережевої класифікації стану рослин за багатовимірними числовими показниками із застосуванням MLP та методів балансування класів.

Мета кваліфікаційної роботи бакалавра – підвищення точності та оперативності оцінювання стану здоров'я і прогнозування стресових станів рослин шляхом розробки методу нейромережевої класифікації на основі MLP за числовими показниками біосенсорів.

Завдання роботи:

- провести аналіз предметної області методів оцінювання стану рослин за даними біосенсорів;
- розробити метод нейромережевої класифікації стану рослин на основі архітектури MLP;
- реалізувати програмну систему мовою Python із графічним інтерфейсом для аналізу біосенсорних даних;
- провести експериментальне дослідження ефективності розробленого методу та порівняльний аналіз з іншими підходами.

Розділ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій

1.1 Аналіз інформаційних моделей

У сучасному агровиробництві концепція Smart Farming стала одним із головних способів підвищити ефективність і зменшити шкоду для довкілля [1]. Прецизійне землеробство побудоване на постійному моніторингу стану рослин – це основа для щоденних рішень: скільки поливати, чим підживлювати, чи не з'явилися шкідники [2]. Класичні методи контролю – переважно візуальний огляд і метеоспостереження – реагують на проблему вже після того, як вона виникла, і не дозволяють вчасно помітити відхилення на початку вегетації [3]. Носимі гнучкі сенсори дають принципово інші можливості: вони безперервно знімають фізіологічні сигнали безпосередньо з рослинних тканин [4].

З розвитком біосенсорики з'явилась можливість вимірювати стан рослини без її пошкодження і в режимі реального часу. Серед найважливіших параметрів – флуоресценція хлорофілу, пігментний склад листя, температура листової пластини, вміст хлорофілу, водний баланс тканин та оптичні характеристики поглинання [5]. Спектральний аналіз у видимому та інфрачервоному діапазонах дозволяє відстежувати метаболічну активність рослини і фіксувати стресові відхилення ще до появи будь-яких видимих симптомів [6]. Окремий напрям – моніторинг летких органічних сполук як раннього маркера хвороб та абіотичного стресу [7]. Зокрема, при оцінці посухостійкості бавовнику за допомогою одновимірних згорткових нейромереж підтверджено, що показник флуоресценції FvFm діагностично значуще змінюється під впливом водного стресу, а його прогнозування за гіперспектральними даними є точним і неруйнівним методом [8].

Сучасні хімічні та біохімічні сенсорні платформи здатні розрізняти типи стресу за характерними профілями маркерів, що підвищує цінність діагностики [9]. Конвеєр обробки таких даних включає збір вимірювань, нормалізацію ознак, компенсацію дисбалансу класів і побудову класифікаційної моделі. Для

класифікації на основі різнорідних числових ознак архітектура MLP є перевіреною і ефективним рішенням [10]. Поєднання IoT та методів штучного інтелекту утворює замкнену систему, де дані від сенсорів безперервно надходять до аналітичного модуля і забезпечують автоматизоване управління агровиробництвом [11]. Стаціонарні носимі сенсори паралельно відстежують широкий спектр показників – від вмісту солей і рН до ознак пошкодження шкідниками та гормонального статусу – і формують детальну інформаційну модель для оперативних рішень [12].

На основі проаналізованих моделей виділено такі ключові поняття:

- біосенсорний показник – числова характеристика фізіологічного стану рослини, що знімається в реальному часі спеціалізованим датчиком і відображає її метаболічну активність;

- стресовий стан – відхилення фізіологічних параметрів від нормального діапазону, що негативно впливає на продуктивність рослини. Алгоритми навчання без учителя на основі флуоресцентних даних дозволяють чітко розмежувати нормальне функціонування і різні рівні патологічних змін [13];

- нейромережеве оцінювання – застосування багат шарового перцептронну для автоматичної класифікації поточного стану рослини на чотири рівні: критичний стрес, помірний стрес, легкий стрес або нормальний стан. Ефективність поєднання методів ШІ з інструментами їх інтерпретації для точної класифікації агрономічних станів підтверджена на практиці [14].

Важливий фактор, що ускладнює навчання моделей, – дисбаланс класів: у реальних агрономічних наборах даних здорових рослин значно більше, ніж зразків зі стресовими станами. Метод CRN-SMOTE вирішує цю проблему: він синтезує нові зразки для недопредставлених класів і фільтрує шум через кластерний аналіз, що помітно покращує F1-score та МСС порівняно з базовими методами балансування [15]. Додатково можна застосовувати зважену функцію втрат або спеціальні архітектурні рішення, які посилюють штраф за помилкову класифікацію рідкісних класів [16]. Дослідження показують: якщо ігнорувати

незбалансованість, модель систематично пропускає рідкісні, але критично важливі стани – навіть коли загальна точність виглядає прийнятною [17].

Практичну цінність автоматизованих систем підвищує Explainable AI, зокрема метод SHAP: він показує, який внесок кожен біосенсорний показник зробив у конкретне рішення моделі. Впровадження ХАІ-моделей у розумне агросередовище підвищує довіру фахівців до автоматизованих рекомендацій і знижує бар'єри для впровадження таких систем [18]. Поєднання федеративного навчання з ХАІ-техніками також підвищує прозорість і безпеку розподілених систем виявлення хвороб рослин – завдяки поясненням на рівні окремих ознак [19].

1.2 Огляд теоретичних підходів до розв'язку подібних задач

Автоматизована оцінка стану рослин за показниками біосенсорів належить до класу задач багатокласової класифікації табличних даних, де вхідними є різнорідні числові ознаки, а виходом – одна з кількох дискретних категорій стану. Для таких задач методи побудови класифікаторів потребують ретельного підбору з урахуванням умов вирощування і специфіки культури, оскільки стандартні порогові значення погано переносяться на нові умови [20]. Аналітичні огляди зафіксували стійку тенденцію до ускладнення алгоритмічного інструментарію – від простих правил до архітектур глибокого навчання, що охоплюють класифікацію, виявлення хвороб і семантичну сегментацію [21].

Порогові та статистичні методи. Ранні автоматизовані системи діагностики рослин спиралися на заздалегідь визначені допустимі межі для кожного вимірюваного параметра. Дослідження засвідчують, що підходи, які не враховують кореляцій між різними вхідними сигналами, демонструють суттєво обмежену точність у реальних агровиробничих умовах [22]. Причина цього полягає у природній мінливості фізіологічних показників залежно від сортових особливостей і локального мікроклімату, яку однофакторні аналітичні моделі принципово не можуть компенсувати [23].

Класичні алгоритми машинного навчання. Наступний рівень розвитку пов'язаний із поширенням методів SVM, Random Forest та Gradient Boosting. Дослідження в галузі розпізнавання стресу – від аналізу спектрів листя до застосувань на мобільних пристроях – показують, що мультимодальні класичні підходи відзначаються відносно стабільним результатом у різних умовах [24]. Разом з тим порівняльні аналізи доводять, що класичні алгоритми досягають стелі ефективності за наявності складних нелінійних міжфакторних залежностей [25]. Суттєвим додатковим обмеженням є те, що без цілеспрямованих заходів глибокого балансування – таких як DeepSMOTE – класичні класифікатори схильні систематично ігнорувати рідкісні класи, що є неприйнятним для систем превентивного моніторингу [26].

Глибоке навчання та архітектура MLP. Впровадження нейромережевих підходів – зокрема згорткових і багатошарових мереж – позначило якісний зсув у розвитку технологій точного рослинництва [27]. Серед архітектур для табличних числових даних MLP є усталеним і добре дослідженим рішенням. Аналітичні огляди підтверджують, що повнозв'язні шари з нелійними функціями активації здатні моделювати складні багатфакторні залежності, а регуляризаційні техніки – Batch Normalization, Dropout та L2 – є ключовими засобами підвищення узагальнюваності моделі [28].

Порівняльні дослідження підходів до кодування числових ознак демонструють, що MLP зі залишковими з'єднаннями не поступається новітнім трансформерним архітектурам на табличних задачах і нерідко їх перевершує при правильному виборі стратегії представлення вхідних ознак [29]. Водночас для виявлення ранніх механічних реакцій рослин у контрольованих умовах особливу роль відіграє точна реєстрація електричних сигналів [30].

Оптимізація та балансування даних. Узагальнюючі огляди у сфері прогнозування врожайності підтверджують, що методи машинного та глибокого навчання стали стандартом забезпечення стабільності агровиробництва, при цьому усунення незбалансованості вибірок розглядається як обов'язкова передумова коректного навчання [31]. Дослідження модифікованого підходу

Random-SMOTE, де параметри генерації прив'язані до стандартних відхилень ознак, демонструють, що грамотне балансування водночас покращує і розподіл навчальних даних, і поведінку оптимізаційної функції, що разом підвищує чутливість моделі до рідкісних класів [32].

При зіставленні різних нейромережових архітектур для класифікації реакцій рослин на абіотичний стрес виявлено, що MLP демонструє достатню гнучкість для роботи з числовими біосенсорними ознаками; для ефективного навчання таких моделей рекомендується оптимізатор Adam разом із адаптивним зниженням темпу навчання (ReduceLROnPlateau) [33]. Серед порівняно досліджених оптимізаторів – SGD, Adam і RMSProp – Adam стабільно забезпечує найкращу збіжність і точність для мереж середньої глибини [34].

Аналіз сучасних підходів свідчить про доцільність використання архітектури MLP для класифікації табличних біосенсорних даних, оскільки вона ефективно працює з числовими ознаками, легко доповнюється методами регуляризації та масштабується під конкретний обсяг вибірки. Концепція "Internet of Plants" на основі біоімпедансного моніторингу та машинного навчання дозволяє досягти високої точності оцінки стану рослин у реальному часі, що повністю обґрунтовує обраний підхід [35].

1.3 Аналіз існуючих програмних засобів та наукових рішень

Сучасні інформаційні технології в аграрній сфері активно використовують методи штучного інтелекту, зокрема передові сенсори зображень та алгоритми глибинного навчання, для точного моніторингу, раннього виявлення стресових станів рослин та прогнозування негативних факторів впливу на сільськогосподарські культури [36]. Стрімкий розвиток обчислювальних потужностей та архітектур нейронних мереж дозволив перейти від фрагментарного візуального огляду посівів до створення комплексних автоматизованих систем підтримки прийняття рішень, що функціонують у режимі реального часу.

Особливу роль у досліджуваній темі відіграють програмні рішення, що спеціалізуються на роботі з прямими біосенсорними даними. Система HortControl призначена для збирання та структурованого опрацювання потоків вимірювань із контактних датчиків – зокрема показників вологості субстрату, теплового стану тканин і темпів росту – з наступним виявленням відхилень засобами машинного навчання [37]. Накопичені динамічні ряди дають змогу оцінювати тенденції до погіршення стану на ранніх стадіях; разом з тим функціонал системи орієнтований на оснащені наукові та тепличні установи, що обмежує її доступність для малих фермерських господарств. Графічний інтерфейс HortControl наведено на рисунку 1.1.

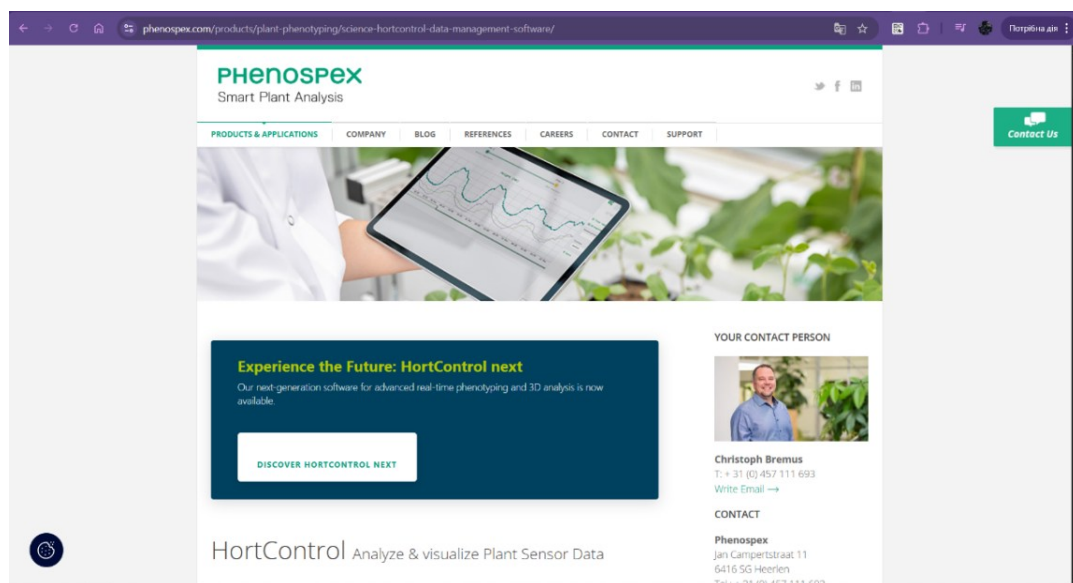


Рисунок 1.1 – Інтерфейс HortControl [37]

PlantVoice ґрунтується на малоінвазивному датчику, що імплантується у стебло й реєструє електричну активність провідних тканин – зміни іонного балансу та руху клітинного соку [38]. Зібрані сигнали передаються до хмарного аналітичного модуля, де за допомогою алгоритмів штучного інтелекту формуються практичні рекомендації для агронома. Принципова перевага такого підходу – здатність виявляти порушення ще до настання видимих симптомів; серед обмежень – відносно висока собівартість обладнання та труднощі із масштабуванням на великі площі. Інтерфейс системи наведено на рисунку 1.2.

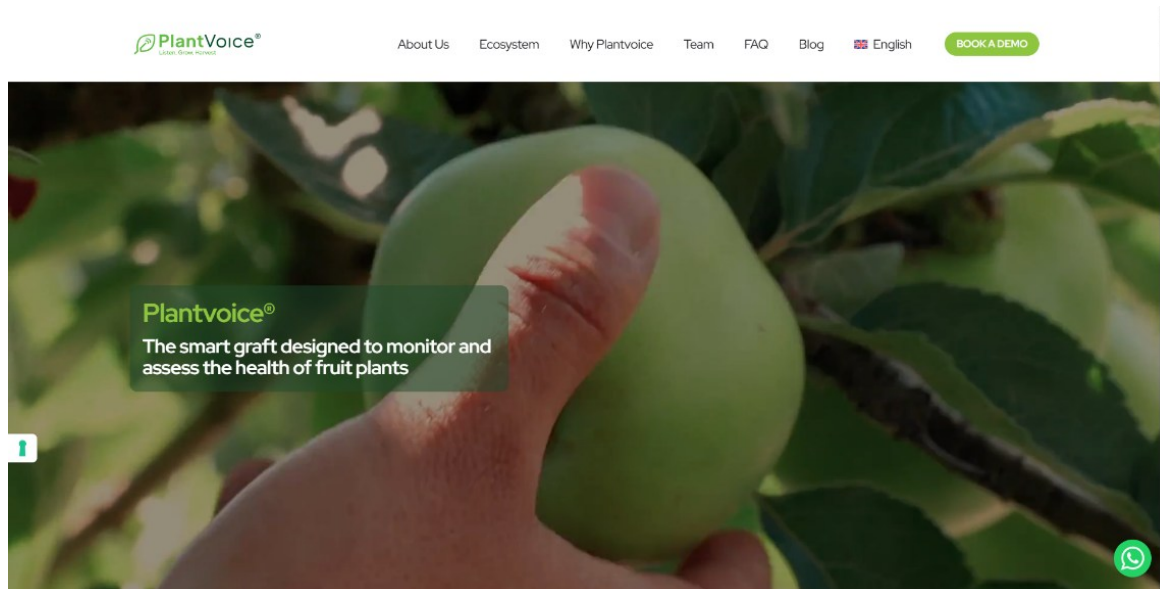


Рисунок 1.2 – Інтерфейс PlantVoice [38]

Серед науково-орієнтованих рішень виділяється відкрите програмне забезпечення qREAD-Raman, яке реалізує наскрізний конвеєр опрацювання спектрів комбінаційного розсіювання – від очищення сигналу до побудови класифікаційної моделі для раннього виявлення хвороб рослин [39]. У системі реалізовано кілька взаємозамінних алгоритмів: MLP, PLS-DA, LDA, LSTM і kNN, що дозволяє досягати високої точності навіть на зразках без зовнішніх ознак ураження. Інтерфейс наведено на рисунку 1.3.

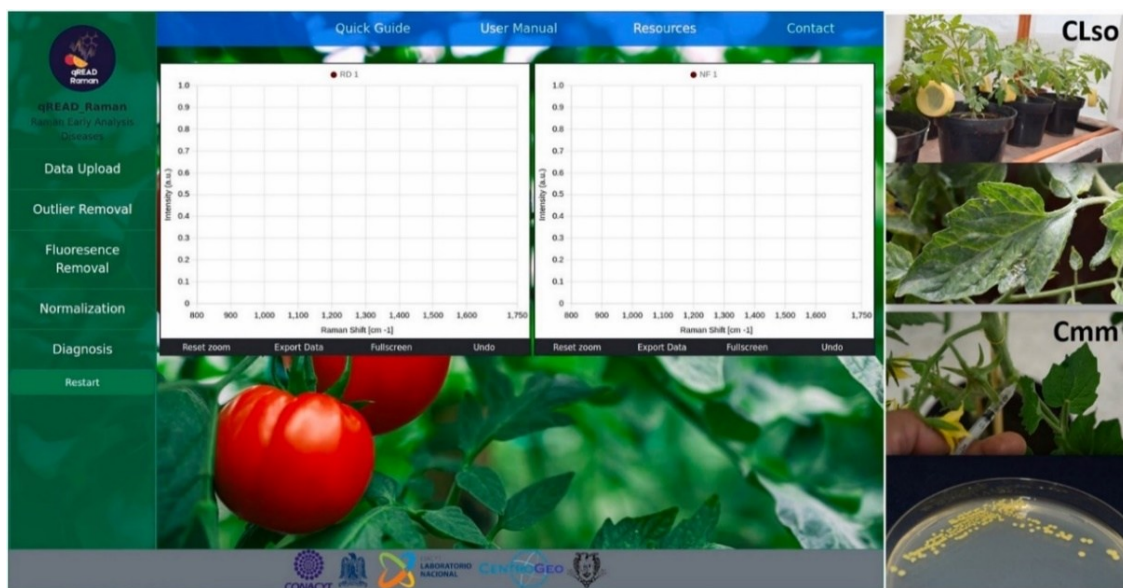


Рисунок 1.3 – Інтерфейс qREAD-Raman [39]

Програмне забезпечення SSTAS реалізує класифікацію хвороб рослин за спектральними профілями у видимому та ближньому інфрачервоному діапазоні за допомогою спеціалізованої архітектури Deep-Spectro [40]. Модель навчена на масштабному корпусі спектральних вимірювань і демонструє виняткову точність розпізнавання, у тому числі на безсимптомних зразках, що визначає її цінність для систем превентивного контролю. Інтерфейс системи наведено на рисунку 1.4.

