

UNIVERSITATEA DE STAT DIN MOLDOVA
ȘCOALA DOCTORALĂ ȘTIINȚE ALE NATURII

Cu titlu de manuscris
C.Z.U.:

ION GANEA

**MODELE, TEHNICI ȘI PRODUSE PROGRAM
DE ANALIZĂ INTELIGENTĂ A DATELOR
ÎN FIZIOLOGIA PLANTELOR**

121.03 Programarea calculatoarelor

Teză de doctor în informatică

Autor:

Ion Ganea

Conducător științific:

Gheorghe Căpățână,
doctor în științe tehnice, prof. univ.

Comisia de îndrumare:

Ana Bîrsan, doctor în biologie,
conf. univ.

Valeriu Ungureanu, doctor în științe
fizico-matematice, conf. univ.

Maria Beldiga, doctor în
informatică, conf. univ.

CHIȘINĂU, 2023

©Ion Ganea, 2023

CUPRINS

ADNOTARE	5
LISTA FIGURILOR	8
LISTA TABELELOR	10
LISTA ABREVIERILOR	11
INTRODUCERE	12
1. PROCESAREA INTELIGENTĂ A DATELOR ÎN BIOLOGIE ȘI AGRICULTURĂ	23
1.1. Fundamentele Tehnologiilor Inteligente	23
1.2. Integrarea Datelor în Cercetarea Avansată	31
1.3. Infrastructuri de Date și Cercetări Avansate	39
1.4. Principalele lacune în implementarea tehnologiilor inteligente	47
1.5. Concluzii la capitolul 1	49
2. INTEGRAREA METODELOR BIOSTATISTICII ȘI A DEPOZITELOR DE DATE ÎN PROCESAREA DATELOR DIN DOMENIUL FIZIOLOGIEI PLANTELOR	51
2.1. Dezvoltarea bazei de cunoștințe	51
2.2. Modelarea datelor	54
2.3. Modelarea biostatistică a datelor din domeniul fiziologiei plantelor	60
2.4. Modelarea datelor experimentale	66
2.5. Depozit de date în agricultură	69
2.6. Concluzii la capitolul 2	84
3. APLICAREA INTELIGENTĂ A METODELOR BIOSTATISTICII ÎN CERCETĂRI DE FIZIOLOGIE A PLANTELOR	87
3.1. Model de analiză biostatistică a datelor privind conținutul de azot, conținutul de proteină, conținutul de fosfor și pentaoxid de fosfor în boabele de soia	87
3.2. Model de analiză statistică a influenței unor substanțe biologice active asupra parametrilor de creștere și dezvoltare a plantelor de soia cultivate pe medii nutritive apoase Knop	95
3.3. Model de analiză biostatistică a datelor referitor la scurgerea electroliților în plantele de soia	104
3.4. Model de analiză biostatistică a termotoleranței la plantele de soia	109
3.5. Concluzii la capitolul 3	111
4. DEPOZITUL DE DATE GRAF: EVIDENȚIEREA CONEXIUNILOR PENTRU OPTIMIZAREA PROCESELOR AGRICOLE	114

4.1. Grafurile pentru procesarea datelor experimentale și a datelor meteorologice	114
Definirea nodurilor și relațiilor în <i>Neo4j</i> pentru datele menționate	114
4.2. Predicția fenomenelor în baza datelor stocate în depozitul de date	118
4.3. Dezvoltarea modelului de predicție	121
4.4. Model de predicție a productivității culturii de soia	138
4.5. Concluzii la capitolul 4.	140
CONCLUZII GENERALE ȘI RECOMANDĂRI	143
BIBLIOGRAFIE	145
CURRICULUM VITAE	164

ADNOTARE

Ion GANEA „Modele, tehnici și produse program de analiză inteligentă a datelor în fiziologia plantelor”. Teza de doctor în informatică, Chișinău, 2023.

Structura tezei: Teza este scrisă în limba română și constă din Introducere, patru capitole, concluzii și recomandări. Bibliografia cuprinde 221 de titluri. Lucrarea conține 132 de pagini de text de bază, 37 de figuri, 17 tabele și 32 de anexe. Rezultatele obținute au fost publicate în 16 lucrări științifice cu un volum de peste 5 coli de autor.

Cuvinte cheie: tehnologii inteligente, fiziologia plantelor, omice, biostatistică, bază de cunoștințe, baze de date graf, depozit de date, modele, limbaj formal, familii de probleme.

Scopul tezei de doctorat este dezvoltarea și furnizarea de asistență cercetătorilor biologi pentru a facilita procesarea informației și identificarea rezultatelor semnificative obținute în cadrul cercetărilor biologice complexe pentru optimizarea procesului de descoperire a cunoștințelor relevante și de dobândire a înțelepciunii în biologie și agricultură.

Obiective: Identificarea informațiilor cheie în cercetările biologice pentru interpretarea datelor; Dezvoltarea și validarea unor modele de analiză inteligentă a datelor; Procesarea datelor pentru cuantificarea efectelor tratamentelor asupra plantelor; Utilizarea tehnologii avansate pentru optimizarea descoperirii rezultatelor relevante; Crearea Depozitului de date eficient pentru cercetările biologice și agricole.

Noutatea și originalitatea științifică: model generic în limbaj formal al domeniului „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor” (MBDFP), o bază de cunoștințe pentru acest domeniu, modele de analiză inteligentă a datelor, depozit de date.

Principala problemă științifică rezolvată: prin modelele dezvoltate și baza de cunoștințe a domeniului de cercetare, cercetătorii biologi sunt asistați în selecția datelor relevante și obținerea de rezultate de înaltă calitate în domeniul fiziologiei plantelor.

Semnificația teoretică: limbajul formal pentru procesarea datelor și baza de cunoștințe furnizează informații structurate, facilitând cercetările, iar modelele de analiză adâncesc înțelegerea datelor.

Valoarea aplicativă: Modelul de analiză inteligentă a datelor aplicat pe depozitul de date prezintă rezultatele în mod intuitiv, prin grafuri ușor de înțeles, optimizând astfel performanța sistemelor biologice și consolidarea rezilienței sistemelor agricole. Printr-o soluție inovativă bazată pe un depozit de date graf, se optimizează performanțele sistemelor biologice și agricole.

Implementarea rezultatelor științifice: Modelele elaborate au fost premiate cu: două medalii de aur, o medalie de argint, una de bronz și diplomă de excelență la saloane internaționale de tehnologie în anii 2021 - 2023. La AGEPI au fost înregistrate două drepturi de autor.

ANNOTATION

Ion GANEA „Models, Techniques, and Software Products for Intelligent Data Analysis in Plant Physiology”. Doctoral Thesis in Computer Science, Chişinău, 2023.

Thesis Structure: The thesis is written in Romanian and consists of an Introduction, four chapters, conclusions, and recommendations. The bibliography includes 221 titles. The work contains 132 pages of main text, 37 figures, 17 tables, and 32 appendices. The obtained results have been published in 16 scientific papers totaling over 5 author sheets.

Keywords: intelligent technologies, plant physiology, omics, biostatistics, knowledge base, graph databases, data repositories, models, problem families.

Research Purpose: Develop and provide assistance to biologists in order to facilitate data processing and identify significant results obtained in complex biological research, optimizing the knowledge discovery process and acquiring wisdom in biology and agriculture.

General Research Objectives: Identifying key information in biological research for data interpretation; Development and validation of intelligent data analysis models; Data processing for quantifying the effects of treatments on plants; Utilizing advanced technologies to optimize the discovery of relevant results; Establishing an efficient Data Repository for biological and agricultural research; Evaluating factors influencing agricultural efficiency, considering complex interactions.

Scientific Novelty and Originality: A generic model in the formal language of the field "Biostatistical Data Modeling in Plant Physiology" (MBDFP), a knowledge base for this domain, and intelligent data analysis models.

Main Scientific Problem Solved: Through the developed model and knowledge base, biological researchers are assisted in selecting relevant data and obtaining high-quality results in the field of plant physiology.

Theoretical Significance: The knowledge base provides structured information, facilitating research, and the analysis models deepen the understanding of data.

Applicative Value: The applied intelligent data analysis model on this repository presents results intuitively, through easily understandable graphs, thus optimizing the performance of biological and agricultural systems and strengthening the resilience of agricultural systems.

Implementation of Scientific Results: The models have been awarded two gold medals, one silver medal, one bronze medal, and a certificate of excellence at international technology exhibitions in the years 2022, 2023. Two copyrights have been registered with AGEPI.

АННОТАЦИЯ

Ион ГАНЯ „Модели, техники и программные продукты для интеллектуального анализа данных в физиологии растений”.

Докторская диссертация по информатике, Кишинэу, 2023.

Структура диссертации: Диссертация написана на румынском языке и состоит из Введения, четырех глав, заключения и рекомендаций. Библиография включает 221 названий. Работа содержит 132 страниц основного текста, 37 иллюстраций, 17 таблиц и 32 приложений. Полученные результаты были опубликованы в 16 научных статьях, общим объемом более 5 авторских листов.

Ключевые слова: интеллектуальные технологии, физиология, биостатистика, база знаний, графовые базы данных, хранилища данных, модели, семьи проблем.

Цель исследования: Разработка и помощь биологическим исследователям с интеллектуальными программными продуктами для обнаружения знаний и результатов в области биологии и сельском хозяйстве на основе математического анализа большого объема данных.

Научные задачи: оптимизация процесса обнаружения результатов, оценка факторов риска и их воздействия на сельскохозяйственную эффективность, разработка и валидация моделей интеллектуального анализа данных, создание хранилища данных.

Научная новизна и оригинальность: Обобщенная модель на формальном языке области "Биостатистическое Моделирование Данных в Физиологии Растений" (MBDFP), база знаний для данной области и модели интеллектуального анализа данных. Основная

Решенная научная проблема: С помощью разработанной модели и базы знаний биологические исследователи получают помощь в выборе соответствующих данных и получении высококачественных результатов в области физиологии растений.

Теоретическое значение: База знаний предоставляет структурированную информацию, облегчая исследования, а модели анализа углубляют понимание данных.

Прикладная ценность: Примененная модель интеллектуального анализа данных на хранилище данных представляет результаты интуитивно, через легко понимаемые графы, тем самым оптимизируя производительность биологических систем.

Внедрение научных результатов: Разработанные модели были удостоены двух золотых медалей, одной серебряной медали, одной бронзовой медали и Диплом отличия на международных выставках технологий в 2022 и 2023 годах. Две авторские права были зарегистрированы в AGEPI.

LISTA FIGURILOR

Figura 1.1 . Resurse de date și instrumente EMBL-EBI	42
Figura 2.1 . Structura plantei	62
Figura 2.2 . Noduri de tip Soiuri	77
Figura 2.3 . Nodurile de tip Coeficientul hidrotermic	78
Figura 3.1 . Conținutul de azot din boabele de soia (<i>soiul Clavera</i> , supus tratamentului foliar cu <i>Reglalg</i> și <i>Biovit</i>), (%)	87
Figura 3.2 . Conținutul de proteină din boabele de soia (<i>soiul Clavera</i> , supus tratamentului foliar cu <i>Reglalg</i> și <i>Biovit</i>), (%)	89
Figura 3.3 . Conținutul de fosfor (P_2) din boabele de soia (<i>soiul Clavera</i> , supus tratamentului foliar cu <i>Reglalg</i> și <i>Biovit</i>), (mg/kg)	91
Figura 3.4 . Conținutul de pentaoxid de fosfor (P_2O_5) din boabele de soia	93
Figura 3.5 . Lungimea rădăcinii principale a plantelor cultivate pe medii apoase suplinite cu substanțe biologice active	96
Figura 3.6 . Volumul rădăcinii plantelor cultivate	97
Figura 3.7 . Lungimea tulpinii plantelor cultivate	98
Figura 3.8 . Biomasa proaspătă a rădăcinii plantelor cultivate pe medii apoase suplinite cu substanțe biologice active	99
Figura 3.9 . Biomasa uscată a rădăcinii plantelor cultivate	100
Figura 3.10 . Biomasa proaspătă a tulpinii plantelor cultivate	102
Figura 3.11 . Biomasa uscată a tulpinii plantelor cultivate	103
Figura 3.12 . Scurgerea electroliților la grupul <i>Martor</i>	106
Figura 3.13 . Scurgerea electroliților la grupul tratat cu <i>Humat</i>	106
Figura 3.14 . Lungimea rădăcinii	109
Figura 4.1 . Graful rezultat (fragment)	118
Figura 4.2 . Proiecția grafului	124
Figura 4.3 . Determinarea nodurilor similare	125
Figura 4.4 . Anii cu <i>htc</i> similar	126