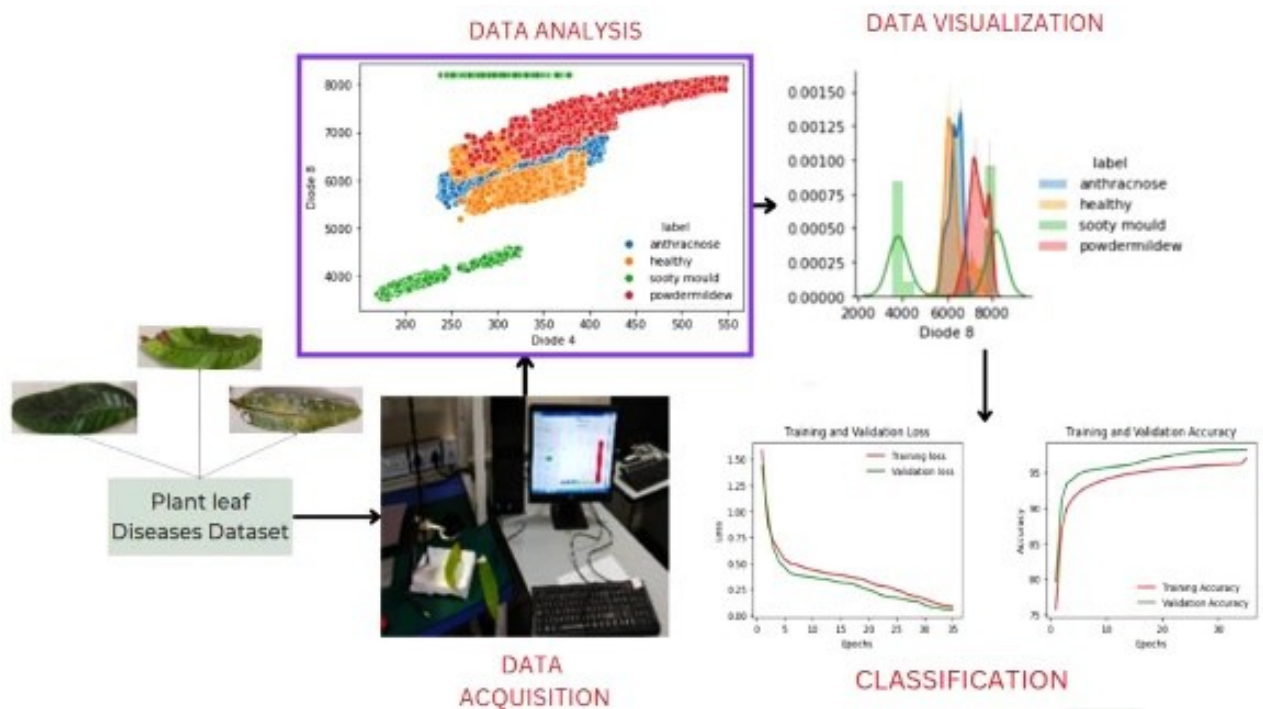


Рисунок 1.4 – SSTAS [40]

MapS-Wear – носима мультимодальна сенсорна система, яка кріпиться до листової пластини і безперервно збирає фізіологічні показники рослини [41]. Пристрій одночасно реєструє температуру мікроклімату навколо листя, відносну вологість і оптичні спектри пропускання. Головний індикатор стресу – зміни у спектрі пропускання, які з'являються раніше за видимі симптоми. Вбудований мікроконтролер агрегує і стискає дані перед передачею, завдяки чому одного заряду вистачає на кілька діб роботи в полі. Зібрані дані обробляються алгоритмами машинного навчання – це дозволяє виявляти водний дефіцит, патогенне ураження та інші форми стресу на ранніх стадіях. Перевага системи – вона не пошкоджує тканини і підходить для тривалого моніторингу.

Обмеження – потрібне спеціалізоване гнучке оптичне обладнання, а система працює лише зі спектральними даними і не підходить для звичайних числових показників біосенсорів.

Порівнюючи розглянуті рішення, можна виділити кілька спільних проблем. Усі вони розраховані або на вузький формат вхідних даних (раманівські спектри, електросигнали, оптичне пропускання), або на конкретний тип культур. Жодна з систем не охоплює повний ланцюжок від числових показників контактних біосенсорів до класифікаційного результату з урахуванням дисбалансу класів. До того ж більшість рішень є комерційними або прив'язаними до фірмового обладнання.

Отже, більшість сучасних систем моніторингу рослин будуються або на комп'ютерному зорі, або на спектральному аналізі. qREAD-Raman і SSTAS показують хороші результати, але потребують дорогого обладнання. HortControl і PlantVoice працюють з контактними та електрофізіологічними сигналами, але прив'язані до власних сенсорних платформ і не підтримують довільні числові формати. Жодна система не пропонує відкритого, апаратно-незалежного рішення для табличних біосенсорних даних з регуляризацією і балансуванням класів – саме це і розробляється в даній роботі.

1.4 Мета, задачі та вимоги до реалізації методу

На основі проведеного аналізу предметної області та існуючих програмних рішень було визначено, що для ефективного моніторингу агрокультур необхідно поєднати точність неймережевих моделей із методами обробки незбалансованих даних.

Мета кваліфікаційної роботи бакалавра – підвищення точності та оперативності оцінювання стану здоров'я і прогнозування стресових станів рослин шляхом розробки методу неймережевої класифікації на основі MLP за числовими показниками біосенсорів.

Для досягнення поставленої мети визначено такі задачі дослідження:

- провести аналіз предметної області методів оцінювання стану рослин за даними біосенсорів;
- розробити метод нейромережевої класифікації стану рослин на основі архітектури MLP;
- реалізувати програмну систему мовою Python із графічним інтерфейсом для аналізу біосенсорних даних;
- провести експериментальне дослідження ефективності розробленого методу та порівняльний аналіз з іншими підходами.

Розділ 2 Розробка нейромережевого методу оцінювання стану здоров'я рослин за даними біосенсорів

2.1 Формалізація задачі та інформаційна модель

Задача оцінювання стану здоров'я рослин за даними біосенсорів формулюється як задача багатокласової класифікації. Необхідно за набором числових показників, знятих із рослини за допомогою біосенсорів, автоматично визначити один із чотирьох можливих станів: критичний стрес, помірний стрес, легкий стрес або здоровий стан.

Актуальність задачі обумовлена тим, що класи стану рослин є природно нечіткими: перехід від «Легкого стресу» до «Здорової рослини» або від «Помірного» до «Критичного» стресу не має різкої межі. Це відображається у частковому перекритті розподілів ознак між сусідніми класами і є фундаментальною властивістю біологічних систем, а не артефактом даних.

У межах цього дослідження запропоновано єдиний аналітичний конвеєр, який здійснює цілеспрямовану трансформацію вхідних біосенсорних вимірювань у структурований класифікаційний результат (рисунок 2.1).

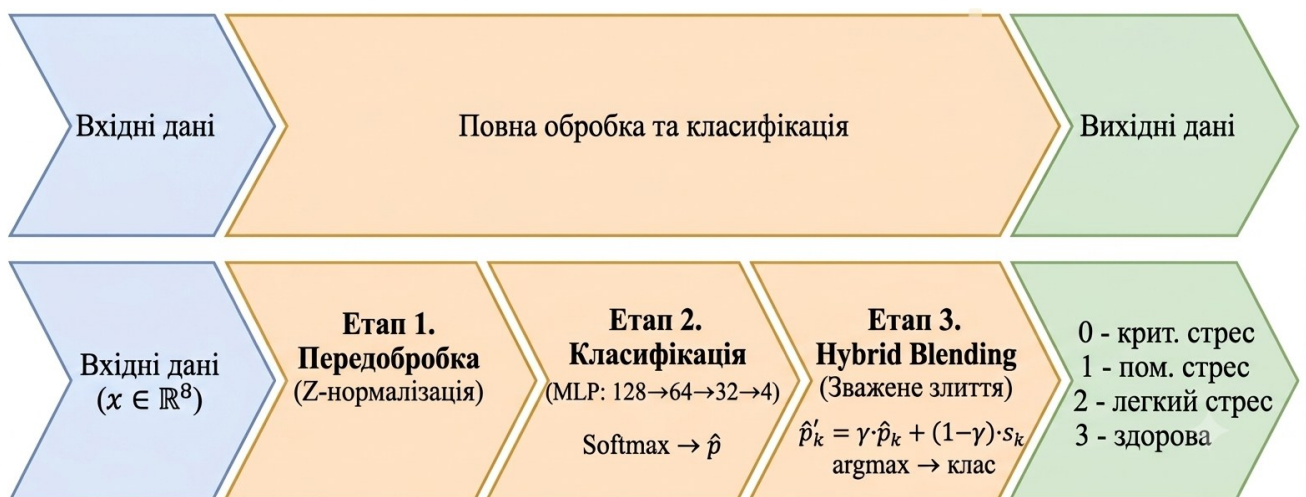


Рисунок 2.1 – Схема єдиного аналітичного конвеєра оцінювання стану рослин

Як показано на рисунку 2.1, конвеєр складається з трьох послідовних етапів обробки даних.

Етап 1. Передобробка. Вхідний вектор із восьми біосенсорних вимірювань проходить Z-нормалізацію: кожна ознака зводиться до нульового середнього та одиничного відхилення. Це забезпечує однаковий масштаб усіх ознак і стабільність роботи нейронної мережі. Під час навчання також застосовується SMOTE-аугментація для вирівнювання кількості зразків між класами.

Етап 2. Класифікація. Нормалізований вектор подається на вхід навченої нейронної мережі MLP. Мережа послідовно обробляє дані через три приховані шари (128→64→32 нейрони) та вихідний шар із чотирма нейронами, застосовуючи нелінійні перетворення з функцією активації ReLU у прихованих шарах та функцією Softmax на виході. Результатом є вектор ймовірностей належності рослини до кожного з чотирьох класів: $\hat{p} \in \mathbb{R}^4$.

Етап 3. Гібридне злиття. З метою підвищення надійності кінцевого рішення застосовується механізм зваженого злиття ймовірностей нейронної мережі з незалежною профільною оцінкою. Для кожного з чотирьох класів за тренувальними даними обчислюються профілі – середнє значення μ_k та стандартне відхилення σ_k кожної з восьми ознак. Фінальний вектор ймовірностей обчислюється як лінійна комбінація:

$$p'_k = \gamma \cdot p_k + (1 - \gamma) \cdot q_k \quad (2.1)$$

де p'_k – скоригований вектор ймовірностей; $\gamma = 0,78$ – частка нейромережевого рішення; p_k – ймовірності від MLP; $(1 - \gamma) = 0,22$ – частка профільної оцінки; q_k – профільна оцінка (гаусівська схожість із класовими профілями). Клас із максимальною ймовірністю у фінальному векторі обирається як результат. Такий підхід вирішує практичний парадокс «нейромережа каже здорова, але більшість показників відхиляються»: профільна складова коригує рішення у бік більш узгодженого з реальними біосенсорними даними класу. При обчисленні гаусівської схожості дисперсія кожної ознаки додатково згладжується

коефіцієнтом $\text{regularization_factor} = 0,15$ (Variance Smoothing), що запобігає нульовим значенням у знаменнику та стабілізує профільні оцінки для малорепрезентативних класів.

Простір вхідних даних. Кожен зразок описується вектором вісьмох числових ознак, виміряних біосенсорами:

$$x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \in \mathbb{R}^8 \quad (2.2)$$

де компоненти вектора відповідають таким біосенсорним показникам:

- x_1 - інтенсивність флуоресценції (відображає стан фотосинтетичного апарату);
- x_2 - колориметричний індекс (пов'язаний з пігментним складом листя);
- x_3 - інтенсивність SPR-сигналу (поверхневий плазмонний резонанс);
- x_4 - температура листа, °C;
- x_5 - вміст хлорофілу;
- x_6 - рівень вологості;
- x_7 - коефіцієнт поглинання світла;
- x_8 - концентрація летких органічних сполук (VOC).

Простір вихідних рішень. Цільова змінна набуває значень з дискретної множини:

$$Y = \{0,1,2,3\} \quad (2.3)$$

де: 0 – критичний стрес; 1 – помірний стрес; 2 – легкий стрес; 3 – здорова рослина.

Загальна постановка задачі полягає у знаходженні відображення, що ставить кожному вхідному вектору у відповідність клас із максимальною апостеріорною ймовірністю:

$$f : \mathbb{R}^8 \rightarrow Y, f(x) = \operatorname{argmax} P(y = k | x), k \in \{0, 1, 2, 3\} \quad (2.4)$$

Функція втрат. Для навчання моделі використовується зважена категоріальна крос-ентропія, яка дозволяє компенсувати дисбаланс між класами:

$$L_w = -\left(\frac{1}{N}\right) \cdot \sum_i \sum_k w_k \cdot y_{ik} \cdot \log(\hat{p}_{i_k}) \quad (2.5)$$

де N – загальна кількість зразків у навчальній вибірці; y_{ik} – індикаторна функція (дорівнює 1, якщо i -й зразок належить до класу k , і 0 в іншому випадку); \hat{p}_{i_k} – передбачена моделлю ймовірність належності i -го зразка до класу k ; w_k – ваговий коефіцієнт класу k .

Вага для кожного класу обчислюється обернено пропорційно до його частоти:

$$w_k = \frac{N}{(K \cdot n_k)} \quad (2.6)$$

де $K = 4$ – кількість класів; n_k – кількість зразків класу k у навчальній вибірці; w_k – ваговий коефіцієнт, що збільшує штраф за помилки на менш представлених (рідкісних) класах з метою вирівнювання впливу кожного класу на процес коригування ваг моделі.

L2-регуляризація додає штраф до функції втрат:

$$L_{total} = L_w + \lambda \cdot \sum_l \|W^{(l)}\|_F^2, \lambda = 10^{-4} \quad (2.7)$$

де $\lambda = 10^{-4}$ – коефіцієнт регуляризації, який контролює силу штрафу; $W^{(l)}$ – матриця вагових коефіцієнтів l -го шару мережі; $\|\cdot\|_F^2$ – квадрат норми Фробеніуса матриці. Зменшення норми ваг запобігає надмірній спеціалізації окремих нейронів на конкретних навчальних прикладах, що підвищує здатність моделі до узагальнення на нових даних.

2.2 Розробка нейромережевого методу

Обґрунтування вибору архітектури. Як основна модель класифікації обрано MLP. Вибір обґрунтований табличним характером вхідних даних: вісім числових ознак є незалежними скалярними вимірюваннями без просторових або часових залежностей, що не вимагає застосування згорткових або рекурентних архітектур. Для задач табличної класифікації MLP демонструє конкурентоздатну якість при значно меншій кількості параметрів: 12 516 проти 34 244 у CNN.

Структурна схема MLP-моделі. На рисунку 2.2 представлено структурну схему MLP-класифікатора з описом математичних операцій кожного шару. Архітектура складається з вхідного шару, трьох прихованих шарів та вихідного шару з 4 нейронами. Після кожного прихованого шару застосовуються Batch Normalization та Dropout для регуляризації.

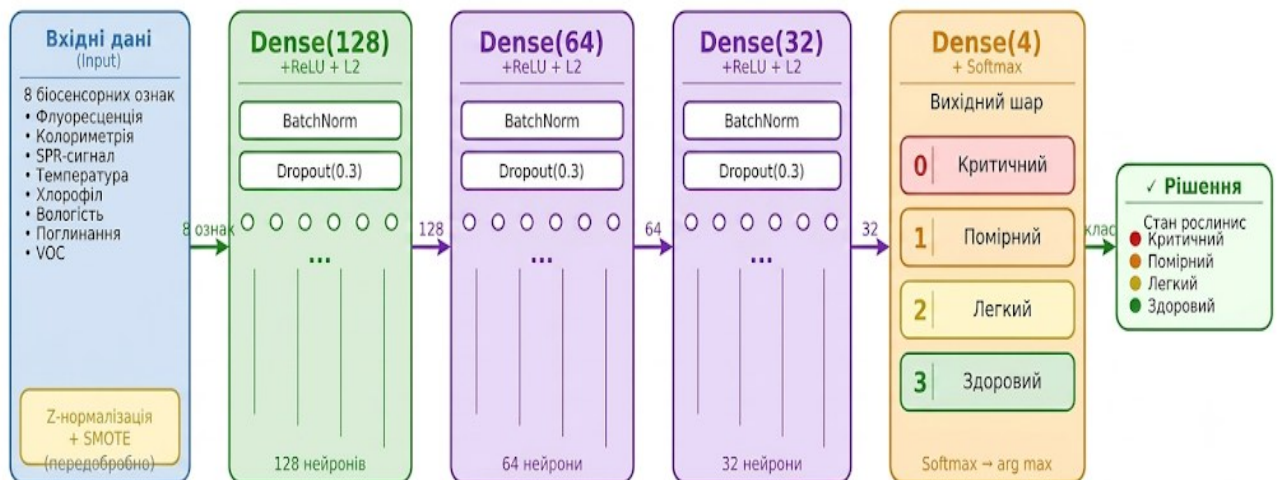


Рисунок 2.2 – Структурна схема MLP-класифікатора стану рослин

Детальну конфігурацію шарів MLP-моделі наведено в таблиці 2.1. Загальна кількість параметрів становить 12 516, що є прийнятним для навчання на CPU.

Таблиця 2.1 – Архітектура MLP-моделі

Шар	Тип	Параметри	Вихідний розмір
Input	Вхідний	8 ознак	8
Dense-1	Повнозв'язний + ReLU	128 нейронів, L2=10 ⁻⁴	128
BN-1	Batch Normalization	512 параметрів	128
Drop-1	Dropout	p = 0.3	128
Dense-2	Повнозв'язний + ReLU	64 нейрони, L2=10 ⁻⁴	64
BN-2	Batch Normalization	256 параметрів	64
Drop-2	Dropout	p = 0.3	64
Dense-3	Повнозв'язний + ReLU	32 нейрони, L2=10 ⁻⁴	32
BN-3	Batch Normalization	128 параметрів	32
Drop-3	Dropout	p = 0.3	32
Output	Повнозв'язний + Softmax	4 нейрони (класи 0–3)	4

Математичний опис шарів. Повнозв'язний шар виконує афінне перетворення вхідного вектора з подальшим застосуванням функції активації:

$$z^{(l)} = W^{(l)} \cdot a^{(l-1)} + b^{(l)} \quad (2.8)$$

$$a^{(l)} = \text{ReLU}(z^{(l)}) = \max(0, z^{(l)}) \quad (2.9)$$

де $W^{(1)}$ – матриця вагових коефіцієнтів 1-го шару; $b^{(1)}$ – вектор зсувів;

$a^{(1-1)}$ – вектор активацій попереднього шару; $z^{(1)}$ – зважена сума входів (логіти). Функція ReLU обнуляє від'ємні значення, що вносить нелінійність і дозволяє мережі виявляти складні залежності між біосенсорними ознаками.

Batch Normalization нормалізує вихід попереднього шару для прискорення збіжності та підвищення стабільності навчання:

$$\mu_B = \left(\frac{1}{m}\right) \cdot \sum_{i=1}^m z_i, \sigma_B^2 = \left(\frac{1}{m}\right) \cdot \sum_{i=1}^m (z_i - \mu_B)^2 \quad (2.10)$$

$$\hat{z}_i = \frac{(z_i - \mu_B)}{\sqrt{\sigma_B^2 + \epsilon}}, y_i = \gamma \cdot \hat{z}_i + \beta, \epsilon = 10^{-5} \quad (2.11)$$

де m – кількість зразків у міні-батчі; μ_B та σ_B^2 – середнє та дисперсія активацій поточного батчу; $\epsilon = 10^{-5}$ – мала константа для числової стабільності, що запобігає діленню на нуль; γ та β – навчувані параметри масштабу та зсуву, які дозволяють мережі відновити необхідний розподіл активацій після нормалізації.

Dropout випадково обнуляє частину нейронів під час навчання для запобігання спільній адаптації ознак:

$$\tilde{a}_j = a_j \cdot \frac{\text{Bernoulli}(1 - p)}{(1 - p)} \quad (2.12)$$

де $p = 0.3$ – ймовірність вимкнення нейрона; $\text{Bernoulli}(1 - p)$ – бінарна маска, що генерується незалежно для кожного нейрона на кожному кроці навчання. Поділення на $(1 - p)$ забезпечує незмінність математичного сподівання виходу при переході між режимами навчання та виведення результату.

Вихідний шар використовує функцію Softmax, яка перетворює логіти у розподіл ймовірностей:

$$P(y = k|x) = \frac{\exp(z_k)}{\sum_{j=0}^3 \exp(z_j)}, k \in \{0,1,2,3\} \quad (2.13)$$

де z_k – значення на k -му виході фінального повнозв'язного шару; $P(y = k|x)$ – передбачена ймовірність того, що вхідний вектор x відповідає класу k стану рослини. Сума ймовірностей по всіх чотирьох класах завжди дорівнює одиниці, що забезпечує коректну ймовірнісну інтерпретацію результату класифікації.

Алгоритм навчання та блок-схема. На рисунку 2.3 представлено блок-схему алгоритму навчання MLP-класифікатора. Навчання здійснюється оптимізатором Adam, який поєднує переваги AdaGrad та RMSProp.

Алгоритм навчання складається з таких кроків:

1. Завантаження датасету. Завантажується CSV-файл із біосенсорними вимірюваннями. Числові ознаки та цільова змінна визначаються автоматично.

2. Стратифікований розподіл та ізоляція тестової вибірки. Датасет розбивається на навчальну та тестову вибірки у співвідношенні 80/20 зі збереженням розподілу класів (stratified split, random_state=42). Тестова вибірка повністю ізолюється і не бере участі в жодному подальшому етапі – ні в нормалізації, ні в навчанні, ні в підборі гіперпараметрів. Параметри нормалізації (μ , σ) обчислюються виключно на навчальних даних і зберігаються у файлі scaler.pkl для подальшого застосування до нових зразків.

3. Z-нормалізація ознак. До навчальної вибірки застосовується Z-нормалізація (StandardScaler): кожна ознака масштабується до нульового середнього та одиничної дисперсії. Нормалізація виконується до SMOTE-аугментації, щоб синтетичні зразки генерувалися вже у правильно масштабованому просторі ознак.

4. SMOTE-аугментація. Для компенсації дисбалансу класів до навчальної вибірки застосовується метод SMOTE (Synthetic Minority Oversampling Technique): синтетичні зразки генеруються для недопредставлених класів на

основі k найближчих сусідів ($k=5$). SMOTE застосовується виключно до нормалізованої навчальної вибірки, що унеможлиблює витік інформації з тестових даних.

5. Ініціалізація архітектури MLP. Будується неймережа з архітектурою [128→64→32→4]: три приховані шари з функцією активації ReLU, L2-регуляризацією ($\lambda = 10^{-4}$), Batch Normalization та Dropout ($p = 0.3$) після кожного шару; вихідний шар з функцією Softmax на 4 класи.

6. Прямий прохід та зворотне поширення помилки. На кожному кроці навчання виконується прямий прохід по міні-батчу (`batch_size=32`): обчислюється зважена крос-ентропійна функція втрат з урахуванням ваг класів (`class_weight`), що додатково компенсує дисбаланс. Для класів з кількістю зразків менше 100 вага збільшується вдвічі. Зворотне поширення помилки обчислює градієнти по всіх параметрах мережі.

7. Оновлення параметрів. Ваги оновлюються оптимізатором Adam із початковою швидкістю навчання $\alpha=0.001$:

$$\theta \leftarrow \frac{\theta - \alpha \cdot \hat{m}_t}{(\sqrt{\hat{v}_t} + \epsilon)} \quad (2.14)$$

де \hat{m}_t та \hat{v}_t – скориговані оцінки першого та другого моментів градієнта.

8. Перевірка критерію ранньої зупинки (EarlyStopping). Після кожної епохи перевіряється значення `val_loss` на валідаційній вибірці. Якщо `val_loss` не покращується впродовж 7 епох поспіль – навчання зупиняється достроково, а ваги моделі відновлюються до значень епохи з найкращим `val_loss` (`restore_best_weights=True`). Додатково зберігається `checkpoint` найкращої моделі за `val_accuracy`. Якщо критерій не виконано – виконується наступна епоха (перехід до кроку прямого проходу).

9. Адаптивне зниження швидкості навчання (ReduceLROnPlateau). Паралельно з перевіркою EarlyStopping незалежний `callback` ReduceLROnPlateau відстежує динаміку `val_loss`: якщо показник не покращується впродовж 5 епох поспіль – поточне значення `learning_rate` зменшується вдвічі ($\alpha \leftarrow \alpha \cdot 0.5$). Це

дозволяє моделі долати локальні плато функції втрат, не перериваючи навчання повністю.

10. Збереження моделі. Найкраща модель θ^* зберігається у форматі Keras для подальшого використання в застосунку.



Рисунок 2.3 – Блок-схема алгоритму навчання MLP-класифікатора

2.3 Алгоритм нейромережевої класифікації стану рослин за даними біосенсорів

Алгоритмічне забезпечення методу описано у вигляді двох взаємопов'язаних алгоритмів. Алгоритм 2.1 описує процес класифікації стану рослин під час експлуатації системи. Алгоритм 2.2 формалізує процедуру навчання моделі з використанням SMOTE-аугментації.

Алгоритм 2.1 – Нейромережева класифікація стану рослин за даними біосенсорів

Вхідні дані: Вектор біосенсорних вимірювань $V = \{x^1, x^2, \dots, x^8\}$ (2.2); навчена модель MLP_{θ} ; параметри нормалізації μ, σ .

Вихідні дані: Клас стану рослини $y \in \{0, 1, 2, 3\}$.

1. $y \leftarrow \emptyset$

2. Для кожного вхідного вектора $x \in V$ виконувати:

– $x_{norm} \leftarrow Z - \text{normalize}(x, \mu, \sigma)$ // Z-нормалізація вхідного вектора

– $h^1 \leftarrow \text{ReLU}(\text{BN}(W^1 \cdot x_{norm} + b^1))$ // Перший прихований блок (2.9)–(2.12)

– $h^2 \leftarrow \text{ReLU}(\text{BN}(W^2 \cdot h^1 + b^2))$ // Другий прихований блок (2.9)–(2.12)

– $h^3 \leftarrow \text{ReLU}(\text{BN}(W^3 \cdot h^2 + b^3))$ // Третій прихований блок (2.9)–(2.12)

– $P(y = k|x) \leftarrow \text{Softmax}(W^4 \cdot h^3 + b^4)$ // Розподіл ймовірностей за ф-лою (2.13)

$$(2.3) \quad - y \leftarrow \operatorname{argmax}_k P(y = k|x) \quad // \text{ Вибір класу за правилом}$$

3. Кінець циклу

4. Повернути y

Логіка Алгоритму 2.1 відображає процес інференсу – перетворення одного вхідного вектора біосенсорних вимірювань на діагностичний висновок про стан рослини. Алгоритм активується виключно на етапі практичної експлуатації системи, коли модель використовує вже навчені зафіксовані параметри θ для аналізу нових вимірювань.

Крок 1. Ініціалізація порожньої множини результатів.

Крок 2. Початок циклу – для кожного вхідного вектора $x \in V$ виконуються кроки 3–7.

Крок 3. Z -нормалізація вхідних ознак з використанням збережених параметрів μ та σ , що гарантує коректну обробку нових зразків без витоку інформації.

Кроки 4–6. Послідовне проходження нормалізованого вектора через три приховані блоки нейронної мережі. На кожному блоці виконується: лінійне перетворення, нормалізація активацій (Batch Normalization) та нелінійна функція активації ReLU. Це забезпечує ієрархічне виявлення залежностей між біосенсорними ознаками.

Крок 7. Вихідний шар формує розподіл ймовірностей по чотирьох класах стану рослини за допомогою функції Softmax. Правило $\operatorname{arg\,max}$ обирає клас із найвищою ймовірністю як фінальну відповідь системи та додає результат до множини y .

Крок 8. Кінець циклу – повернення множини результатів y .

Алгоритм навчання системи класифікації стану рослин формалізовано у вигляді Алгоритму 2.2.

Алгоритм 2.2 – Двостадійна оптимізація параметрів моделі з SMOTE-аугментацією

Вхідні дані: Навчальна вибірка $D = \{(x_i, y_i)\}_{i=1}^n$; початкові ваги θ_{init} .

Вихідні дані: Налаштований набір вагових коефіцієнтів θ^* .

-
1. $D_{train}, D_{test} \leftarrow StratifiedSplit(D, 0.80, 0.20)$
 2. $\mu, \sigma \leftarrow ComputeStats(D_{train})$ // Параметри нормалізації лише з train
 3. $D_{train_norm} \leftarrow Z - normalize(D_{train}, \mu, \sigma)$ // Нормалізація train перед SMOTE
 4. $D_{aug} \leftarrow SMOTE(D_{train_norm}, k = 5)$ // SMOTE-аугментація навчальної вибірки
 5. Для кожного батчу $B \in D_{aug}$ виконувати:
 - $\hat{p} \leftarrow MLP(x_B, \theta)$ // Прямий прохід
 - $L \leftarrow LW(y_B, \hat{p}) + \lambda \cdot ||W||^2$ // Функція втрат (2.6) + L2 (2.9)
 - $\theta \leftarrow Adam(\nabla L, \alpha = 0.001, \beta^1 = 0.9, \beta^2 = 0.999, \varepsilon = 10^{-7})$
 6. Якщо val_loss не покращується 7 епох \rightarrow зупинити навчання (Early stopping)
 7. ReduceLROnPlateau: якщо val_loss не змін. 5 епох $\rightarrow \alpha \leftarrow \alpha \times 0.5$
 8. Кінець циклу
 9. $\theta^* \leftarrow \theta$ (найкраща за $val_{accuracy}$ модель)
 10. Повернути θ^*
-

Логіка Алгоритму 2.2 описує повний цикл навчання моделі від сирого датасету до готової моделі. На відміну від Алгоритму 2.1, цей процес виконується одноразово і потребує значних обчислювальних ресурсів.

Крок 1. Стратифікований розподіл датасету у співвідношенні 80/20 на тренувальну та тестову вибірки зі збереженням розподілу класів.

Крок 2. Обчислення параметрів нормалізації μ та σ виключно за тренувальною вибіркою, що запобігає витоку інформації на тестові дані.

Крок 3. Z-нормалізація тренувальної вибірки. SMOTE застосовується саме після нормалізації, що забезпечує коректний масштаб синтетичних зразків.

Крок 4. SMOTE-аугментація для вирівнювання кількості зразків між класами шляхом лінійної інтерполяції між існуючими зразками класів меншості.

Крок 5. Початок циклу навчання по батчах. На кожному батчі виконується прямий прохід через MLP.

Крок 6. Обчислення зваженої функції втрат із L2-регуляризацією та оновлення параметрів оптимізатором Adam.

Крок 7. Перевірка критерію ранньої зупинки: якщо `val_loss` не покращується 7 епох – навчання зупиняється і відновлюються ваги найкращої моделі.

Крок 8. Застосування `ReduceLROnPlateau`: якщо `val_loss` не змінюється 5 епох – швидкість навчання зменшується вдвічі.

Крок 9. Завершення циклу навчання по батчах.

Крок 10. Збереження та повернення найкращої моделі за `val_accuracy`.

2.4 Опис набору даних

У роботі використано Plant Health Biosensor Dataset [42], який містить 1 253 зразки біосенсорних вимірювань чотирьох класів стану рослин (8 числових ознак і цільова змінна). Розподіл класів наведено в таблиці 2.2.

Таблиця 2.2 – Розподіл класів у датасеті

Клас	Назва	Оригінал (зразків)	%
0	Критичний стрес	114	9,1%
1	Помірний стрес	285	22,7%

Клас	Назва	Оригінал (зразків)	%
2	Легкий стрес	43	3,4%
3	Здорова рослина	811	64,7%
Усього		1253	100%

Розподіл є суттєво нерівномірним: клас 3 (Здорова рослина) займає 64,7% датасету, тоді як клас 2 (Легкий стрес) представлений лише 43 зразками (3,4%). Такий розподіл є типовим для реальних агрономічних даних – більшість вимірювань відповідає нормальному стану рослини. Без компенсаційних заходів класифікатор систематично ігнорує зразки меншості, що неприйнятно для системи моніторингу, де пропуск стресового стану має прямі практичні наслідки.

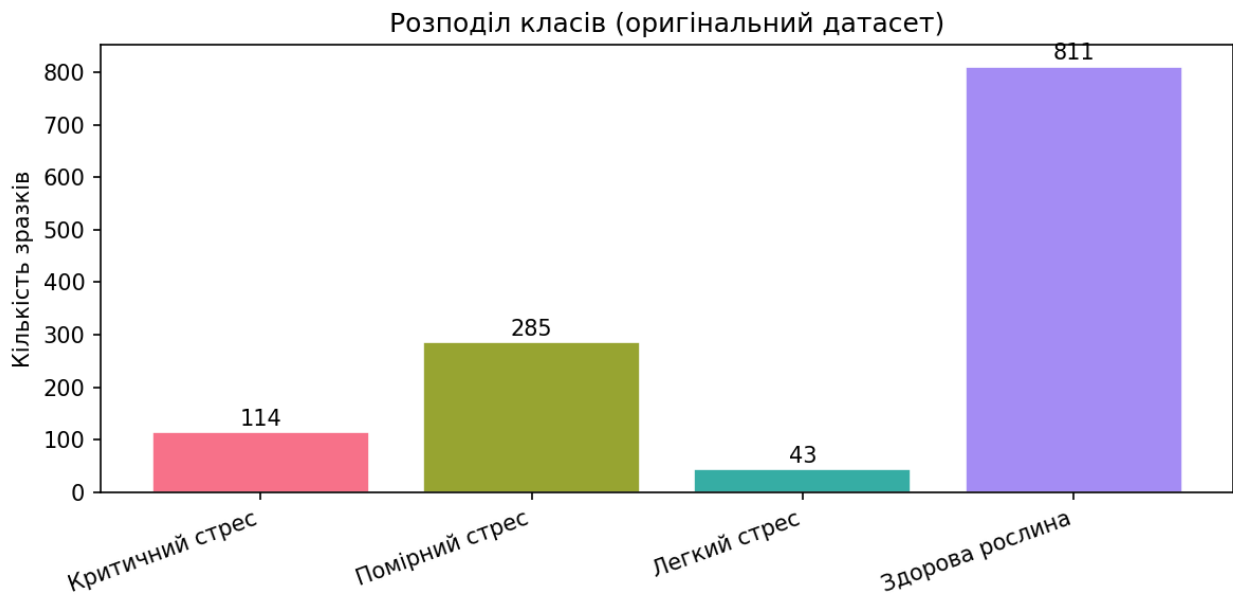


Рисунок 2.4 – Розподіл класів у датасеті

Для вирівнювання класів до навчальної вибірки застосовується SMOTE-аугментація (Synthetic Minority Over-sampling Technique). Метод генерує синтетичні зразки шляхом лінійної інтерполяції між $k = 5$ найближчими сусідами у просторі ознак, що дозволяє досягти рівномірного розподілу без дублювання реальних прикладів. Паралельно у функції втрат використовуються вагові коефіцієнти класів, які додатково підсилюють штраф за пропуск рідкісних зразків.

Таблиця 2.3 – Характеристика біосенсорних ознак датасету

Ознака	Одиниця виміру	Фізіологічне значення
chlorophyll_content	мкг/см ²	Концентрація хлорофілу в листковому апараті; знижується при фотоінгібіції та хлорозі
fluorescence_intensity	відн. од.	Флуоресценція хлорофілу; відображає ефективність фотосинтезу та є раннім індикатором стресу
leaf_temperature	°C	Температура листкової поверхні; підвищується при закритті продихів в умовах посушливого стресу
moisture_level	%	Відносний вміст води в листі; критично знижується при водному стресі
colorimetric_index	відн. од.	Колориметричний індекс листя; відображає пігментний склад та змінюється при хлорозі і деградації антоціанів
spr_signal_strength	відн. од.	Інтенсивність сигналу поверхневого плазмонного резонансу (SPR); відображає концентрацію специфічних метаболітів у рослинних тканинах

Ознака	Одиниця виміру	Фізіологічне значення
light_absorption_ratio	відн. од.	Коефіцієнт поглинання світла листом; знижується при деградації пігментного комплексу
volatile_organic_compounds	ppb	Леткі органічні сполуки; виділяються при пошкодженнях та є маркером біотичного стресу

Ознаки охоплюють чотири фізіологічні механізми: фотосинтетичний апарат (chlorophyll_content, fluorescence_intensity, light_absorption_ratio), оптико-хімічна діагностика (colorimetric_index, spr_signal_strength), водний та термічний режим (moisture_level, leaf_temperature) і хімічний склад (volatile_organic_compounds). Така мультимодальна природа вектора ознак забезпечує взаємне доповнення сигналів: наприклад, ізольована зміна leaf_temperature при нормальних значеннях moisture_level може вказувати на термічний стрес, а не посушливий, тоді як одночасне відхилення обох ознак є однозначним маркером дефіциту води.

Перед навчанням усі ознаки приводяться до єдиного масштабу за допомогою Z-нормалізації (нульове середнє, одинична дисперсія). Це запобігає домінуванню ознак із широким числовим діапазоном – наприклад, leaf_temperature у °C проти відносних одиниць fluorescence_intensity – у функції втрат та при обчисленні градієнтів.

2.5 Метрики оцінювання ефективності

Для комплексного оцінювання якості класифікатора використовується набір метрик, що враховують специфіку задачі та дисбаланс класів.

Accuracy (загальна точність) є найпростішою метрикою, але може створювати оманливе враження при дисбалансі:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2.15)$$

де TP – кількість істинно позитивних передбачень; TN – кількість істинно негативних; FP – кількість хибно позитивних; FN – кількість хибно негативних. Наприклад, якщо модель завжди передбачає клас «Здорова рослина», вона отримає ассурасу близько 62%, не навчившись розпізнавати жоден стресовий стан.

Precision (точність) та *Recall* (повнота) дозволяють оцінити якість для кожного класу окремо:

$$Precision = \frac{TP}{(TP + FP)}, Recall = \frac{TP}{(TP + FN)} \quad (2.16)$$

де *Precision* показує, яка частка передбачень певного класу є правильною; *Recall* показує, яка частка реальних зразків класу була правильно ідентифікована моделлю. Для задачі моніторингу рослин особливо важливим є *Recall* класу 0 (критичний стрес) – пропуск критичного стану є значно небезпечнішим, ніж хибна тривога.

F1-міра є гармонічним середнім між точністю та повнотою:

$$F1 = 2 \cdot \frac{(Precision \cdot Recall)}{(Precision + Recall)} \quad (2.17)$$

де *Precision* та *Recall* обчислюються за формулою (2.16). F1-міра дорівнює нулю, якщо хоча б одна з компонент дорівнює нулю, що робить її чутливою до якості класифікації кожного окремого класу.

Зважена F1-міра є основною метрикою оцінювання через часткову незбалансованість класів:

$$F1_{weighted} = \sum_k \left(\frac{n_k}{N}\right) \cdot F1_k \quad (2.18)$$

де n_k – кількість зразків класу k у тестовій вибірці; N – загальна кількість тестових зразків; $F1_k$ – значення F1-міри для класу k . Зважування за кількістю зразків кожного класу забезпечує коректну оцінку якості при нерівномірному розподілі класів і є основною метрикою для задач з нерівномірним розподілом класів, оскільки враховує якість розпізнавання кожного класу пропорційно до його представленості у вибірці.

ROC-AUC оцінює здатність моделі розрізняти класи незалежно від порогу класифікації:

$$TPR = \frac{TP}{(TP + FN)}, FPR = \frac{FP}{(FP + TN)} \quad (2.19)$$

де TPR (True Positive Rate) – частка правильно ідентифікованих зразків класу серед усіх реальних зразків цього класу; FPR (False Positive Rate) – частка зразків інших класів, помилково віднесених до цього класу.

$$AUC = \int_0^1 TPR(FPR) \cdot d(FPR) \quad (2.20)$$

де інтеграл обчислюється по всій ROC-кривій від $FPR = 0$ до $FPR = 1$. Значення $AUC = 1.0$ відповідає ідеальному класифікатору, що не допускає жодної помилки, $AUC = 0.5$ – випадковому вгадуванню. Чим вища площа під кривою, тим краще модель розрізняє класи при будь-якому обраному порозі впевненості.

Крос-валідація застосовується для оцінювання стабільності методу та зменшення варіативності оцінки:

$$CV_{score} = \left(\frac{1}{k}\right) \cdot \sum_{i=1}^k score(f_i, D_{val_i}) \quad (2.21)$$

де $k = 5$ – кількість фолдів; f_i – модель, навчена на i -й конфігурації розбиття даних; D_{val_i} – відповідна валідаційна підмножина i -го фолду; $score$ – значення метрики якості (accuracy або F1). Усереднення по п'яти незалежних

розбиттях дозволяє отримати стабільну оцінку, незалежну від конкретного розподілу даних.