Figura 4.5 . Nodurile similare	127
Figura 4.6 . Incorporarea nodurilor	127
Figura 4.7 . Proprietățile înscrise în noduri	128
Figura 4.8 . Crearea conductei	129
Figura 4.9 . Adăugarea unui nod de proprietate	130
Figura 4.10 . Instrucțiunea și rezultatul adăugării caracteristicilor	131
Figura 4.11 . Divizarea datelor	132
Figura 4.12 . Adăugarea unui model de regresie logistică	133
Figura 4.13 . Adăugarea unui model de pădure aleatoriu.	134
Figura 4.14 . Model de perceptron multistrat cu pierdere focală	135
Figura 4.15 . Reglarea automată a parametrilor	136
Figura 4.16 . Rezultatul instruirii conductei	137
Figura 4.17 . Rezultatul testării predicției	138
Figura 4.18 . Instruirea modelului.	140
Figura 4.19 . Rezultatul predicției productivității.	140

LISTA TABELELOR

Tabelul 3.1 . Variația conținutului de azot (N_2)	88
Tabelul 3.2 . Rezultatele influenței tratamentului foliar asupra conținutului de azot din boabele de soia (<i>soiul Clavera</i>)	88
Tabelul 3.3 . Variația conținutului de proteine în boabele de soia (<i>soiul Clavera</i>)	90
Tabelul 3.4 . Rezultatele influenței tratamentului asupra conținutului de proteină în boabele de soia (<i>soiul Clavera</i>)	90
Tabelul 3.5 . Variația conținutului de fosfor (P_2) din boabele de soia (<i>soiul Clavera</i>)	92
Tabelul 3.6 . Rezultatele influenței tratamentului asupra conținutului de fosfor (P_2) din boabele de soia (<i>soiul Clavera</i>)	92
Tabelul 3.7 . Variația conținutului de pentoxid de fosfor (P_2O_5) în boabele de soia (<i>soiul Clavera</i>)	93
Tabelul 3.8 . Rezultatele influenței tratamentului asupra	94
Tabelul 3.9 . Rezultatele influenței compușilor biologic activi asupra lungimii rădăcinii	97
Tabelul 3.10 . Rezultatele influenței compușilor biologic activi asupra lungimii tulpinii	99
Tabelul 3.11 . Rezultatele influenței compușilor biologic activi asupra biomasei proaspete a rădăcinii	100
Tabelul 3.12 . Rezultatele influenței compușilor biologic activi asupra biomasei uscate a rădăcinii	101
Tabelul 3.13 . Influența tratamentului asupra biomasei proaspete a tulpinii	102
Tabelul 3.14 . Influența tratamentului asupra biomasei uscate a tulpinii	104
Tabelul 3.15 . Rezultatul scurgerii electroliților	107
Tabelul 3.16 . Variația lungimii rădăcinii	110
Tabelul 3.17 . Influența tratării termice asupra rădăcinii	111

LISTA ABREVIERILOR

AGEPI - Agenția de Stat pentru Proprietatea Intelectuală
AIns – Artificial Insight (Perspicacitate Artificială)
APOC - Awesome Procedures On Cypher
AW – Artificial Wisdom (Înțelepciune artificială)
BAC - Cromozomul artificial bacterian
BLAST - Basic Local Alignment Search Tool
CHT – Coeficientul hidrotermic
CNN - Convolutional Neural Networks (Rețele Neuronale Convoluționale)
DC – Domeniu de cunoaștere
DL - Deep Learning (Învățare Profundă)
EMBL - Laboratorul European de Biologie Moleculară
EST - Etichetele de secvență exprimată
FP – Familii de probleme
GDS – Graph Data System
GISD - Baza de date globală pentru specii invazive
GRIIS - Registrul Global al speciilor introduse și invazive
IA – Inteligență artificială
IABIN - Rețeaua Interamericană De Informații Despre Biodiversitate
IBIS - Baza de date pentru biodiversitate și speciile invazive
IC - Inteligența Computațională
IoT – Internet of Things (Internetul lucrurilor)
KB - Knowledge Base (Bază de cunoștințe)
KDD – Knowledge Discovery
MBDFP – Modelare Biostatstică a Datelor de Fiziologie a Plantelor
ML – Machine Learning (Învățare Automată)
OFP – Orientat pe familii de probleme
PDB - Protein Data Bank
SGBD – Sistem de Gestiune a Bazelor de Date
SSD – Sisteme suport Decizii
TI – Tehnologia informației
TL – Transfer Learning (Învățare prin Transfer)
WEKA - Waikato Environment for Knowledge Analysis

INTRODUCERE

Actualitatea și importanța temei

Actualmente, omenirea se confruntă cu probleme globale complexe ce pun în pericol dezvoltarea și existența sa. În zonele sărace creșterea rapidă a populației generează lipsă de resurse și impact asupra mediului. În țările dezvoltate, îmbătrânirea populației și migrația au propriile provocări.

Problemele globale derivă din activități umane nesustenabile, epuizarea resurselor, impactul negativ al producției industriale, conflicte pentru resurse, globalizare și inegalități. Exploatarea irațională a resurselor generează poluarea și degradarea mediului, afectând sistemele biologice (plante, animale, om), inclusiv prin schimbările climatice.