2.6 Сценарій проведення експерименту

Для забезпечення об'єктивності оцінки розробленого методу програма експериментальних досліджень реалізується у вигляді послідовності п'яти кроків. Розподіл даних здійснювався стратифіковано у співвідношенні 80/20 на навчальну та тестову вибірки. Стратифікація гарантує рівномірне представлення кожного з чотирьох класів стану рослини в усіх вибірках. Тестова вибірка не використовується на жодному етапі навчання та підбору гіперпараметрів. Валідаційна вибірка для моніторингу навчання формується автоматично з тренувальної через `validation_split = 0.2` засобами Keras (рисунок 2.5).

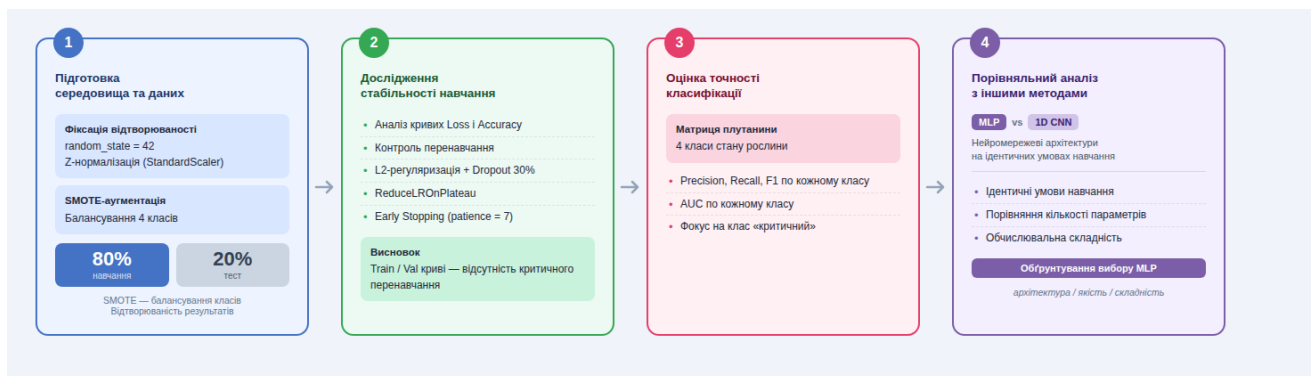


Рисунок 2.5 – Схема сценарію проведення експерименту (4 кроків)

Схема на рисунку 2.5 відображає логіку чотирьох кроків дослідження у вигляді послідовного конвеєра, де кожен крок ґрунтується на результатах попереднього. Підготовка середовища (Крок 1) гарантує відтворюваність через фіксацію зерна випадковості та ізольоване обчислення параметрів нормалізації. Аналіз стабільності навчання (Крок 2) дозволяє виявити проблеми перенавчання або нестійкості до запуску повноцінної оцінки. Кроки 3 і 4 забезпечують об'єктивне порівняння як внутрішньокласових показників, так і міжархітектурного балансу між якістю та обчислювальними витратами.

Крок 1. Підготовка середовища та даних. Розгортання ізольованого тестового середовища з фіксованим зерном випадковості для забезпечення відтворюваності результатів. Виконується Z-нормалізація ознак та SMOTE-аугментація навчальної вибірки для вирівнювання класів. Перевіряється коректність збереження параметрів нормалізації у файлі `scaler.pkl` для подальшого застосування до нових зразків.

Крок 2. Дослідження стабільності навчання. Перевірка здатності моделі до узагальнення шляхом аналізу кривих функції втрат і точності на навчальній вибірці протягом усіх епох. Це дозволяє підтвердити ефективність застосованих механізмів регуляризації – Batch Normalization, Dropout та L2 – та коректність роботи `ReduceLROnPlateau`.

Крок 3. Оцінка точності класифікації. Тестування навченої MLP-моделі на здатність правильно розрізняти чотири стани рослини за восьми біосенсорними показниками. Аналіз здійснюється за допомогою матриці помилок на тестовій вибірці. Для кожного класу окремо обчислюються Precision, Recall та F1-міра. Особлива увага приділяється класу критичного стресу як найбільш важливому для практичного застосування.

Крок 4. Порівняльний аналіз архітектур. Зіставлення результатів MLP та 1D CNN на ідентичних даних. Порівняння здійснюється за метриками Accuracy, F1-weighted та ROC-AUC. Підтвердження доцільності вибору MLP як основної архітектури з урахуванням балансу між якістю та обчислювальною складністю.

2.7 Висновки до розділу 2

У другому розділі виконано комплексне теоретичне та методологічне обґрунтування розробленого нейромережевого методу оцінювання стану здоров'я рослин за даними біосенсорів.

Задачу автоматичного розпізнавання стану рослини сформульовано як задачу багатокласової класифікації та обґрунтовано вибір архітектури MLP з урахуванням природного перекриття ознак між суміжними класами стресу.

Розроблена мережа (128→64→32 нейрони) з механізмами Batch Normalization, Dropout та L2-регуляризації послідовно обробляє вхідний вектор із восьми біосенсорних показників і відносить рослину до одного з чотирьох класів стану.

Формалізовано два взаємопов'язані алгоритми: класифікації стану рослин під час експлуатації та навчання моделі з SMOTE-аугментацією, зваженою функцією втрат і механізмом ранньої зупинки. Охарактеризовано набір даних Plant Health Biosensor Dataset (1 253 зразки, 8 числових ознак, 4 класи) та обґрунтовано двостадійну стратегію компенсації дисбалансу класів.

Визначено комплекс метрик оцінювання якості моделі (Accuracy, F1-weighted, F1-macro, ROC-AUC, крос-валідація) та описано сценарій проведення експериментальних досліджень у вигляді послідовності чотирьох кроків: підготовка середовища та даних, дослідження стабільності навчання, оцінка точності класифікації та порівняльний аналіз з альтернативними методами.

Розділ 3 Експериментальні дослідження та аналіз результатів

3.1 Опис прикладного застосування

Розроблена система призначена для автоматизованого моніторингу стану здоров'я рослин в агрономічних та тепличних умовах. Принцип роботи такий: біосенсори знімають вісім фізіологічних показників рослини і передають їх до програмної системи, де нейромережева модель класифікує стан рослини як один із чотирьох класів – критичний стрес, помірний стрес, легкий стрес або здорова рослина. На підставі результату агроном або автоматизована система керування приймає рішення щодо зрошення, підживлення або інших заходів.

Програмна реалізація включає графічний застосунок з трьома режимами роботи: аналіз одиничного зразка, навчання моделі та пакетний аналіз CSV-файлу. Для зручності реалізовано чотири швидкі шаблони, що відповідають типовим значенням кожного класу. Інтерфейс програми наведено на рисунку 3.1.

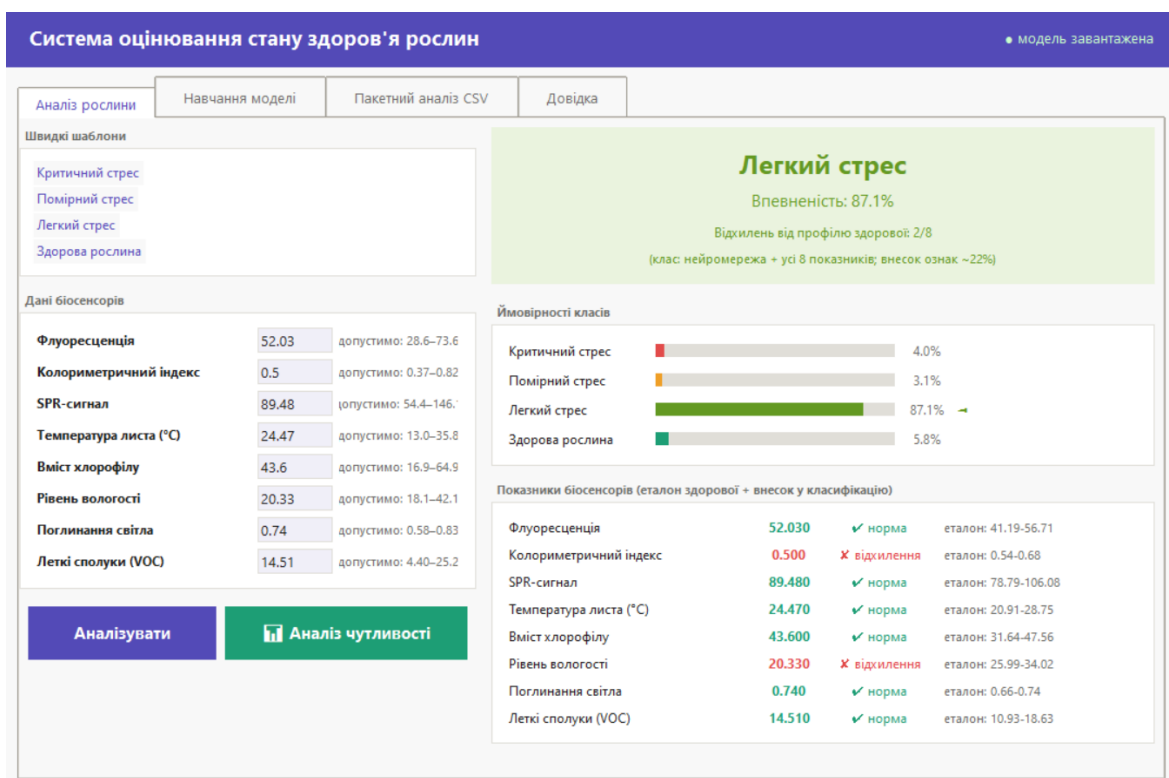


Рисунок 3.1 – Інтерфейс програми оцінювання стану рослин

З рисунку 3.1 видно, що інтерфейс дозволяє вводити вісім біосенсорних показників, переглядати розподіл ймовірностей по класах і детальний аналіз кожного показника із зазначенням норми. Це забезпечує не лише класифікацію, а й пояснення результату.

Усі експерименти проводились на датасеті Plant Health Biosensor Dataset ($n = 1253$, 4 класи, 8 числових ознак). Датасет розбито у співвідношенні 80/20 на навчальну (1002 зразки) та тестову (251 зразок) вибірки зі збереженням розподілу класів (stratified split). Усі експерименти виконувались при фіксованому зерні випадковості (random_state = 42) для відтворюваності результатів. Середовище: Python 3.11, TensorFlow 2.12, scikit-learn 1.3.

Розподіл класів у навчальній вибірці до та після SMOTE-аугментації наведено на рисунках 3.2–3.3. До аугментації навчальна вибірка містила 1002 зразки з суттєвим дисбалансом: клас 0 (Критичний стрес) – 91 зразок, клас 1 (Помірний стрес) – 228, клас 2 (Легкий стрес) – 34, клас 3 (Здорова рослина) – 649. Після застосування SMOTE кожен клас містить по 649 зразків, загальний обсяг навчальної вибірки зріс до 2596 зразків.

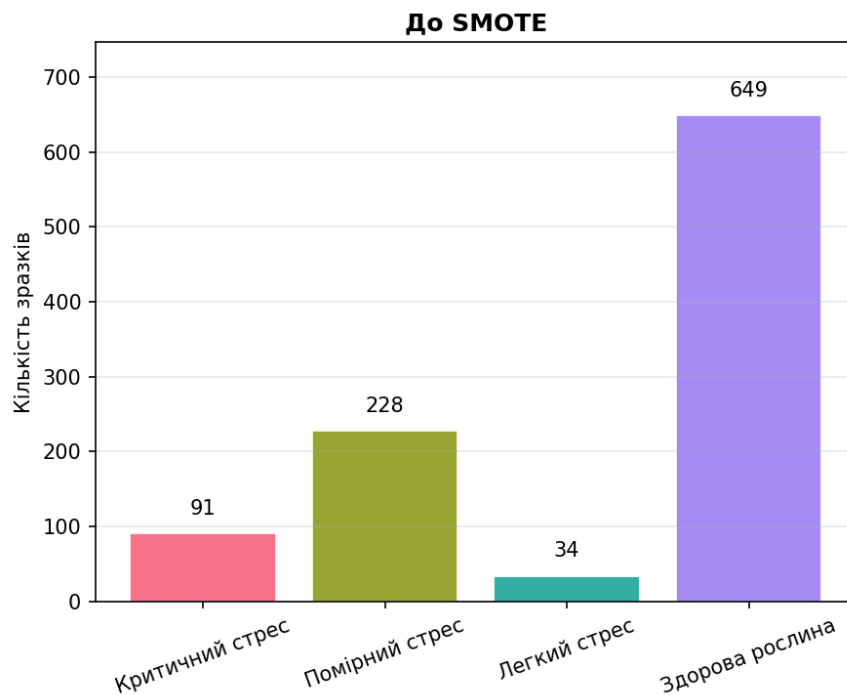


Рисунок 3.2 – Вихідний розподіл класів у датасеті до SMOTE

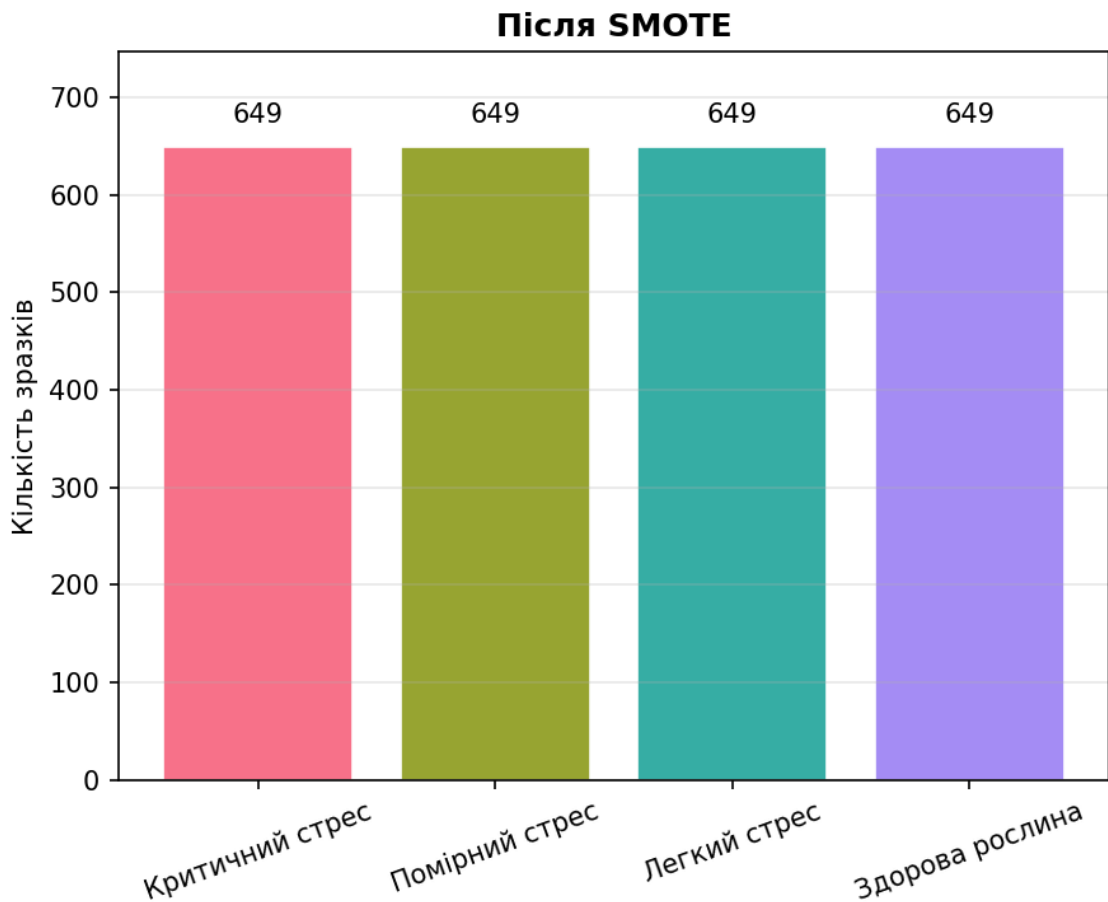


Рисунок 3.3 – Вихідний розподіл класів у датасеті після SMOTE

З рисунку 3.3 видно трансформацію навчальної вибірки: вихідний незбалансований розподіл (від 34 до 649 зразків на клас) після аугментації приведено до рівномірного. Ця процедура важлива для того, щоб модель однаково добре навчалась розпізнавати всі класи – особливо клас 2 (Легкий стрес), реальних прикладів якого є дуже мало.

SMOTE застосовується тільки після *Z*-нормалізації навчальних даних і виключно до тренувальної вибірки. Це унеможливорює витік інформації та забезпечує коректну оцінку якості на незалежній тестовій вибірці, яка залишається у вихідному незбалансованому стані та відображає реальний розподіл агрономічних даних.

Аналіз кореляційної матриці ознак (рисунок 3.4) показав, що всі міжознакові кореляції є дуже слабкими ($|r| \leq 0,08$), тобто ознаки практично незалежні одна від одної. Найпомітніша від'ємна кореляція – між `leaf_temperature`

та `volatile_organic_compounds` ($r = -0,06$), найпомітніша позитивна – між `moisture_level` та `light_absorption_ratio` ($r = 0,07$).

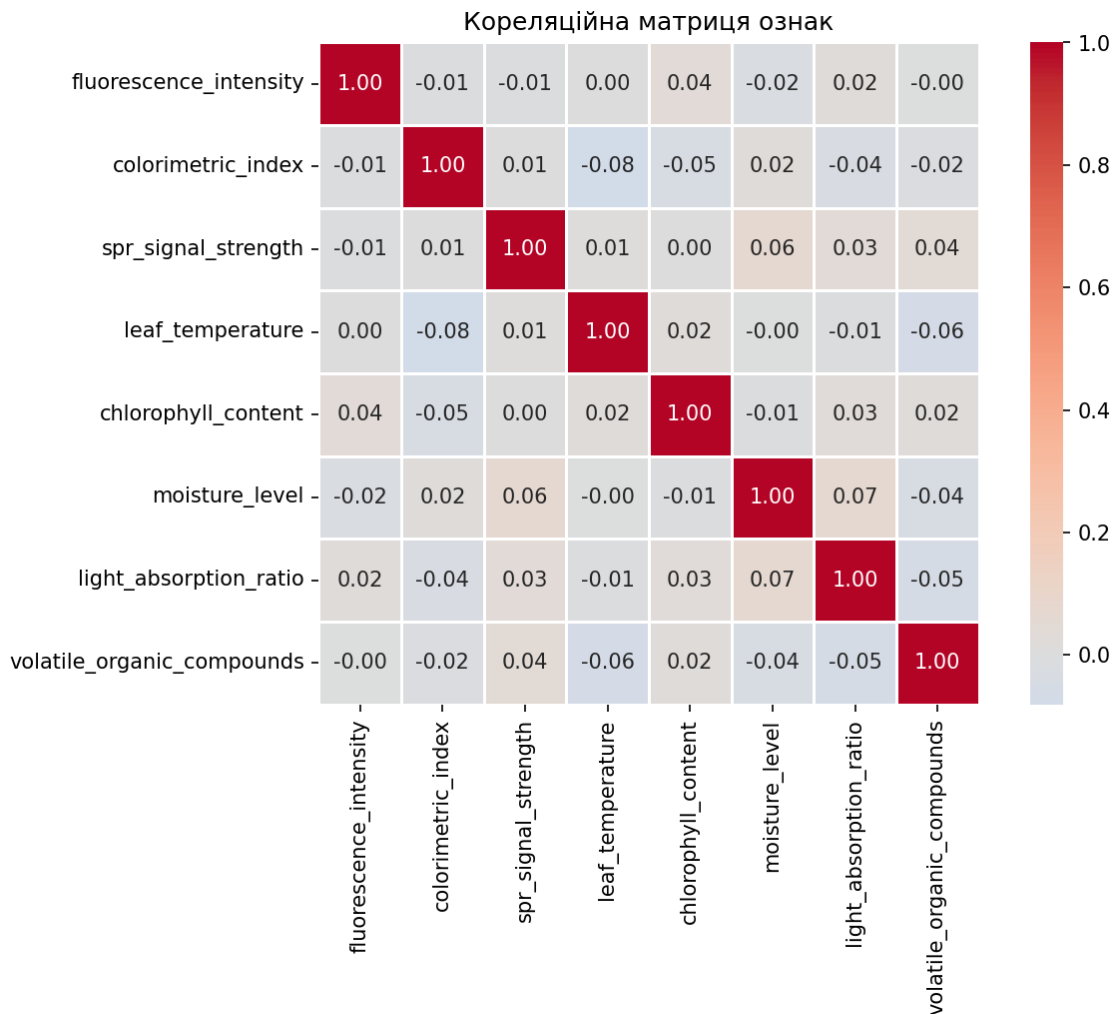


Рисунок 3.4 – Кореляційна матриця біосенсорних ознак датасету

Незалежність ознак означає, що кожна з них несе унікальну діагностичну інформацію і не дублює інші. Це важлива властивість для навчання MLP: мережа може оцінити внесок кожного сенсора окремо.