În același timp, dezvoltarea societății umane vizează satisfacerea durabilă a nevoilor umane fundamentale [1]. Dezideratele esențiale ale omului, ce trebuie îndeplinite pentru evoluția personalității sunt asigurarea sănătății și fericirii individuale și dezvoltarea durabilă [2]. Dezvoltarea durabilă a societății umane depinde de satisfacerea nevoilor umane prin intermediul tehnologiilor și utilizarea eficientă a resurselor disponibile, fie naturale fie produse de om [3]. Aceste probleme necesită noi concepte pentru dezvoltarea economică durabilă, evaluând impactul asupra mediului. Schimbările climatice necesită culturi agricole adaptabile, sprijinite de cercetările transdisciplinare și tehnologii inteligente, astfel încât protejarea naturii să coexiste cu dezvoltarea. Tehnologiile reprezintă metode esențiale pentru obținerea produselor și îndeplinirea cerințelor umane [4-5].

Analiza inteligentă a datelor în domeniul fiziologiei plantelor reprezintă o zonă de cercetare în creștere care reprezintă o combinație între domeniul biologiei și cel al informaticii. Acest subiect îmbină cercetarea în domeniul științelor plantelor cu abordări moderne de analiză și prelucrare a datelor, utilizând tehnici și algoritmi de inteligență artificială (IA) și învățare automată (ML) pentru a extrage informații relevante despre homeostazia plantelor (procesul fiziologic prin care organismele mențin un mediu intern stabil (temperatura, pH-ul și nivelurile de nutrienți) în ciuda modificărilor externe, creșterea și dezvoltarea acestora și răspunsul lor la acțiunea factori nefavorabili. Aceasta poate aduce avantaje semnificative în cercetarea și managementul plantelor, a agroecosistemelor, dar și a mediului în ansamblu. În ultimii ani, se atestă o creștere semnificativă a interesului pentru aplicarea tehnologiilor informaționale în studiul plantelor și a proceselor lor fiziologice. În domeniul fiziologiei plantelor, au fost dezvoltate tehnologii avansate pentru a colecta date cu referire la compoziția biochimică și genetică a plantelor, procese fiziologice de creștere, starea de sănătate în dependență/corelație cu

mediul de existență etc. Senzorii IoT (*Internet of Things* - *Internetul lucrurilor*) și alte dispozitive pot furniza o cantitate mare de date în timp real:

1. *Internetul lucrurilor (IoT) în agricultură* oferă oportunități de integrare a dispozitivelor IoT în domeniul agricol și permite colectarea automată a datelor, comunicarea în timp real cu plantele, ajustarea automată a condițiilor de creștere, pentru a maximiza producția și sănătatea plantelor. Integrarea senzorilor în plante pentru a măsura parametrii fiziologici în dependență de umiditatea solului, conținutul de nutrienți, lumină, temperatură și alți parametri, în combinație cu tehnologiile IoT, permite colectarea și transmiterea datelor în timp real către sistemele de analiză. Aceasta contribuie la înțelegerea mai profundă a necesităților fiziologice ale plantelor și a răspunsurilor lor la diferite condiții. Integrarea acestor date într-un sistem informatic permite analiza în timp real a stării plantelor și optimizarea condițiilor de creștere.
2. În *Agricultura de precizie*, utilizarea datelor colectate prin senzori în timp real din terenurile agricole oferă oportunități uriase pentru monitorizarea și gestionarea eficientă a plantelor. Prin tehnici de analiză inteligentă a acestor date, cum ar fi învățarea automată și analiza big data, fermierii pot lua decizii informate pentru irigație, aplicarea precisă a îngrășămintelor și a pesticidelor, ceea ce duce la sporirea recoltelor și reducerea impactului asupra mediului. Integrarea analizei inteligente a datelor în agricultura de precizie permite agricultorilor să ia decizii mai informate cu privire la momentul optim de plantare, fertilizare, combaterea dăunătorilor etc., ceea ce duce la creșterea randamentului culturilor.
3. *Analiza imaginilor și recunoașterea plantelor* vizează utilizarea algoritmilor de procesare a imaginilor pentru a identifica și urmări dezvoltarea plantelor. Aceasta poate include recunoașterea bolilor, deficiențelor nutriționale și a altor probleme fiziologice prin analiza detaliată a imaginilor. Tehnicile utilizate pot dezvălui informații esențiale despre starea de sănătate a plantelor, precum deficiențele nutriționale, boli sau stresul hidric, contribuind la intervenții rapide și precise.
4. *Big Data și Complexitate*: Colectarea masivă a datelor din fiziologia plantelor poate duce la provocări legate de volumul, variabilitatea și complexitatea datelor. Analiza acestor seturi mari de date necesită tehnici avansate de procesare, cum ar fi învățarea automată și analiza datelor. Cu colectarea continuă a datelor de la diverse experimente și culturi de plante, apare provocarea gestionării și analizării eficiente a volumelor mari de date. Tehnologiile de analiză a datelor masive, cum ar fi tehnici de stocare eficientă, procesare paralelă și metode de reducere a dimensionalității datelor, devin tot mai importante în

acest context. Creșterea volumului de date din fiziologia plantelor a dus la necesitatea dezvoltării unor baze de date centralizate și platforme integrate care să faciliteze stocarea, gestionarea și partajarea datelor între cercetători din întreaga lume. În acest sens sunt depozitele de date și, mai recent, lacurile de date.