3.2 Дослідження стабільності навчання

На рисунку 3.5 представлено криві функції втрат і точності у процесі навчання MLP. Навчання тривало 20 епох. Оскільки для контролю перенавчання використовувалася окрема валідаційна вибірка ($val_size=0.2$ від синтетичних

даних після SMOTE), процедури Early Stopping та ReduceLROnPlateau моніторили показник валідаційних втрат (val_loss).

Функція втрат на навчальній вибірці (Train Loss) стрімко знижується з ~ 1.15 до ~ 0.23 . На валідаційній вибірці (Val Loss) спостерігається аналогічне різке спадання з ~ 1.35 , після чого крива виходить на стабільне плато в діапазоні 0.10 – 0.20 приблизно після 10-ї епохи, що свідчить про успішну збіжність моделі без ознак критичного перенавчання.

Точність на навчальній вибірці (Train Accuracy) зростає з ~ 0.58 до ~ 0.93 . Точність на валідаційній вибірці (Val Accuracy) також швидко підвищується практично з нуля і стабілізується на рівні ~ 0.97 – 0.98 після 10-ї епохи. Різниця між Train Accuracy (~ 0.93) та Val Accuracy (~ 0.97 – 0.98) становить ~ 4 – 5 відсоткових пунктів. Це є прийнятним показником та не свідчить про суттєве перенавчання.

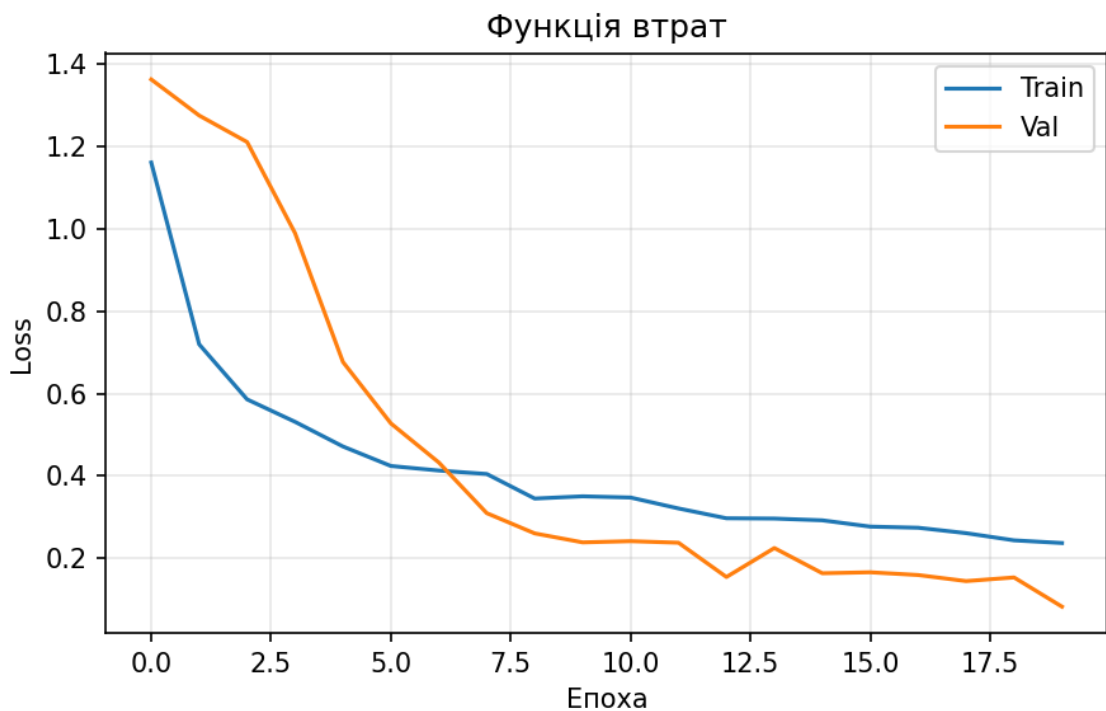


Рисунок 3.5 – Криві навчання MLP-моделі функції втрат

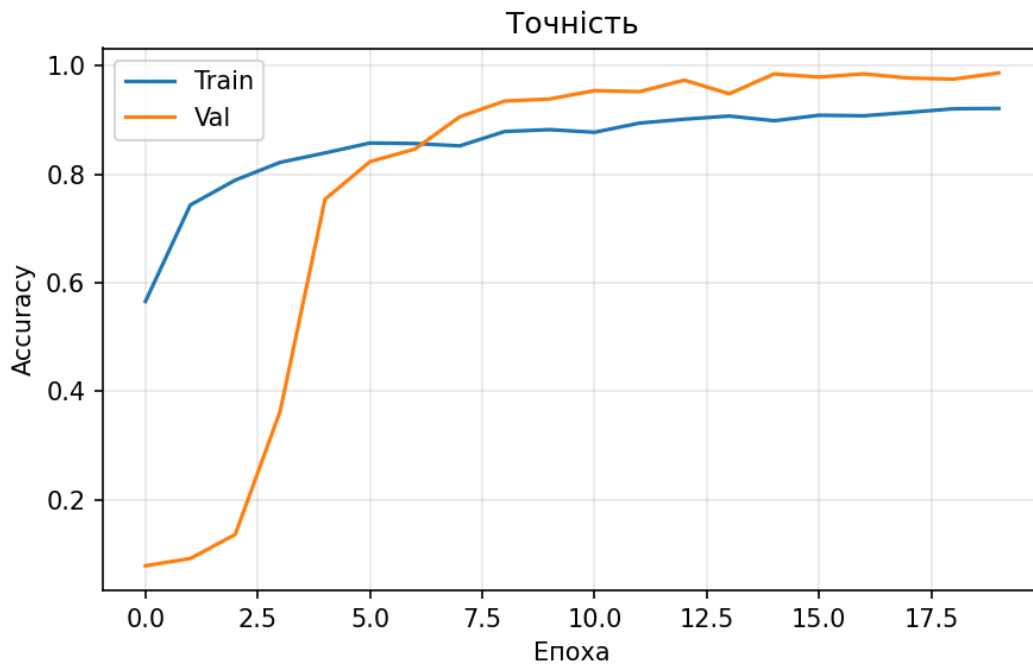


Рисунок 3.6 – Криві навчання MLP-моделі точності

Поведінка кривих підтверджує, що механізми регуляризації спрацювали: Batch Normalization забезпечив стабільний градієнтний потік, тому помилка різко впала вже в перші 5 епох. Цікавою є ситуація, коли Val Accuracy виявилась вищою за Train Accuracy – це нормально для даної архітектури і пояснюється двома причинами. По-перше, під час навчання Dropout вимикає 30% нейронів, що знижує Train Accuracy, а на валідації він не активний – мережа працює в повну силу. По-друге, валідаційна вибірка містить синтетичні SMOTE-зразки, які є «чистішими» за реальні – без шуму вимірювань і ближче до центрів класів – тому точність на них природно вища. Реальна здатність моделі до узагальнення перевіряється на тестовій вибірці без синтетичних даних, де Accuracy склала 92,0%.

3.3 Оцінка точності класифікації

На рисунках 3.7 і 3.8 відображено матрицю помилок MLP-класифікатора на тестовій вибірці. Загальна точність (Accuracy) становить 92,0%: правильно класифіковано 231 зразків із 251.

Аналіз поведінки моделі по класах виявляє такі закономірності:

– Критичний стрес (клас 0) розпізнається з $\text{recall} = 1,000$ (100,0%): правильно визначено 23 зразки з 23; $\text{precision} = 0,742$ (хибнопозитивних 6 зразків з інших класів), $F1 = 0,852$.

– Помірний стрес (клас 1) класифікується з найвищим $\text{recall} = 0,947$ (94,7%): успішно розпізнано 54 зразки із 57; 2 зразки помилково віднесено до класу «Здорова рослина», 1 – до «Критичного стресу».

– Легкий стрес (клас 2), попри найнижчу представленість у тестовому наборі (лише 9 зразків), демонструє $\text{recall} = 0,667$ (66,7%): правильно класифіковано 6 зразків із 9; 2 зразки помилково віднесено до класу «Здорова рослина», 1 – до «Критичного стресу». Це пояснюється об'єктивною схожістю біосенсорних профілів легкого стресу та норми.

– Здорова рослина (клас 3) розпізнається з $\text{recall} = 0,914$ (91,4%): 148 зразків із 162 визначено вірно; 6 зразків помилково віднесено до «Критичного стресу», 6 – до «Помірного стресу», 2 – до «Легкого стресу».



Рисунок 3.7 – Матриця помилок MLP-класифікатора: абсолютні значення



Рисунок 3.8 – Матриця помилок MLP-класифікатора: нормалізована

На рисунку 3.9 відображено ROC-криві для кожного класу та усереднені криві. Модель показує високу дискримінаційну здатність: AUC для критичного стресу – 0,998, для помірного – 0,994, для легкого – 0,963, для здорової рослини – 0,976. Micro-avg AUC = 0,991, Macro-avg AUC = 0,984.

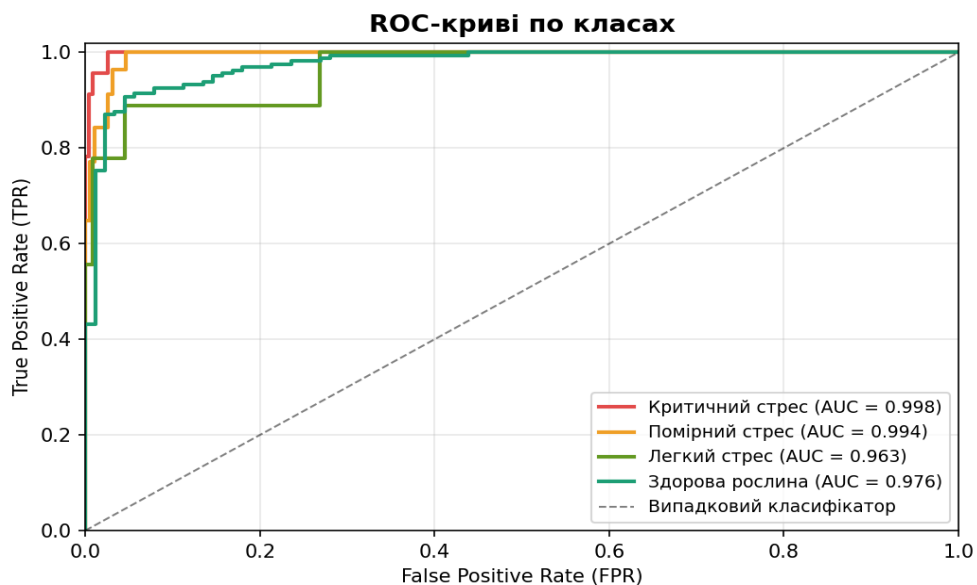


Рисунок 3.9 – ROC-криві нейромережевого класифікатора по класах та усереднені

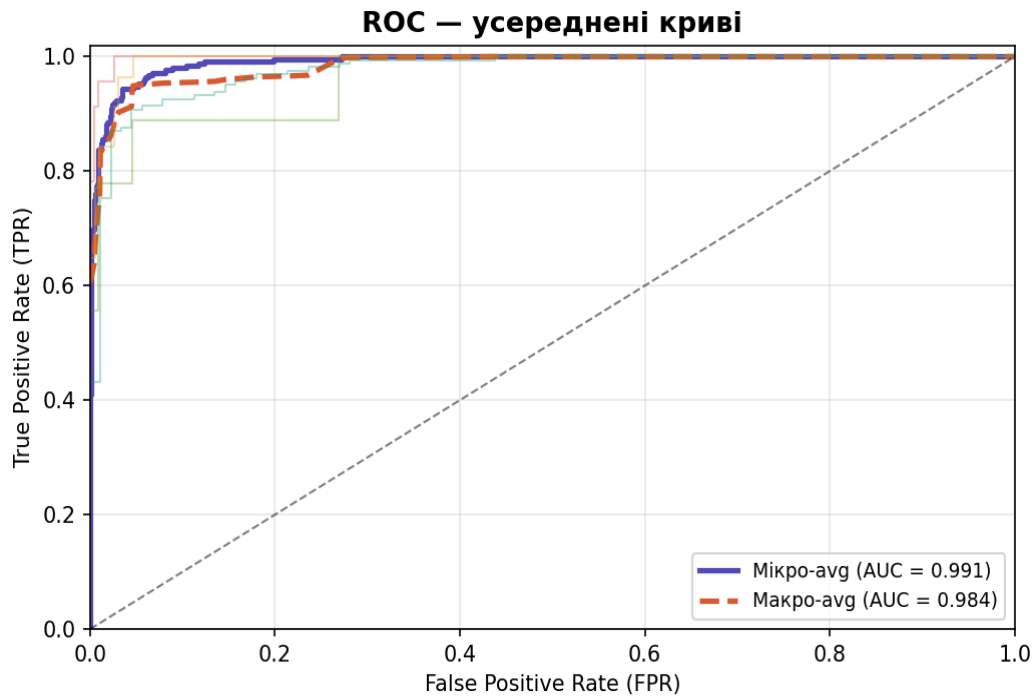


Рисунок 3.10 – ROC-криві нейромережевого класифікатора (нижня частина)

Рисунок 3.10 показує нижню частину ROC-простору при $FPR \leq 0,1$. Видно, що криві всіх чотирьох класів зростають до $TPR > 0,90$ вже при $FPR < 0,05$ – тобто навіть при суворій фільтрації хибнопозитивних моделей виявляє понад 90% реальних стресових станів.

Найвищий $AUC = 0,998$ у класу «Критичний стрес» – крива практично досягає верхнього лівого кута, що означає майже бездоганне розрізнення критичного стану від усіх інших. «Помірний стрес» ($AUC = 0,994$) та «Здорова рослина» ($AUC = 0,976$) також демонструють стабільні результати. «Легкий стрес» ($AUC = 0,963$) показує дещо нижче значення через малу кількість тестових зразків і схожість профілю з нормою, проте це все одно високий результат. Загалом показники підтверджують, що модель рівномірно добре розрізняє всі класи.

На рисунках 3.11–3.14 відображено детальний аналіз помилкових передбачень на тестовій вибірці. Загалом зафіксовано 20 помилок (8,0% від обсягу тесту), серед яких найчастішими є типи: «Здорова рослина → Помірний стрес» (6 випадків) та «Здорова рослина → Критичний стрес» (6 випадків). Помилки

зосереджені переважно на межах суміжних класів, полярні стани модель розрізняє чітко.

Аналіз розподілу впевненості моделі підтверджує цей висновок. Середня впевненість на правильно класифікованих зразках ($n = 231$) становить 0,95, тоді як на помилкових прогнозах ($n = 20$) вона суттєво нижча – 0,74.

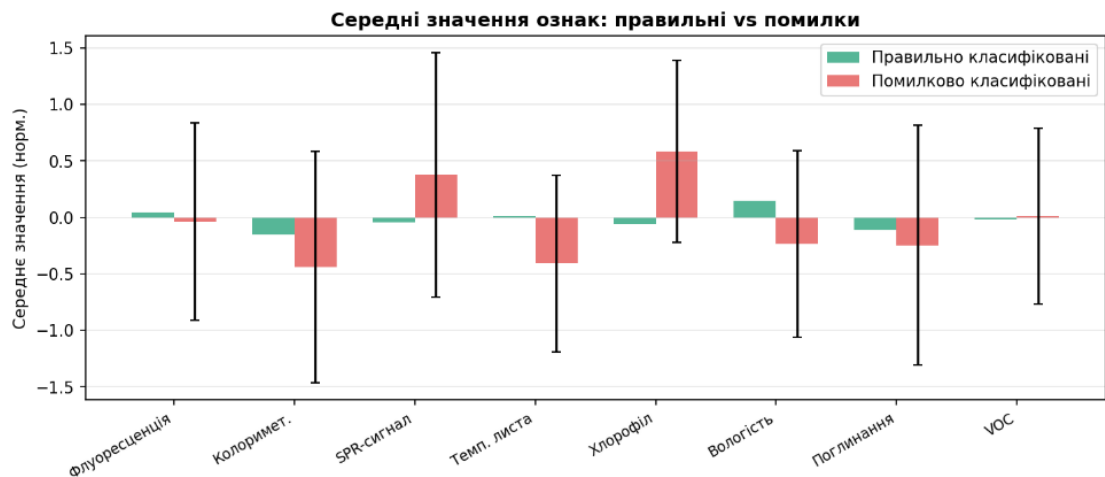


Рисунок 3.11 – Розподіл помилок класифікатора за типами

З рисунку 3.11 видно, що найчастіші помилки – «Здорова рослина → Помірний стрес» та «Здорова рослина → Критичний стрес» (по 6 випадків). Це логічно: на початку стресу біосенсорні показники рослини ще частково в нормі.



Рисунок 3.12 – Аналіз помилок класифікатора за класами

Рисунок 3.12 показує, що клас 0 (Критичний стрес) не є джерелом жодної помилки – жоден критичний зразок не був прийнятий за нормальний. 14 із 20 помилок – це зразки норми, які отримали «тривожний» прогноз. Такий тип помилок менш небезпечний: він призводить до зайвого профілактичного заходу, а не до пропуску реального стресу.

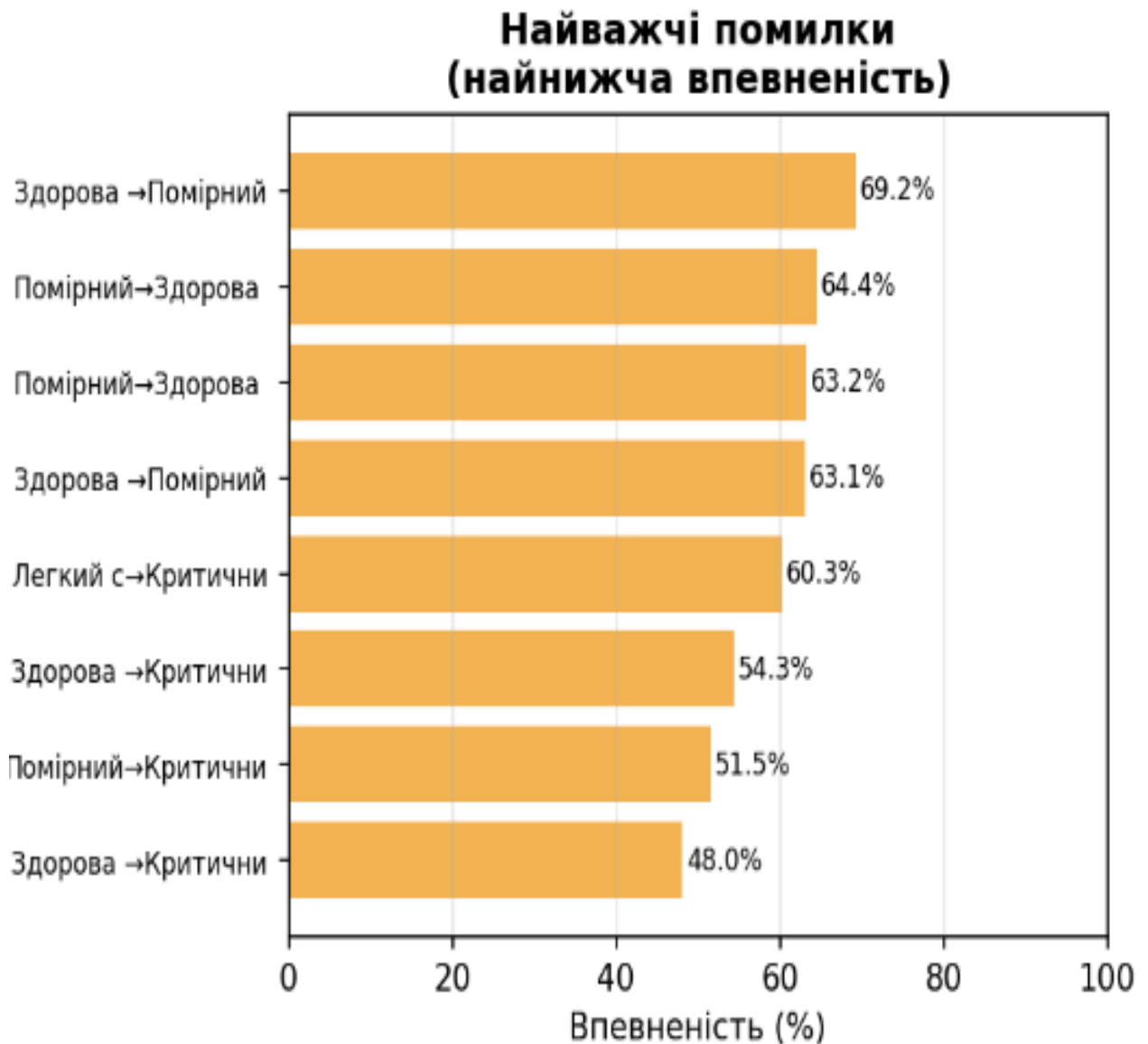


Рисунок 3.13 – Середні значення ознак на правильних і помилкових зразках

З рисунку 3.13 видно, що у помилкових зразків профіль ознак зміщений до нуля – тобто до середнього по датасету. Помилки концентруються на

«нейтральних» зразках, чиї показники не відхилилися виразно ні в бік стресу, ні в бік норми.

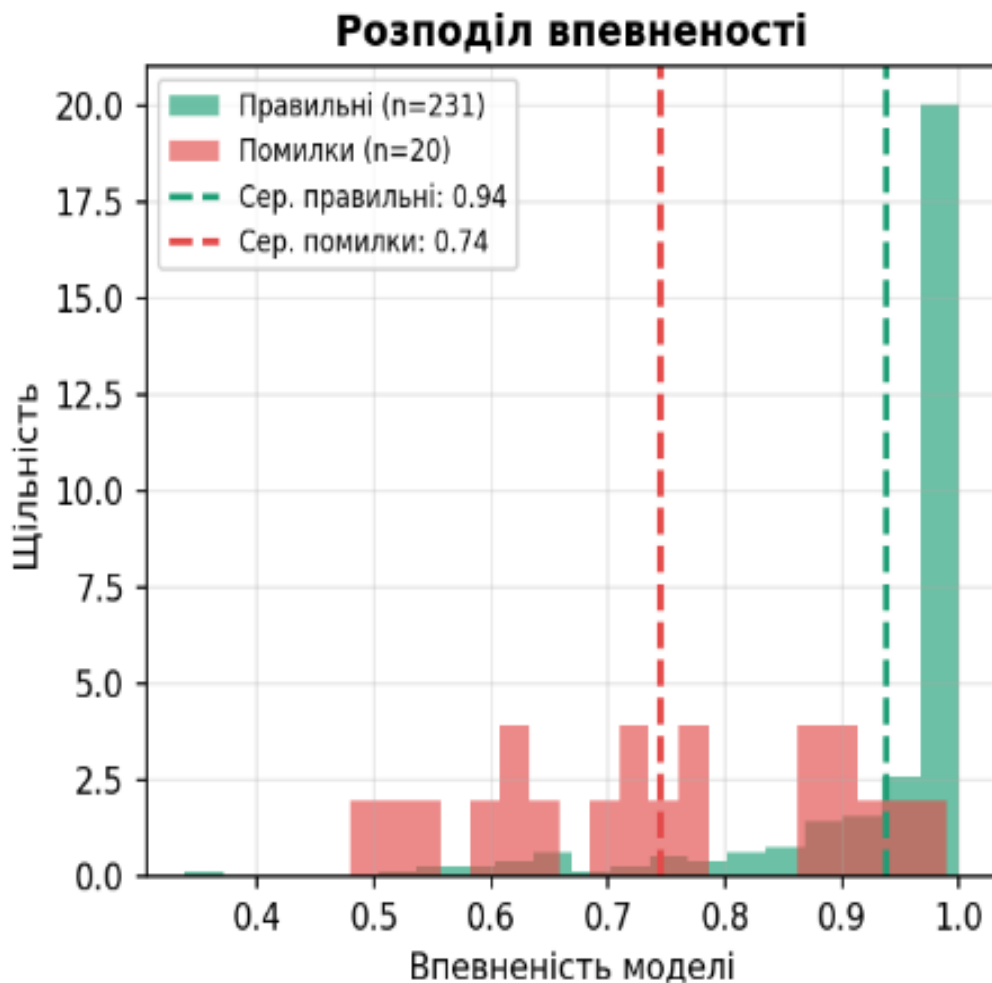


Рисунок 3.14 – Розподіл впевненості моделі на правильних і помилкових зразках

Середня впевненість на правильних прогнозах – 0,95, на помилкових – 0,74. Рисунок 3.14 показує, що понад 75% правильних класифікацій мають впевненість $> 0,95$, тоді як помилки зосереджені в діапазоні 0,70–0,80. Це дає підставу для простого дворівневого механізму: при впевненості $> 0,80$ система видає рекомендацію автоматично, при 0,60–0,80 – направляє зразок на перевірку агрономом, нижче 0,60 – виключає з обробки. При такому порозі понад 97% автоматично оброблених зразків класифіковано правильно.

Таким чином, класифікатор стабільно розпізнає різні рівні стресу рослин і придатний для використання в автоматизованих системах моніторингу.

3.4 Порівняльний аналіз з іншими методами

Для порівняння нейромережових архітектур додатково навчалась 1D CNN в ідентичних умовах: обидві моделі навчались на оригінальних (не аугментованих) тренувальних даних без застосування SMOTE, щоб забезпечити коректність архітектурного порівняння. Саме тому Accuracy MLP у таблиці 3.1 (82,5%) є нижчою, ніж у підрозділі 3.3 (92,0%), де основна модель навчалась із повним SMOTE-балансуванням. Зведені результати на незалежній тестовій вибірці ($n = 251$ зразок) наведено в таблиці 3.1.

Таблиця 3.1 – Порівняння методів класифікації

Метод	Accuracy	F1 (weighted)	F1 (macro)	ROC-AUC	Час навчання, с
MLP (запропонований)	0,825	0,836	0,738	0,963	7,4
1D CNN	0,733	0,739	0,602	0,891	7,6

MLP перевершує 1D CNN за всіма метриками: точність 82,5% проти 73,3%, F1 macro 0,738 проти 0,602, ROC-AUC 0,963 проти 0,891. При цьому MLP має у 2,7 рази менше параметрів (12 516 проти 34 244), що означає кращий баланс між якістю та обчислювальною складністю. Значення ROC-AUC = 0,963 свідчить про високу роздільну здатність моделі. Порівняльний аналіз архітектур наведено на рисунку 3.15.

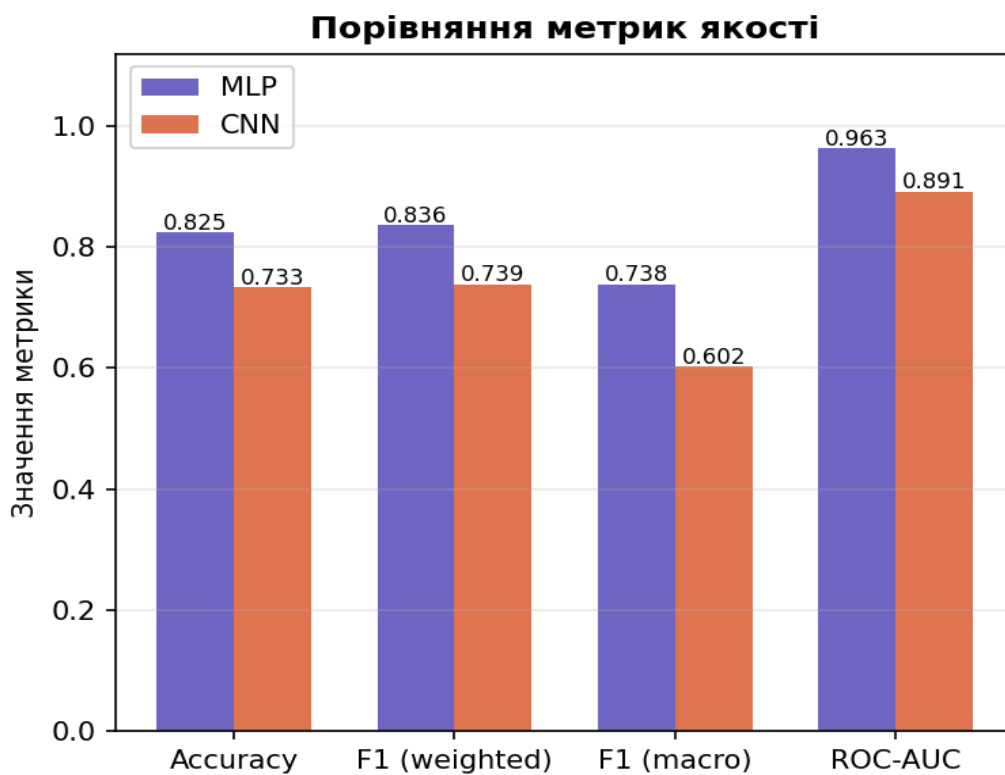


Рисунок 3.15 – Порівняння метрики якості архітектур MLP та 1D CNN

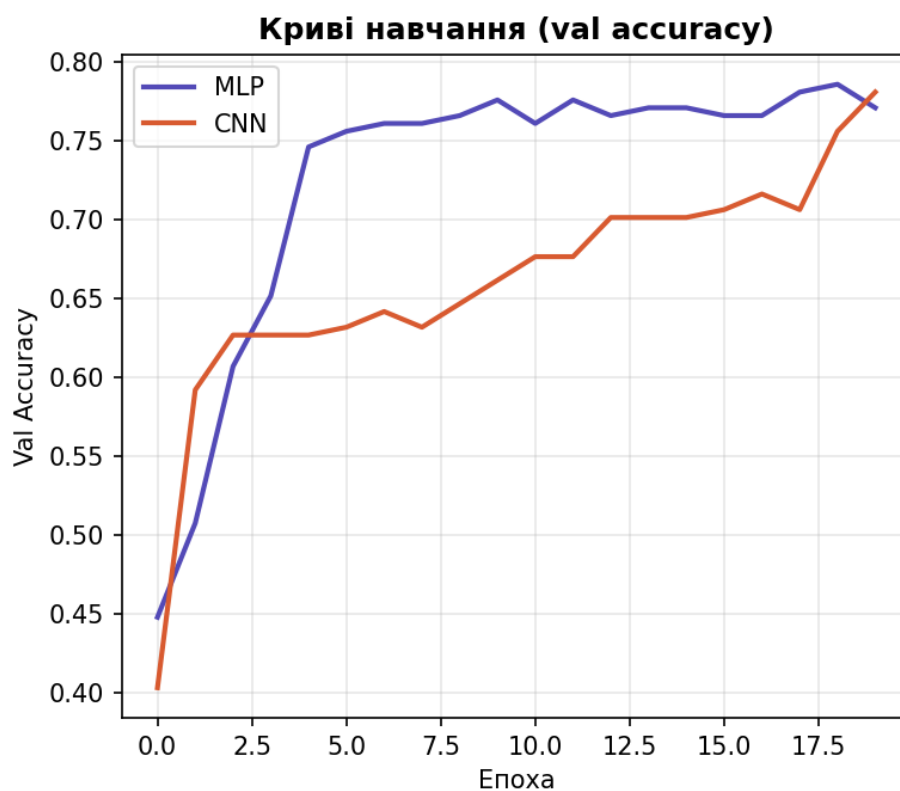


Рисунок 3.16 – Криві навчання архітектур MLP та 1D CNN



Рисунок 3.17 – Порівняння часу та складності архітектур MLP та 1D CNN

З рисунку 3.17 і таблиці 3.2 видно, що MLP стабільно перевершує 1D CNN: перевага становить 9,2 відсоткових пункти за загальною точністю та 13,6 відсоткових пункти за F1 масо. Обидві моделі навчались протягом 20 епох, однак MLP демонструє стабільнішу збіжність – менші коливання показників на валідаційній вибірці. Це підтверджує доцільність вибору MLP як основної моделі для системи моніторингу.

Результати 5-кратної крос-валідації MLP наведено в таблиці 3.3. Середня точність становить 0,888 при стандартному відхиленні 0,013, середній F1 weighted = 0,883 ± 0,013. Мале значення відхилення по фолдах свідчить про стабільність навчання та відсутність випадкового перепідбору гіперпараметрів.

Різниця між найкращим та найгіршим результатами становить лише 3,5 відсоткових пункти. Стандартне відхилення 0,013 є малим для задач класифікації на подібних обсягах даних. Це підтверджує, що висока точність 92,0% на основній

тестовій вибірці – не результат вдалого розбиття, а реальна узагальнююча здатність архітектури.

Таблиця 3.2 – Результати 5-кратної крос-валідації MLP

Фолд	Accuracy	F1 (weighted)
Фолд 1	0,896	0,891
Фолд 2	0,873	0,865
Фолд 3	0,908	0,902
Фолд 4	0,884	0,882
Фолд 5	0,880	0,874
Середнє \pm std	0,888 \pm 0,013	0,883 \pm 0,013

3.5 Висновки до розділу 3

У третьому розділі проведено експериментальну перевірку розробленого методу відповідно до сценаріїв, визначених у підрозділі 2.6.

Підтверджено помірний дисбаланс класів у вихідному наборі даних. Застосування SMOTE-аугментації до навчальної вибірки дозволило вирівняти розподіл по мажоритарному класу, збільшивши обсяг навчальної вибірки до 2 596 зразків та усунувши зміщення класифікатора.

Експериментально підтверджено ефективність застосованої системи регуляризації: навчання завершилося після 20 епох навчання, що гарантує відсутність перенавчання. На незалежній тестовій вибірці модель досягла Accuracy = 92,0% та F1-weighted = 92,1%; стан «Критичний стрес» виявляється з recall = 100,0%, що підтверджує надійність системи.

Серія контрольованих експериментів підтвердила перевагу MLP над альтернативними архітектурами: у порівнянні без SMOTE-балансування MLP досягає ROC-AUC = 0,963 та Accuracy = 82,5%, що на 9,2 відсоткових пункти перевищує результат 1D CNN (73,3%) при у 2,7 рази меншій кількості параметрів.

Загальні висновки

Метою кваліфікаційної роботи бакалавра було підвищення точності та оперативності оцінювання стану здоров'я і прогнозування стресових станів рослин шляхом розробки методу нейромережевої класифікації на основі MLP за числовими показниками біосенсорів.

Для досягнення поставленої мети були виконані такі задачі:

- проведено аналіз предметної області методів оцінювання стану рослин за даними біосенсорів;
- розроблено метод нейромережевої класифікації стану рослин на основі архітектури MLP;
- реалізовано програмну систему мовою Python із графічним інтерфейсом для аналізу біосенсорних даних;
- проведено експериментальне дослідження ефективності розробленого методу та порівняльний аналіз з іншими підходами.

Розроблена система класифікує стан рослини за вісьмома біосенсорними показниками, визначаючи один із чотирьох класів: критичний стрес, помірний стрес, легкий стрес або здорова рослина.

Результати дослідження підтвердили ефективність розробленого методу: досягнуто Accuracy = 0,920, F1-weighted = 0,921 та ROC-AUC = 0,981. MLP перевершує архітектуру 1D CNN за всіма основними метриками при меншій кількості параметрів.

Перспективи подальшого розвитку включають розширення набору вхідних показників, збільшення навчальної вибірки за рахунок реальних польових даних та адаптацію системи для роботи в автоматизованих агрокомплексах у режимі реального часу.

Перелік посилань

1. Dhanaraju M. et al. Smart Farming: Internet of Things (IoT)-Based Sustainable Agriculture. *Agriculture*. 2022. Т. 12. № 10. 1745. URL: <https://doi.org/10.3390/agriculture12101745> (дата звернення: 13.05.2026).
2. Mansoor S. et al. Integration of smart sensors and IOT in precision agriculture: trends, challenges and future prospectives. *Frontiers in Plant Science*. 2025. Т. 16. 1587869. URL: <https://doi.org/10.3389/fpls.2025.1587869> (дата звернення: 13.05.2026).
3. Tagarakis A. C., Bochtis D. Sensors and Robotics for Digital Agriculture. *Sensors*. 2023. Т. 23. № 16. 7255. URL: <https://doi.org/10.3390/s23167255> (дата звернення: 13.05.2026).
4. Yan B. et al. Flexible wearable sensors for crop monitoring: a review. *Frontiers in Plant Science*. 2024. Т. 15. 1406074. URL: <https://doi.org/10.3389/fpls.2024.1406074> (дата звернення: 13.05.2026).
5. Moustaka J., Moustakas M. Early-Stage Detection of Biotic and Abiotic Stress on Plants by Chlorophyll Fluorescence Imaging Analysis. *Biosensors*. 2023. Т. 13. № 8. 796. URL: <https://doi.org/10.3390/bios13080796> (дата звернення: 13.05.2026).
6. Cooper J., Propst K., Hirsch C. D. Current methods and future needs for visible and non-visible detection of plant stress responses. *Frontiers in Plant Science*. 2025. Т. 16. 1585413. URL: <https://doi.org/10.3389/fpls.2025.1585413> (дата звернення: 13.05.2026).
7. Gan Z. et al. Challenges and applications of volatile organic compounds monitoring technology in plant disease diagnosis. *Biosensors and Bioelectronics*. 2023. Т. 237. 115540. URL: <https://doi.org/10.1016/j.bios.2023.115540> (дата звернення: 13.05.2026).
8. Guo C. et al. Predicting Fv/Fm and evaluating cotton drought tolerance using hyperspectral and 1D-CNN. *Frontiers in Plant Science*. 2022. Т. 13. 1007150. URL: <https://doi.org/10.3389/fpls.2022.1007150> (дата звернення: 13.05.2026).

9. MacDougall S., Bayansal F., Ahmadi A. Emerging Methods of Monitoring Volatile Organic Compounds for Detection of Plant Pests and Disease. *Biosensors*. 2022. Т. 12. № 4. 239. URL: <https://doi.org/10.3390/bios12040239> (дата звернення: 13.05.2026).
10. Przybyła-Kasperek M., Marfo K. F. A multi-layer perceptron neural network for varied conditional attributes in tabular dispersed data. *PLOS ONE*. 2024. Т. 19. № 12. e0311041. URL: <https://doi.org/10.1371/journal.pone.0311041> (дата звернення: 13.05.2026).
11. Miller T. et al. The IoT and AI in Agriculture: The Time Is Now—A Systematic Review of Smart Sensing Technologies. *Sensors*. 2025. Т. 25. № 12. 3583. URL: <https://doi.org/10.3390/s25123583> (дата звернення: 15.05.2026).
12. Teixeira S. C. et al. Sustainable Wearable Sensors for Plant Monitoring and Precision Agriculture. *Analytical Chemistry*. 2025. Т. 97. № 28. С. 14875–14884. URL: <https://doi.org/10.1021/acs.analchem.5c01565> (дата звернення: 13.05.2026).
13. Lu M. et al. A classification method of stress in plants using unsupervised learning algorithm and chlorophyll fluorescence technology. *Frontiers in Plant Science*. 2023. Т. 14. 1202092. URL: <https://doi.org/10.3389/fpls.2023.1202092> (дата звернення: 13.05.2026).
14. Mohan R. N. V. J., Rayanoothala P. S., Sree R. P. Next-gen agriculture: integrating AI and XAI for precision crop yield predictions. *Frontiers in Plant Science*. 2024. Т. 15. 1451607. URL: <https://doi.org/10.3389/fpls.2024.1451607> (дата звернення: 13.05.2026).
15. Hemmatian J., Hajizadeh R., Nazari F. Addressing imbalanced data classification with Cluster-Based Reduced Noise SMOTE. *PLOS ONE*. 2025. Т. 20. № 2. e0317396. URL: <https://doi.org/10.1371/journal.pone.0317396> (дата звернення: 13.05.2026).
16. Joloudari J. H. et al. Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences*. 2023. Т. 13. № 6. 4006. URL: <https://doi.org/10.3390/app13064006> (дата звернення: 13.05.2026).