5. *Înțelegerea fenomenelor complexe*: Prin aplicarea analizei inteligente a datelor, cercetătorii pot identifica modele și corelații care pot fi dificil de observat cu metode tradiționale—ceea ce ajută la înțelegerea mai profundă a răspunsului plantelor la diverși stimuli determinați de schimbările de mediu, boli, deficiențe nutriționale etc.
6. *Selecția plantelor*: Studiul fiziologiei plantelor are, de asemenea, implicații în dezvoltarea de produse sau substanțe active derivate din plante. Analiza datelor poate ajuta la identificarea plantelor cu compuși valoroși și la optimizarea condițiilor de creștere pentru a maximiza producția acestora.
7. *Modelarea, simulările și predicția comportamentului plantelor*: Dezvoltarea modelelor matematice complexe și algoritmilor de învățare automată pentru simularea proceselor fiziologice ale plantelor poate ajuta la prezicerea comportamentului plantelor în funcție de diferiți factori precum condițiile de mediu, tipurile de sol și nivelurile de nutrienți. Acești factori contribuie la proiectarea unor strategii agricole optime și în evaluarea impactului schimbărilor climatice asupra culturilor. Modelele dezvoltate pot ajuta agricultorii să ia decizii informate legate de gestionarea culturilor.
8. *Învățarea automată și inteligența artificială*: Tehnologiile inteligente pot fi utilizate pentru a analiza seturi mari de date generate de experimente fiziologice, pentru a identifica modele, corelații și tendințe ascunse. De asemenea, pot contribui la dezvoltarea unor algoritmi de diagnostic precoce a bolilor sau a altor probleme de sănătate ale plantelor.
9. *Colaborarea interdisciplinară*: Subiectul implică o colaborare strânsă între informaticieni și experți în inteligență artificială, biologi, agronomi, și alți specialiști în diverse domenii. Comunicarea transdisciplinară și schimbul de cunoștințe sunt esențiale pentru progresul în cercetarea acestui domeniu. Integrarea informaticii și a tehnologiilor inteligente în studiul domeniului fiziologiei plantelor demonstrează un impact semnificativ asupra agriculturii, ecologiei și mediului înconjurător. Cu instrumentele adecvate, se pot obține înțelegeri mai profunde a proceselor fiziologice ale plantelor și dezvoltarea de strategii mai eficiente pentru gestionarea resurselor agricole într-un mod sustenabil.
10. *Optimizarea resurselor agricole*: Utilizarea tehnologiilor de analiză a datelor pentru optimizarea utilizării resurselor precum apă, energie și substanțe chimice în agricultură.

Prin aceste metode, cercetătorii și agricultorii pot contribui la dezvoltarea durabilă a agriculturii, atenuând schimbările climatice și adaptând-o noilor condiții.

Scopul tezei de doctorat este dezvoltarea și furnizarea de asistență cercetătorilor biologi pentru a facilita procesarea informației și identificarea rezultatelor semnificative obținute în cadrul cercetărilor biologice complexe pentru optimizarea procesului de descoperire a cunoștințelor relevante și de dobândire a înțelepciunii în biologie și agricultură.

Obiectivele de cercetare propuse sunt următoarele:

1. Identificarea informațiilor esențiale în contextul cercetărilor biologice pentru facilitarea procesului de selecție și interpretare a datelor relevante.
2. Elaborarea de modele matematice pentru evaluarea influenței anumitor preparate asupra plantelor, cu scopul de a determina efectele acestora în contextul agricol.
3. Procesarea datelor pentru determinarea gradelor de influență ale preparatelor studiate, astfel încât să se poată cuantifica și compara impactul diferitelor tratamente asupra plantelor.
4. Modelarea datelor. Optimizarea procesului de descoperire a rezultatelor relevante prin utilizarea tehnologiilor inteligente și algoritmilor avansați de analiză a datelor și a modelelor inteligente.
5. Dezvoltarea și validarea unor modele de analiză inteligentă a datelor, cu scopul de a oferi un suport semnificativ cercetătorilor în înțelegerea informațiilor complexe din domeniul biologic și agricol.
6. Dezvoltarea produselor program:
 - 6.1. Crearea *Depozitului de Date* care să permită accesul și stocarea eficientă a datelor pentru cercetările biologice și agricole.
 - 6.2. Elaborarea unor modele de procesare a datelor cu accent pe adaptarea plantelor la noile condiții climatice și dezvoltarea agriculturii reziliante la schimbările de mediu.
7. Procesarea algoritmilor pentru evaluarea factorilor care influențează eficiența agriculturii, luând în considerare interacțiunile complexe dintre plante, sol, climă și preparatele utilizate.
8. Evaluarea și formularea de recomandări privind dezvoltarea agriculturii în Republica Moldova, bazate pe rezultatele obținute în cadrul acestor cercetări, în vederea promovării unei agriculturi durabile și competitive în regiune.

Metodologia cercetării

În teză au fost utilizate metode și tehnici avansate pentru cercetarea fiziologiei plantelor și agriculturii, prin combinarea biostatisticii și depozitelor de date graf. Această abordare

interdisciplinară a permis obținerea perspicacității valoroase, informații și cunoștințe relevante în domeniu.

Prin utilizarea biostatisticii, au fost aplicate o serie de analize și teste statistice riguroase pentru a investiga relațiile dintre diverse variabile din fiziologia plantelor și din agricultură. Aceste metode au ajutat la identificarea de modele, corelații și diferențe semnificative în datele colectate, oferindu-ne înțelegerea profundă a fenomenelor studiate.

De asemenea, au fost explorate depozitele de date graf pentru vizualizarea rezultatelor analizelor și pentru a aduce în prim-plan relațiile complexe dintre diferitele variabile. Vizualizările sub formă de grafuri au reprezentat o modalitate eficientă de a comunica rezultatele într-un mod accesibil și ușor de înțeles, facilitând asimilarea informațiilor de către publicul țintă.

A fost elaborat un depozit de date bazat pe grafuri, utilizând SGBD Neo4j. Autorul deține certificare profesională în acest domeniu (anexa 10). În contextul cercetării de fiziologie a plantelor, această metodologie a permis obținerea unei perspective mai clare asupra mecanismelor care dirijează creșterea și dezvoltarea plantelor, precum și asupra interacțiunilor dintre plante și factorii de mediu din agricultură. Prin combinarea biostatisticii și depozitelor de date graf, s-a reușit evidențierea importanței și potențialul acestei abordări holistice în dezvoltarea cunoștințelor în domeniul fiziologiei plantelor și agriculturii, deschizând noi direcții de cercetare și aplicând rezultatele cercetării în îmbunătățirea practicilor agricole și a sustenabilității mediului.

Datele biologice și îndrumările teoretice referitoare la fiziologia plantelor în cadrul cercetărilor efectuate, au fost oferite de către dna conferențiar universitar, dr. Ana Bîrsan, Departamentul Biologie și Ecologie, Facultatea de Biologie și Geoștiințe, Universitatea de Stat din Moldova.

Noutatea și originalitatea științifică: Pentru domeniul de cercetare MBDFP a fost dezvoltat un limbaj formal bazat pe familii de probleme, o bază de cunoștințe, modele de analiză biostatistică a datelor, un depozit de date, modele predictive referitoare la secetă și productivitatea unor culturi agricole. Depozitul de date permite totodată aplicarea modelului de analiză inteligentă a datelor și prezintă rezultatele prin grafuri intuitive, ușor de utilizat și de înțeles. Printr-o soluție inovativă bazată pe un depozit de date graf, se optimizează performanțele sistemelor biologice și agricole prin tehnologii avansate și analize biostatistice. Scopul este de a găsi soluții eficiente pentru provocările din agricultură, precum schimbările climatice, degradarea solurilor și producția sustenabilă. Cu ajutorul IoT, IA și Big Data, cercetătorii și agricultorii colectează și analizează date din surse diverse. Abordarea permite luarea deciziilor mai informate pentru practicile agricole în timp real. Analiza biostatistică ajută la identificarea

corelațiilor între variabilele plantelor, facilitând selecția plantelor reziliente și adaptate la mediu. Astfel, soluțiile tehnologice și analiza biostatistică a datelor biologice optimizează durabilitatea și productivitatea sistemelor biologice și agricole, contribuind la avansarea cunoașterii în agricultura inteligentă și sustenabilă.

Semnificația teoretică: Avansarea cunoașterii și înțelegerii în domeniul tehnologiilor inteligente aplicate în biologie și agricultură. Cercetările în domeniul tehnologiilor inteligente aplicate în biologie și agricultură constă în descoperirea de noi cunoștințe și înțelegerea mai profundă a interacțiunilor complexe dintre factorii biologici și mediul înconjurător. Aceste cercetări contribuie la dezvoltarea și consolidarea fundamentelor teoretice ale fiziologiei plantelor, agriculturii inteligente și analizei biostatistice a datelor biologice, extinzând astfel cunoștințele în științele agricole. Prin intermediul cercetărilor, sunt identificate și validate modelele și mecanismele de acțiune ale tehnologiilor inteligente, în contextul biologiei și agriculturii. Aceasta aduce în fața comunității științifice și a specialiștilor în domeniu informații esențiale despre cum aceste tehnologii pot fi aplicate pentru a optimiza performanțele și a spori sustenabilitatea sistemelor agricole.

Valoarea aplicativă a cercetărilor: Dezvoltarea soluțiilor concrete și inovatoare pentru optimizarea producției agricole, adaptarea la schimbările climatice și sporirea sustenabilității sistemelor agricole. Modelele predictive sunt cruciale în gestionarea riscurilor climatice și adaptarea agriculturii la schimbările climatice în agricultură și îmbunătățirea planificării producției și alocarea resurselor. Informațiile avansate și soluțiile tehnologice cresc productivitatea, reduc impactul asupra mediului și asigură securitatea alimentară pe termen lung. Soluțiile concrete și inovatoare dezvoltate în cercetare pot fi implementate efectiv în practicile agricole, generând beneficii economice și sociale pentru agricultori și comunități. Utilizarea tehnologiilor inteligente facilitează monitorizarea și gestionarea mai eficientă a resurselor naturale, prevenind problemele de mediu și contribuind la un mediu agricol echilibrat și durabil pe termen lung.

Rezultate științifice înaintate spre susținere:

1. A fost studiat domeniul de activitate „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor” (abreviat MBDFP).
2. A fost elaborat un limbaj formal al domeniului de activitate MBDFP, orientat pe familii de probleme. Limbajul MBDFP este de complexitate semnificativă. Aplicând tehnologiile convenționale de programare, limbajul poate fi realizat în cadrul unui sistem de automatizare a programării dezvoltat de către un colectiv de informaticieni.

3. Utilizarea instrumentelor de inteligență artificială, a bazelor de date non-relaționale și a metodologiei rezolvării problemelor orientate pe familii de probleme a permis realizarea limbajului creat de autor în termenul stabilit în Planul cercetării doctorale.
4. Au fost dezvoltate *modele generice (abstracte) și modele specifice orientate pe familii de probleme*;
5. Au fost elaborate 9 modele de analiză biostatistică a influenței unor substanțe biologice active asupra creșterii și dezvoltării unor fenotipuri de soia (*Glycine max* (L.) Merrill.);
6. A fost elaborat un depozit de date graf pentru stocarea și procesarea unor cantități mari de date.
7. Au fost elaborate 2 modele predictive referitoare la probabilitatea secetei și productivitatea culturilor agricole.

Aprobarea rezultatelor tezei:

Rezultatele obținute în urma cercetării au fost prezentate și discutate la 11 conferințe naționale și internaționale:

1. European Exhibition of Creativity and Innovation „EuroInvent”, 15-th edition. 16 -18 Mai 2023, Iași, Romania. **Diplomă de excelență.**
2. International Exhibition „Inventcor” 2022, 15-17 decembrie 2022, Deva, Romania. **Medalie de aur.**
3. International Exhibition „Inventcor” 2022, 16-18 decembrie 2021, Deva, Romania. **Medalie de aur.**
4. International Exhibition of Inventions and Innovations „Traian Vuia” 2022, 8-10 decembrie 2022, Timișoara, Romania. **Medalie de bronz.**
5. European Exhibition of Creativity and Innovation „EuroInvent”, 14-th edition. 26 -28 Mai 2022, Iași, Romania. **Medalie de argint.**
6. Conferința națională cu participare internațională „Integrare prin Cercetare și Inovare” 10 – 11 .11.2022, Universitatea de Stat din Moldova, Chișinău, Republica Moldova.
7. The International Student Innovation and Scientific Research Exhibition Cadet INOVA’22 Academia Forțelor Terestre Nicolae Bălcescu, 7 - 9 aprilie 2022, Sibiu, România.
8. International Students' Conference StudMath-IT 2021 „Aurel Vlaicu” University of Arad, Faculty of Exact Science Arad, Romania, November 18-19, 2021.
9. International Virtual Conference „Mathematics & IT: Research and Education (MITRE-2021), July 01-03, 2021, Universitatea de Stat din Moldova, Chișinău, Republica Moldova.