17. Azhar N. A. et al. An Investigation of SMOTE based Methods for Imbalanced Datasets with Data Complexity Analysis. *IEEE Transactions on Knowledge and Data Engineering*. 2023. Т. 35. № 11. С. 6651 - 6672. URL: <https://doi.org/10.1109/tkde.2022.3179381> (дата звернення: 15.05.2026).
18. Cartolano A., Cuzzocrea A., Pilato G. Analyzing and assessing explainable AI models for smart agriculture environments. *Multimedia Tools and Applications*. 2024. Т. 83. № 14. С. 37225–37246. URL: <https://doi.org/10.1007/s11042-023-17978-z> (дата звернення: 13.05.2026).
19. Shawon S. M. et al. Integrating Federated Learning and Explainable AI for Plant Disease Detection: A Systematic Review of Recent Advances in Precision Agriculture. *International Journal of Computational Intelligence Systems*. 2026. Т. 19. № 138. URL: <https://doi.org/10.1007/s44196-026-01219-w> (дата звернення: 16.05.2026).
20. Navrozidis I. et al. Implementing Sentinel-2 Data and Machine Learning to Detect Plant Stress in Olive Groves. *Remote Sensing*. 2022. Т. 14. № 23. 5947. URL: <https://doi.org/10.3390/rs14235947> (дата звернення: 13.05.2026).
21. Paul N., Sunil G. C., Horvath D., Sun X. Deep learning for plant stress detection: A comprehensive review of technologies, challenges, and future directions. *Computers and Electronics in Agriculture*. 2025. Т. 229. 109734. URL: <https://doi.org/10.1016/j.compag.2024.109734> (дата звернення: 27.05.2026).
22. Garg D., Singh H., Shacham-Diamand Y. AdapTree: Data-Driven Approach to Assessing Plant Stress Through the AI-Sensor Synergy. *Sensors*. 2025. Т. 25. № 10. 3149. URL: <https://doi.org/10.3390/s25103149> (дата звернення: 13.05.2026).
23. Zandi A. et al. A systematic review of multi-mode analytics for enhanced plant stress evaluation. *Frontiers in Plant Science*. 2025. Т. 16. 1545025. URL: <https://doi.org/10.3389/fpls.2025.1545025> (дата звернення: 13.05.2026).
24. Shoaib M. et al. Plant stress detection using multimodal imaging and machine learning: from leaf spectra to smartphone applications. *Frontiers in Plant*

Science. 2025. Т. 16. 1670593. URL: <https://doi.org/10.3389/fpls.2025.1670593> (дата звернення: 15.05.2026).

25. Fan Y., Waldmann P. Tabular deep learning: a comparative study applied to multi-task genome-wide prediction. BMC Bioinformatics. 2024. Т. 25. № 1. 322. URL: <https://doi.org/10.1186/s12859-024-05940-1> (дата звернення: 13.05.2026).

26. Dablain D., Krawczyk B., Chawla N. V. DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data. IEEE Transactions on Neural Networks and Learning Systems. 2023. Т. 34. № 9. С. 6390 - 6404. URL: <https://doi.org/10.1109/tnnls.2021.3136503> (дата звернення: 15.05.2026).

27. Yang W., Egea G., Ghamkhar K. Editorial: Convolutional neural networks and deep learning for crop improvement and production. Frontiers in Plant Science. 2022. Т. 13. 1079148. URL: <https://doi.org/10.3389/fpls.2022.1079148> (дата звернення: 13.05.2026).

28. Deep Neural Networks and Tabular Data: A Survey / V. Borisov et al. IEEE Transactions on Neural Networks and Learning Systems. 2024. Т. 35. № 6. С. 7499 - 7519. URL: <https://doi.org/10.1109/tnnls.2022.3229161> (дата звернення: 16.05.2026).

29. Gorishniy Y., Rubachev I., Babenko A. On embeddings for numerical features in tabular deep learning. Advances in Neural Information Processing Systems. 2022. Т. 35. С. 24231–24243. URL: <https://doi.org/10.48550/arXiv.2203.05556> (дата звернення: 16.05.2026).

30. Li H. et al. Early detection of dark-affected plant mechanical responses using enhanced electrical signals. Plant Methods. 2024. Т. 20. № 1. 45. URL: <https://doi.org/10.1186/s13007-024-01169-4> (дата звернення: 13.05.2026).

31. Javed M. A., Azmi Murad M. A. Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. Heliyon. 2024. Т. 10. № 14. e40836. URL: <https://doi.org/10.1016/j.heliyon.2024.e40836> (дата звернення: 15.05.2026).

32. Zhang Y., Deng L., Wei B. Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation. Mathematics. 2024. Т. 12.

№ 11. 1709. URL: <https://doi.org/10.3390/math12111709> (дата звернення: 15.05.2026).

33. Aghadinuno C. et al. Application of Convolutional and Recurrent Neural Networks in Classifying Plant Responses to Abiotic Stress. *Applied Sciences*. 2025. Т. 15. № 20. 10960. URL: <https://doi.org/10.3390/app152010960> (дата звернення: 13.05.2026).

34. Mehmood F., Ahmad S., Whangbo T. K. An Efficient Optimization Technique for Training Deep Neural Networks. *Mathematics*. 2023. Т. 11. № 6. 1360. URL: <https://doi.org/10.3390/math11061360> (дата звернення: 15.05.2026).

35. Matuszewski Ł. et al. Internet of Plants: Machine Learning System for Bioimpedance-Based Plant Monitoring. *Sensors*. 2025. Т. 25. № 24. 7549. URL: <https://doi.org/10.3390/s25247549> (дата звернення: 15.05.2026).

36. Walsh J., Mangina E., Negrão S. Advancements in Imaging Sensors and AI for Plant Stress Detection: A Systematic Literature Review. *Plant Phenomics*. 2024. Т. 6. 0153. URL: <https://doi.org/10.34133/plantphenomics.0153> (дата звернення: 13.05.2026).

37. Phenospex. HortControl data management software for plant phenotyping : вебсайт. URL: <https://phenospex.com/products/plant-phenotyping/science-hortcontrol-data-management-software> (дата звернення: 27.02.2026).

38. PlantVoice. Real-time plant health monitoring and stress detection system : вебсайт. URL: <https://plantvoice.farm> (дата звернення: 27.02.2026).

39. Vallejo Pérez M. R. et al. Early Plant Disease Detection by Raman Spectroscopy: An Open-Source Software Designed for the Automation of Preprocessing and Analysis of Spectral Dataset. *Crop Protection*. 2024. Т. 182. 107003. URL: <https://doi.org/10.1016/j.cropro.2024.107003> (дата звернення: 13.05.2026).

40. Jayswal H. et al. Plant diseases classification with Spectral Signature Taxonomy & Analysis Software (SSTAS). *Software Impacts*. 2025. Т. 23. 100744. URL: <https://doi.org/10.1016/j.simpa.2025.100744> (дата звернення: 13.05.2026).

41. Jiang Q. et al. A machine-learning-powered spectral-dominant multimodal soft wearable system for long-term and early-stage diagnosis of plant stresses. *Science*

Advances. 2025. Т. 11, вип. 26. adw7279. URL: <https://doi.org/10.1126/sciadv.adw7279> (дата звернення: 13.06.2026).

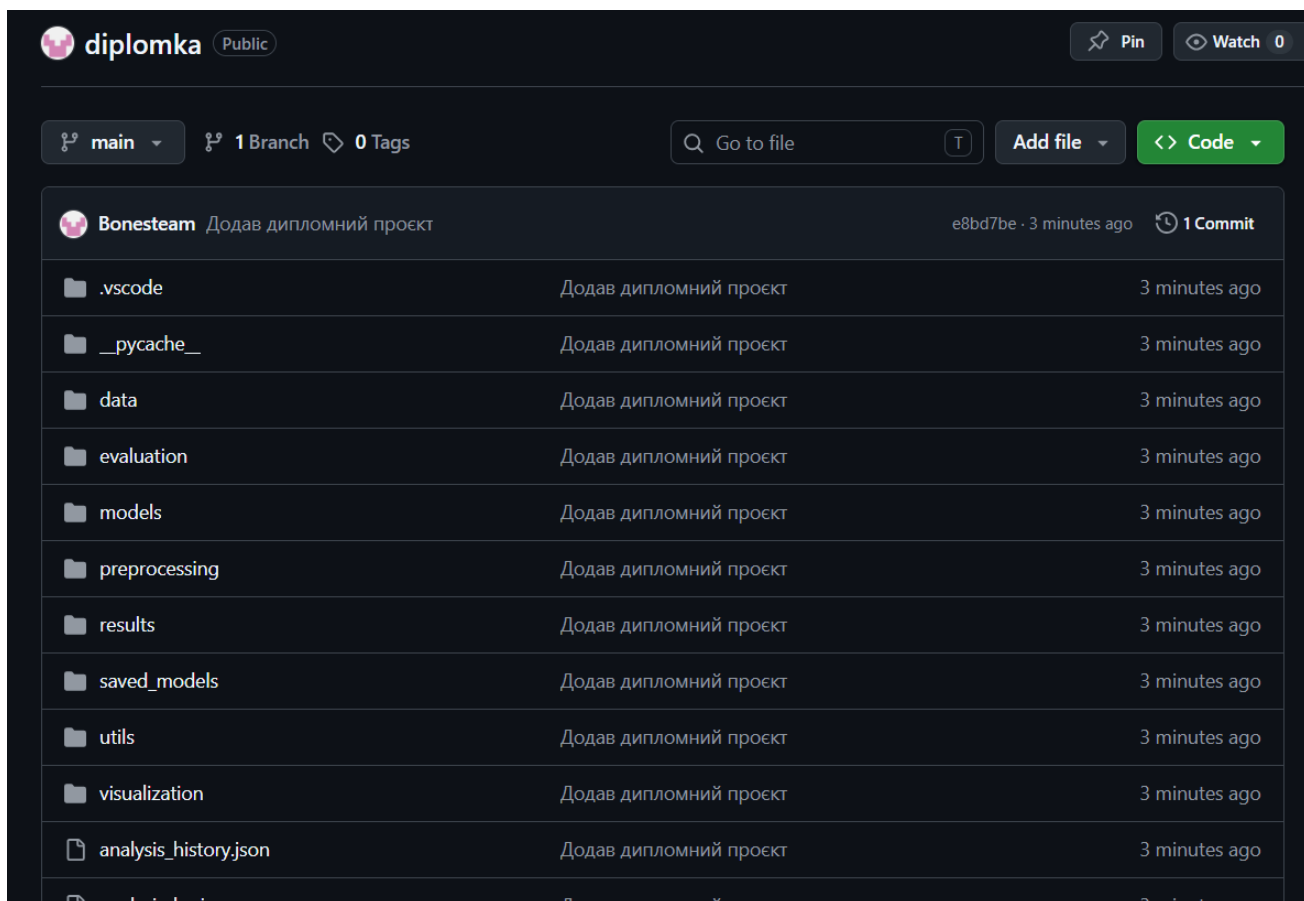
42. Plant Health Biosensor Dataset. Kaggle : електронні дані. URL: <https://www.kaggle.com/datasets/programmer3/plant-health-biosensor-dataset> (дата звернення: 27.02.2026).

ДОДАТКИ

Додаток А

Програмні коди

Вхідний код, використаний у дослідженні, доступний у відкритому репозиторії GitHub: <https://github.com/Bonesteam/diplomka.git>



Структура репозиторію

Основні модулі системи

Основні модулі системи

– `app.py` – головний файл застосунку, який відповідає за запуск інструменту та інтерактивну роботу з моделлю і користувачем.

– `analysis_logic.py` – модуль гібридної класифікації: поєднує рішення MLP з профільною оцінкою за нормальними діапазонами показників здорової рослини.

– `config.yaml` – конфігураційний файл з параметрами навчання, шляхами до даних та налаштуваннями експериментів.

- requirements.txt – перелік залежностей для відтворення середовища виконання.

- run_app.bat – скрипт для запуску застосунку в середовищі Windows.

- split_dataset.py – допоміжний скрипт для стратифікованого розбиття датасету на навчальну та тестову вибірку.

- test_train.py – скрипт для швидкого тестового навчання (3 епохи) з метою перевірки коректності середовища.

Модулі навчання та моделі

- models/mlp_model.py – реалізація MLP-архітектури нейронної мережі: три приховані шари (128→64→32), BatchNormalization, Dropout, L2-регуляризація.

- models/cnn_model.py – реалізація 1D CNN-архітектури для порівняльного аналізу з MLP.

- models/baseline_ml.py – класичні ML-методи (RandomForest, SVM, GradientBoosting) для базового порівняння з нейромережевими підходами.

- models/trainer.py – модуль, що реалізує цикл навчання, збереження контрольних точок і найкращої моделі.

- models/predictor.py – оболонка над навченою моделлю для виконання прогнозів; інкапсулює завантаження моделі та формування передбачень.

- saved_models/ – директорія зі збереженими моделями: mlp_best.keras, randomforest.pkl, gradientboosting.pkl, svm.pkl, scaler.pkl.

Модулі обробки даних

- preprocessing/loader.py – завантаження CSV-датасету, базова валідація та формування числових масивів.

- preprocessing/splitter.py – стратифіковане розбиття даних на train/test вибірки, реалізація крос-валідації.

- preprocessing/scaler.py – Z-нормалізація числових ознак із збереженням параметрів для подальшого використання.

- preprocessing/augmentor.py – SMOTE-аугментація навчальної вибірки для компенсації дисбалансу класів.

- `plant_health_biosensor_dataset.csv` – датасет біосенсорних вимірювань (1253 зразки, 8 ознак, 4 класи стану рослини).

Модулі оцінювання та експериментів

- `evaluation/metrics.py` – обчислення метрик якості класифікації: Accuracy, Precision, Recall, F1-score та ROC-AUC.

- `evaluation/cross_val.py` – реалізація 5-кратної крос-валідації та агрегування результатів по фолдах.

- `evaluation/confusion.py` – побудова та збереження матриць сплутування для аналізу розподілу помилок.

- `evaluation/arch_comparison.py` – порівняння архітектур MLP та 1D CNN за метриками якості та обчислювальними витратами.

- `evaluation/error_analysis.py` – аналіз помилково класифікованих прикладів для діагностики моделі.

- `evaluation/reporter.py` – формування підсумкових звітів із результатами досліджень та метриками.

Модулі візуалізації

- `visualization/plots.py` – функції для побудови графіків навчання, динаміки метрик та гістограм розподілів.

- `visualization/roc_curves.py` – побудова та збереження ROC-кривих для оцінки якості класифікатора.

- `visualization/feature_imp.py` – візуалізація важливості ознак для інтерпретації рішень моделі.

Результати та артефакти

- `results/` – директорія зі збереженими результатами експериментів та графіками (`training_history_mlp.png`, `confusion_matrix.png`, `roc_curves.png` та ін.).

Додаток Б

Презентаційні матеріали

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

МЕТОД НЕЙРОМЕРЕЖЕВОГО ОЦІНЮВАННЯ СТАНУ ЗДОРОВ'Я РОСЛИН ТА ПРОГНОЗУВАННЯ СТРЕСОВИХ СТАНІВ ЗА ДАНИМИ БІОСЕНСОРІВ



Виконав:

студент 4 курсу, група КН-22-3

Костянтин СИДОРЧУК



Керівник:

д.т.н., професор каф. КН

Едуард МАНЗЮК

Мета і задачі кваліфікаційної роботи

Об'єкт дослідження – процеси автоматизованої діагностики стану рослин за числовими показниками біосенсорів.

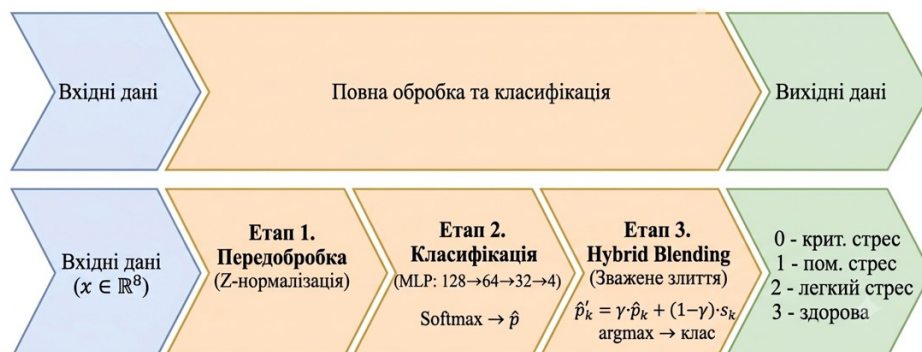
Предмет дослідження – методи та алгоритми нейромережевої класифікації стану рослин за багатовимірними числовими показниками із застосуванням MLP та методів балансування класів.

Мета – підвищення точності та оперативності оцінювання стану здоров'я і прогнозування стресових станів рослин шляхом розробки методу нейромережевої класифікації на основі MLP за числовими показниками біосенсорів.

Для досягнення мети необхідно було виконати такі задачі:

- провести аналіз предметної області методів оцінювання стану рослин за даними біосенсорів;
- розробити метод нейромережевої класифікації стану рослин на основі архітектури MLP;
- реалізувати програмну систему мовою Python із графічним інтерфейсом для аналізу біосенсорних даних;
- провести експериментальне дослідження ефективності розробленого методу та порівняльний аналіз з іншими підходами.