10. Conferința științifică națională a doctoranzilor dedicată aniversării a 75 de ani al USM „Metodologii contemporane de cercetare și evaluare”, 22-23 aprilie 2021, Universitatea de Stat din Moldova, Chișinău, Republica Moldova.
11. International Students' Conference StudMath-IT 2020 „Aurel Vlaicu” University of Arad, Faculty of Exact Science Arad, Romania, November 26-27, 2020.

Implementarea rezultatelor științifice:

Modelele elaborate au fost premiate cu:

- Două medalii de aur la Salonul Internațional al Inovării și Cercetării Științifice Deva în anii 2021 și 2022 [200 - 201]
- Medalia de argint și Diploma de Excelență la Salonul Internațional „EuroInvent”, Iași 2022, 2023 [200]
- Medalia de bronz la Salonul Internațional de Invenții și Inovații „Traian Vuia” 2022 [200].
- La AGEPI au fost înregistrate drepturile de autor asupra lucrărilor „Model de analiză statistică a datelor privind conținutul de azot și fosfor în boabele de soia” (Seria 0, Nr. 7091 din 08.12.2021) (anexa 8) și „Model de predicție a productivității unor culturi agricole” (Seria 0Ș, Nr. 7552 din 21.06.2023), (anexa 9).

Publicații științifice

În cadrul domeniului abordat de această teză de doctorat, au fost înregistrate 15 contribuții științifice (conform Anexei 32). Acestea se distribuie în următoarele categorii: 6 lucrări cu autor unic, 7 lucrări cu autor principal și 1 lucrare cu rol de coautor. Din totalul acestora, se remarcă prezența a 2 publicații în reviste de specialitate din Moldova, clasificate ca fiind de nivel B+. De asemenea, în contextul conferințelor științifice, au fost prezentate și publicate 9 articole, inclusiv rezumatele aferente, dintr-un total de 14 articole în calitate de autor principal și 1 în calitate de coautor.

1. **Ganea, I.** Influența unor substanțe biologice active asupra creșterii și dezvoltării plantelor de soia cultivate pe medii nutritive apoase Knop. In: StudMath-IT 2020 Students Conference organized by the Faculty of Exact science within "Aurel Vlaicu" University of Arad, 13th edition, p. 12-13.
2. **Ganea, I.** Bîrsan, A. Smântâna, L. Căpățână, Gh. Model de analiză statistică a datelor privind conținutul de azot și fosfor în plante. Drept de autor. Seria O Nr. 7091 din 08.12.2021.

3. **Ganea, I.** Elaborarea bazei de date de tip graf pentru analiza datelor biologice. Volumul lucrărilor: International Students' Conference „StudMath-IT 2021”, „Aurel Vlaicu” University of Arad, Faculty of Exact Science Arad, Romania, November 18-19, 2021
4. **Ganea, I.** Sistem suport pentru studiul dezvoltării plantelor cultivate pe medii M1. Conferința științifică națională a doctoranzilor „Metodologii contemporane de cercetare și evaluare” 22-23 aprilie 2021. Chișinău, 2022, CEP USM. CZU:004.89:[57.087.1:631.53], pag. 103 - 107.
5. **Ganea, I.** Use of bio-statistics in data analysis of the biological active substances on plants. Mathematics & IT: Research and education (MITRE – 2021), International Conference. Abstracts. Chișinău, 2021. CEP USM, 2021. ISBN 978-9975-158-19-0. Pag. 104.
6. **Ganea, I.** Bîrsan, A. Biological data processing. International Exhibition InventCor. – II-nd edition. Posters. ISBN 978-606-35-0467-9, p. 44. **Medalie de Aur.**
7. **Ganea, I.** Bîrsan, A. Căpățână, G. Using the neo4j pearson similarity algorithm for the analysis of protein content in soybean. The Official Catalogue of the “Cadet INOVA” Exhibition Research and Innovation in the Vision of Young Researchers. “Nicolae Bălcescu” Land Forces Academy of Sibiu. Scientific Bulletin Addendum No. 7/2022 April 7-9, 2022. Academia Forțelor Terestre „Nicolae Bălcescu” din Sibiu. ISSN 2501-3157. Pag. 120 – 131.
8. **Ganea, I.** Bîrsan, A. Use of Neo4j algorithms in biological data processing. Proceedings of the 14th edition of Euroinvent European Exhibition of Creativity and Innovation, Iași, 2022. ISSN Print: 2601-4564, Online: 2601-4572. Pag. 167 – 168. **Medalie de argint.**
9. **Ganea, I.** Bîrsan, A. Use of neo4j link prediction algorithms in biological data processing. În „Salonul Internațional de Invenții, Inovații „Traian Vuia””. Timișoara, 08-10 octombrie 2022 / coord. : Romi Rădulescu. Timișoara, Editura Politehnica, 2022, p. 77, ISBN 978-606-35-0496-9. **Medalie de bronz.**
10. **Ganea, I.** Bîrsan, A. Căpățână, G. Model de analiză a scurgerii electroliților la plantele de soia. Prezentat la Conferința științifică națională cu participare internațională „Integrare prin Cercetare și Inovare”, 10.11.2022.
11. **Ganea, I.** Model de predicție a productivității unor culturi agricole. Drept de autor. Seria OȘ Nr. 7552 din 21.06.2023
12. Bîrsan, A. Sprincean, V. **Ganea, I.** Oportunități de utilizare în practică a tehnologiilor inteligente în evaluarea unor variabile biologice. În: Studia Universitatis Moldaviae, Nr. 1(2023), pp. 53-64. DOI: [https://doi.org/10.59295/sum1\(171\)2023_07](https://doi.org/10.59295/sum1(171)2023_07).