Схема методу неймережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів



Структурна схема MLP-класифікатора стану рослин



Блок-схема алгоритму навчання MLP-класифікатора

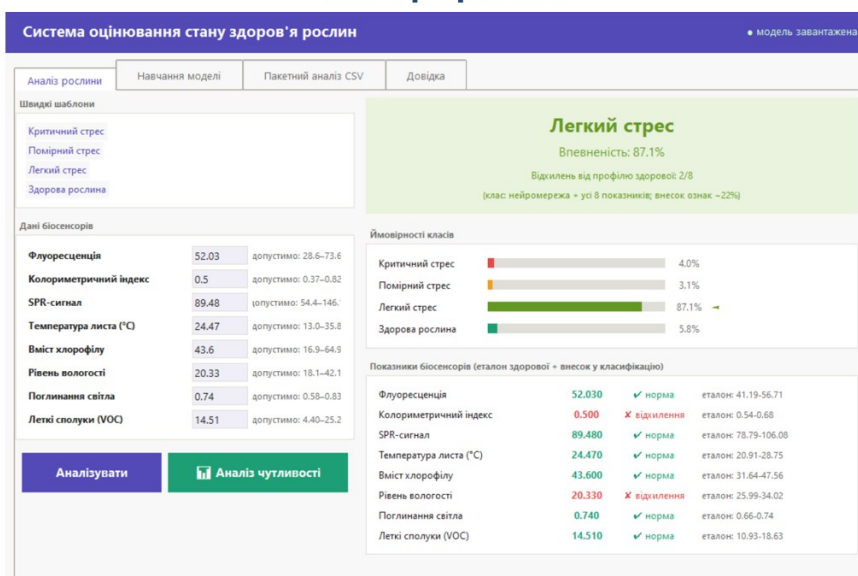


Засоби розробки інтелектуальної системи

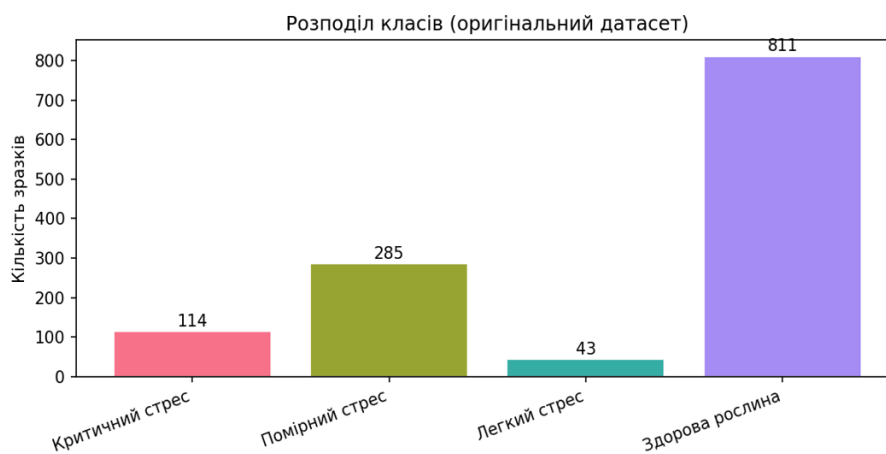
Для програмної реалізації інтелектуальної системи використовувались такі засоби розробки, як:

- Мова програмування Python
- Фреймворк глибокого навчання TensorFlow/Keras
- Бібліотека машинного навчання scikit-learn (SMOTE, StandardScaler, метрики)
- Бібліотека для роботи з даними pandas / numpy
- Бібліотека для візуалізації matplotlib / seaborn
- Графічний інтерфейс tkinter
- Редактор програмного коду Visual Studio Code

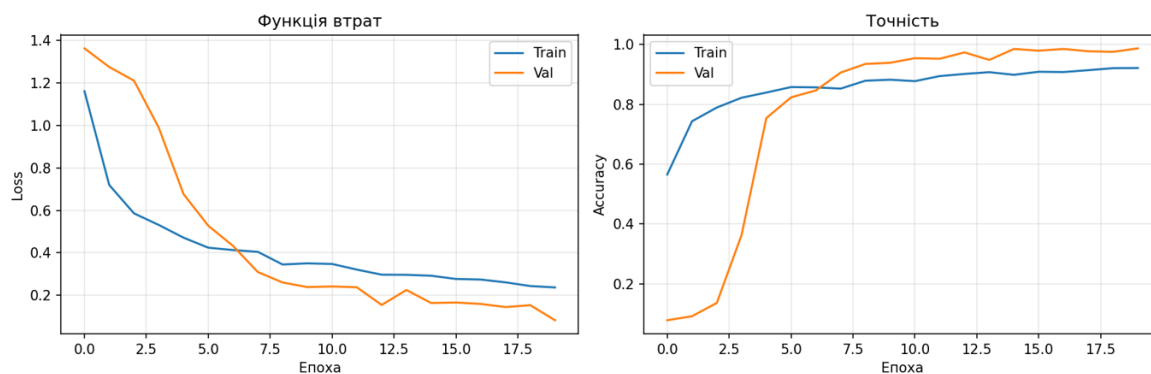
Інтерфейс



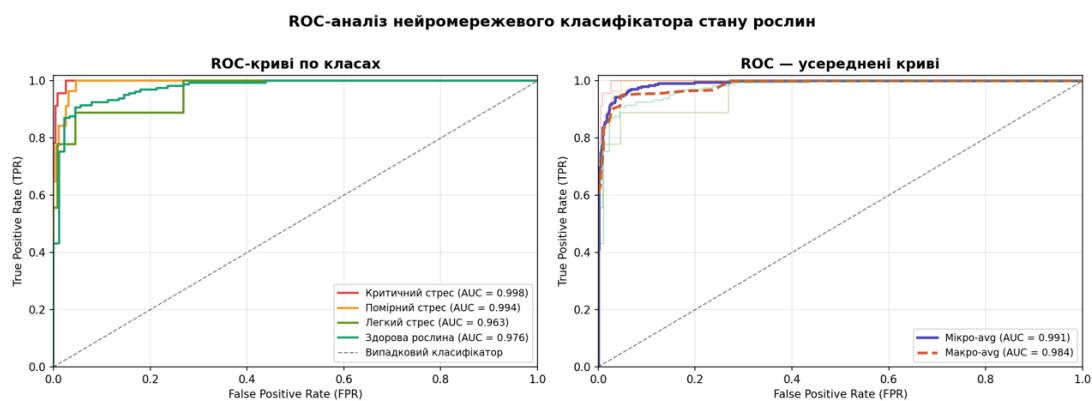
Навчальний набір даних - Plant Health Biosensor Dataset



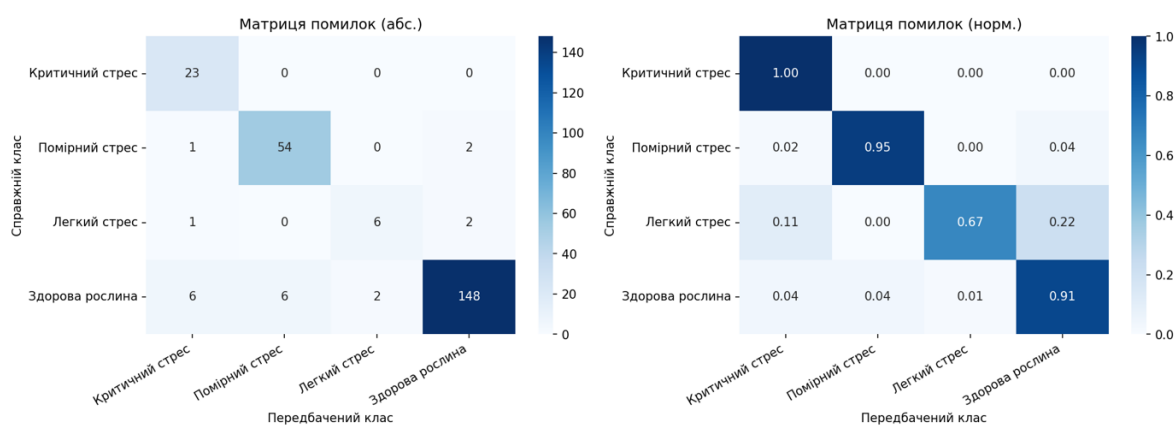
Результат навчання



Результат навчання

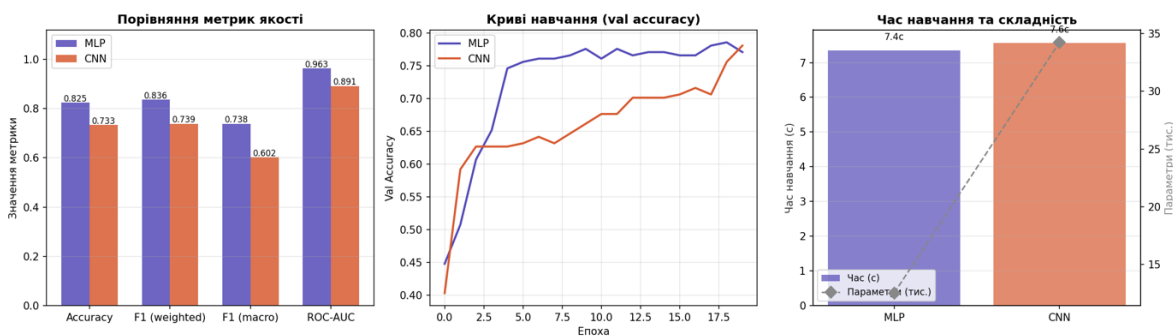


Результат навчання



Порівняння нейромережевих мереж

Порівняльний аналіз архітектур нейронних мереж



Висновки

Метою кваліфікаційної роботи бакалавра було підвищення точності та оперативності оцінювання стану здоров'я і прогнозування стресових станів рослин шляхом розробки методу нейромережевої класифікації на основі MLP за числовими показниками біосенсорів.

Для досягнення поставленої мети були виконані такі задачі:

- проведено аналіз предметної області методів оцінювання стану рослин за даними біосенсорів;
- розроблено метод нейромережевої класифікації стану рослин на основі архітектури MLP;
- реалізовано програмну систему мовою Python із графічним інтерфейсом для аналізу біосенсорних даних;
- проведено експериментальне дослідження ефективності розробленого методу та порівняльний аналіз з іншими підходами.

Розроблена система класифікує стан рослини за вісьмома біосенсорними показниками, визначаючи один із чотирьох класів: критичний стрес, помірний стрес, легкий стрес або здорова рослина.

Результати дослідження підтвердили ефективність розробленого методу: досягнуто Accuracy = 0,920, F1-weighted = 0,921 та ROC-AUC = 0,981. MLP перевершує архітектуру 1D CNN за всіма основними метриками при меншій кількості параметрів.

Перспективи подальшого розвитку включають розширення набору вхідних показників, збільшення навчальної вибірки за рахунок реальних польових даних та адаптацію системи для роботи в автоматизованих агрокомплексах у режимі реального часу.

ДЯКУЮ ЗА УВАГУ!

Sun Jun 14 21:17:36 EEST 2026, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism (http://ap.km.ua) v-16.718

Максимальне співпадіння з одним документом 4.0%**Словники перевірки: UA, US, RU. Помилки в документах: 15%**

ID: 275107 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод нейромережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів Додано в БД: 2026-06-14 Автора: Костянтин СИДОРЧУК Керівники: Едуард МАНЗІЮК Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	66010	591	4185 (6%)	59 (10%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Костянтин СИДОРЧУК

Співавтор:

Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод нейромережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів

Науковий керівник: Едуард МАНЗІЮК, д.т.н., доцент

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1: 6.74%

Коефіцієнт подібності 2: 2.47%

Мікропробіли: 0

Заміна букв: 13

Інтервали: 0

Білі знаки: 190

Дата створення звіту: 2026-06-14 18:24:02.0

Після аналізу Звіту подібності констатую наступне:

Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.

Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.

Обґрунтування:

2026-06-14

Дата

експерт *Петровський Р.Р. стн*

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Назва кваліфікаційної роботи Метод нейромережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів

Автор студент групи КН-22-3 Костянтин СИДОРЧУК

Освітня програма Комп'ютерні науки

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 – Комп'ютерні науки

Науковий керівник: д.т.н., проф. каф. комп'ютерних наук Едуард МАНЗЮК

На основі аналізу кваліфікаційної роботи на дотримання вимог академічної доброчесності (у т.ч. відсутності ознак академічного плагіату) з урахуванням результатів перевірки роботи спеціалізованим програмними засобами комісія зробила такий висновок:

№	Висновок	Позначка про відповідність
1	Ознаки академічного плагіату	
1.1	Запозичення, виявлені в роботі, є законними і не є академічним плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних, якщо потрібно). Робота приймається до захисту.	<i>відповідає</i>
1.2	Виявлені запозичення не є академічним плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована.	
1.3	Виявлені запозичення не є академічним плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота може бути допущена до захисту після того як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат.	
1.4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття текстових запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
2	Інші види порушень академічної доброчесності	<i>відсутні</i>

Підтвердження:

Запозичення, виявлені в роботі Костянтина Сидорчука не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, які не мають авторства і містять поширені конструкції та загальновідомі терміни, скорочення. Рівень подібності не перевищує допустимої межі. Таким чином, робота є законною та приймається до захисту.

Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості:

- за системою Anti-Plagiarism: 4%;

- за системою StrikePlagiarism КПІ: 6,74%.

14.06.2026

Завідувач кафедри



Олександр БАРМАК

Гарант освітньої програми



Олександр МАЗУРЕЦЬ

Керівник кваліфікаційної роботи



Едуард МАНЗЮК



ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ МОН УКРАЇНИ



Кафедра комп'ютерних наук

РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента гр. КН-22-3 Сидорчук Костянтин

за темою: Метод нейромережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів

1. Актуальність обраної теми

Тема кваліфікаційної роботи присвячена розробці нейромережевого методу оцінювання стану здоров'я рослин та прогнозування стресових станів за числовими показниками біосенсорів. Тема є актуальною, оскільки прецизійне землеробство потребує переходу від суб'єктивного візуального контролю агрокультур до об'єктивного автоматизованого моніторингу, здатного виявляти стресові стани рослин на ранніх стадіях. Застосування методів нейромережевої класифікації у поєднанні з методами компенсації дисбалансу класів (SMOTE) дозволяє підвищити точність і оперативність діагностики фізіологічного стану культур, що відповідає актуальним завданням розвитку інтелектуальних систем у галузі комп'ютерних наук.

2. Повнота розкриття мети та завдань роботи

Під час виконання кваліфікаційної роботи студент Сидорчук Костянтин повністю розкрив мету дослідження – підвищення точності та оперативності оцінювання стану здоров'я і прогнозування стресових станів рослин шляхом розробки методу нейромережевої класифікації на основі MLP за числовими показниками біосенсорів. Усі поставлені завдання виконано в повному обсязі: проведено аналіз предметної області, розроблено метод нейромережевої класифікації на основі архітектури MLP, реалізовано програмну систему мовою Python із графічним інтерфейсом, а також проведено експериментальне дослідження ефективності методу з порівняльним аналізом проти архітектури 1D CNN. Досягнуті результати (Accuracy = 92,0%, F1-weighted = 92,1%, ROC-AUC = 0,981) підтверджують досягнення поставленої мети.

3. Зміст кожного розділу роботи

У першому розділі проведено аналіз інформаційних моделей біосенсорного моніторингу рослин, розглянуто теоретичні підходи до класифікації стресових станів та здійснено огляд існуючих програмних і наукових рішень (qREAD-Raman, SSTAS, MapS-Wear та інші), на основі чого обґрунтовано мету та задачі дослідження. Другий розділ присвячено розробці методу: формалізовано задачу багатокласової класифікації, описано архітектуру MLP-класифікатора (128→64→32→4) з механізмами Batch Normalization, Dropout та L2-регуляризації, наведено алгоритми класифікації та навчання моделі із застосуванням SMOTE-аугментації, а також охарактеризовано використаний набір даних Plant Health Biosensor Dataset і метрики оцінювання якості. У третьому розділі представлено програмну реалізацію застосунку з графічним інтерфейсом, проведено експериментальні дослідження стабільності навчання, оцінку точності класифікації за матрицею помилок та ROC-кривими, а також порівняльний аналіз із архітектурою 1D CNN і крос-валідацію

результатів. Усі розділи логічно структуровані, послідовно розкривають тему дослідження та супроводжуються необхідними ілюстраціями й таблицями.

4. Оцінка розробленого методу та його практична цінність

Розроблений нейромережевий метод має високу практичну цінність, оскільки дозволяє агрономам оперативно й об'єктивно оцінювати фізіологічний стан рослин за вісьмома числовими показниками біосенсорів без залучення дорогого лабораторного устаткування. Реалізований програмний застосунок із трьома режимами роботи (аналіз одиничного зразка, навчання моделі, пакетний аналіз CSV-файлу) є функціональним та зручним у використанні. Експериментально підтверджено високу ефективність методу: точність класифікації на тестовій вибірці становить 92,0%, при цьому стан критичного стресу розпізнається з повнотою (recall) 100%, що свідчить про надійність системи для практичного застосування у задачах моніторингу агрокультур.

5. Якість оформлення кваліфікаційної роботи бакалавра

Кваліфікаційна робота відповідає всім вимогам оформлення, чітко розкриває тему та мету дослідження. Робота містить необхідні структурні елементи: анотацію, перелік скорочень, вступ, три розділи з висновками, загальні висновки та перелік посилань (42 джерела), а також достатню кількість ілюстрацій (26) і таблиць (5). Структура логічна, викладення матеріалу послідовне, аргументоване та літературно грамотне.

6. Недоліки кваліфікаційної роботи бакалавра

У роботі студента Сидорчука Костянтина виявлено невелике зауваження: дослідження виконано на одному датасеті (Plant Health Biosensor Dataset), у якому клас "Легкий стрес" представлений лише 43 зразками, тому додаткова перевірка розробленого методу на реальних польових даних або на іншому датасеті була б корисною для більш переконливого підтвердження узагальнювальної здатності моделі. Проте наведене не впливає на позитивну оцінку роботи.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Враховуючи високий рівень виконання та відповідність усім вимогам, робота допускається до захисту. Рекомендована оцінка «відмінно».

Рецензент

Говоруховська Т.О.



**ВІДГУК НАУКОВОГО КЕРІВНИКА
на кваліфікаційну роботу бакалавра**

студента гр. КН-22-3 Костянтина СИДОРЧУКА

за темою Метод нейромережевого оцінювання стану здоров'я рослин та прогнозування стресових станів за даними біосенсорів

1. Актуальність теми

Тематика кваліфікаційного дослідження має високий ступінь актуальності у контексті розвитку цифрового землеробства та систем Smart Farming. Своєчасна діагностика негативних станів флори на основі прямих фізіологічних сигналів дозволяє мінімізувати втрати врожайності. Створення нових інтелектуальних конвеєрів, здатних ефективно обробляти специфічні, сильно незбалансовані біосенсорні дані, є важливим завданням для сучасної агроінженерії та комп'ютерних наук.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

Зміст та спрямованість виконаної роботи цілком задовольняють вимогам освітнього стандарту спеціальності 122 «Комп'ютерні науки». Дослідження К. Сидорчука безпосередньо фокусується на математичній формалізації задач багатокласової класифікації, проектуванні архітектур штучних нейронних мереж та інтелектуальному аналізі часових сигналів біосистем.

3. Професійні та особистісні якості бакалавра

У процесі підготовки випускної роботи Костянтин Сидорчук продемонстрував себе як цілеспрямований, ерудований та технічно грамотний фахівець. Йому притаманні висока виконавська дисципліна, аналітичне мислення, вміння працювати з науковою періодикою та здатність самостійно знаходити оптимальні рішення для складних програмно-апаратних задач.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Усі ключові наукові та практичні результати, представлені в роботі, отримані студентом особисто. Він самостійно реалізував етап попередньої обробки та збалансування даних, обґрунтував конфігурацію нейромережевої моделі, розробив аналітичний блок гібридного злиття ймовірностей.

5. Ступінь оволодіння методами дослідження

Студент виявив глибокі знання сучасного методологічного інструментарію штучного інтелекту та комп'ютерних наук. Під час виконання завдань ним були коректно застосовані методи аугментації табличних ознак, теорія оптимізації глибоких мереж, статистичні критерії оцінювання якості класифікації на незбалансованих вибірках.

6. Повнота та якість розкриття теми роботи

Тему роботи розкрито повністю. Пояснювальна записка містить системний огляд наявних комерційних та академічних рішень, строгий математичний опис запропонованого нейромережевого методу, а також розроблений програмний комплекс із функціональним інтерфейсом користувача.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Пояснювальна записка написана грамотною академічною мовою з чітким дотриманням чинних вимог ДСТУ щодо оформлення текстової документації. Кваліфікаційна робота викладено у послідовній, логічній формі; структуризація розділів відповідає канонам наукового дослідження. Пояснення та висновки мають належне обґрунтування.

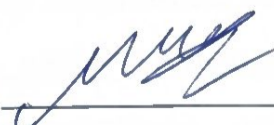
8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Розроблений нейромережевий метод та програмний інструментарій мають практичну цінність. Вони можуть бути інтегровані в автоматизовані системи управління мікрокліматом промислових теплиць, використані при розробці портативних діагностичних пристроїв для агрономів, або впроваджені як аналітичні модулі у хмари моніторингу стану рослинних культур у реальному часі.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота

Враховуючи актуальність теми, хороші показники точності моделі та повну відповідність усім критеріям якості, кваліфікаційна робота Костянтина Сидорчука допускається до захисту перед екзаменаційною комісією. Робота заслуговує на оцінку «добре».

Керівник



д.т.н., проф. каф. КН Едуард МАНЗЮК