13. Ganea, I. Model of statistical data analysis on nitrogen content in soybeans (Glicine max Merrill) in Clavera variety. Journal of Engineering Science 2023, 30 (1), 99. 165-177. [https://doi.org/10.52326/jes.utm.md.2023.30\(1\).14](https://doi.org/10.52326/jes.utm.md.2023.30(1).14).
14. **Ganea, I., Căpățână, Gh.** A predictive model for determination probabilities of drought (în tipar).
15. **Ganea, I.** Prediction Model for Soybean Productivity. Database Systems Journal (în tipar).

Conținutul principal al tezei

Teza este compusă din introducere, patru capitole, concluzii și recomandări finale, bibliografie, 31 anexe. Conținutul tezei include 132 de pagini, 37 figuri și 17 tabele. Datorită faptului că anexele, în număr de 32, ocupă un spațiu considerabil, au fost incluse într-un volum separat.

În **Introducere** este prezentată actualitatea și importanța temei. A fost identificat scopul tezei de doctorat, obiectivele cercetării, metodologiile de cercetare, rezultatele științifice obținute și valoarea lor aplicativă.

Capitolul I reflectă stadiul actual al cunoștințelor și cercetărilor holistice din domeniul implementării tehnologiilor inteligente în biologie și agricultură. Au fost analizate studiile și experimentele relevante care au fost realizate pentru a explora potențialul acestor tehnologii în îmbunătățirea performanțelor și sustenabilității sistemelor biologice și agricole. De asemenea, au fost determinate avantajele și beneficiile aduse de tehnologiile inteligente în diferite aspecte ale biologiei și agriculturii, cum ar fi optimizarea proceselor de producție, monitorizarea și gestionarea eficientă a resurselor, precum și anticiparea și prevenirea unor probleme specifice. Prin înțelegerea stadiului actual și a direcțiilor viitoare, cercetarea în acest domeniu va contribui la avansarea cunoașterii și la aducerea de îmbunătățiri semnificative în practicile agricole și în înțelegerea proceselor biologice complexe.

Capitolul II conține descrierea detaliată a metodelor de cercetare și a resurselor utilizate, precum și a abordărilor metodologice aplicate în cercetare. Acest capitol descrie limbajul formal modelele generice și specifice create pentru procesarea datelor. Dezvoltarea bazei de cunoștințe reprezintă un pas esențial pentru a obține informații relevante și consistente pentru analizele biostatistice a datelor de fiziologie a plantelor și depozitele de date graf. În acest sens, se descriu metodele biostatistice folosite în analiza datelor experimentale din fiziologia plantelor. Un aspect important abordat în teză se referă la modelarea datelor experimentale, explicând cum se colectează, prelucreează și interpretează datele pentru a extrage informații semnificative cu referire la variația unor parametri morfofiziologici și biochimici ai plantelor, dar și despre derularea unor procese fiziologice din sistemele vegetale, în corelație cu mediul de viață. Acest

proces este fundamental pentru înțelegerea mai aprofundată a fenomenelor fiziologice și pentru fundamentarea deciziilor în agricultură. De asemenea, teza se concentrează pe utilizarea depozitelor de date graf în agricultură, evidențiind aspectele generale ale dezvoltării depozitului de date și rolul său în gestionarea datelor specifice domeniului agricol.

Capitolul III sunt descrise unele modele elaborate de analiză biostatistică a datelor experimentale, obținute în laboratorul de Securitate biologică din cadrul facultății „Biologie și Geoștiințe” a Universității de Stat din Moldova. Au fost identificate și descrise rezultatele principale ale influenței unor substanțe biologice active asupra creșterii și dezvoltării plantelor de soia (*Glycine max* (L.) Merrill.). Prin intermediul analizei biostatistice riguroase, s-au identificat corelații semnificative între efectele unor compuși biologici activi și procesele fiziologice ale plantelor, oferind o mai bună înțelegere a impactului lor în contextul condițiilor specifice din Republica Moldova. Un aspect important abordat în acest capitol este evidențierea genotipurilor cu potențial util și reziliență la condițiile climatice din Republica Moldova, în contextul schimbărilor climatice actuale. Aceste genotipuri prezintă o importanță deosebită pentru selectarea varietăților de soia care pot supraviețui și se pot dezvolta optim în condițiile de mediu variabile.

Capitolul IV al tezei prezintă detaliat depozitul de date graf creat pentru a evidenția conexiunile relevante în vederea optimizării proceselor agricole. În acest capitol, este explorat graful folosit pentru procesarea datelor meteorologice și definirea nodurilor și relațiilor în *Neo4j* pentru datele menționate. Unul dintre aspectele importante abordate în acest capitol este modelul de predicție a fenomenelor meteorologice bazat pe datele stocate în depozitul de date. Acest model permite anticiparea evoluției condițiilor meteorologice, fiind esențial în gestionarea și optimizarea activităților agricole. De asemenea, este redat rezultatul modelului de predicție a productivității unor produse agricole, utilizând datele disponibile în depozitul de date graf. Acest model are potențialul de a sprijini agricultorii în luarea deciziilor, optimizând randamentele culturilor agricole.

1. PROCESAREA INTELIGENTĂ A DATELOR ÎN BIOLOGIE ȘI AGRICULTURĂ

1.1. Fundamentele Tehnologiilor Inteligente

1.1.1. Tehnologii inteligente în biologie și agricultură

Factori globali precum schimbările climatice, modificările demografice și problemele securității alimentare determină cercetătorii și producătorii agricoli să caute noi abordări inovatoare pentru dezvoltarea culturilor agricole și sporirea randamentului acestora. Cele mai recente progrese în tehnologia și practicile agricole sunt descrise în literatura de specialitate prin termenii de *Agricultura 4.0* și *Agricultura 5.0*.

Agricultura 4.0, cunoscută și sub denumirea de *Agricultura de precizie*, implică utilizarea unor tehnologii avansate, cum ar fi senzori, drone și cartografiere GPS pentru a colecta și analiza date despre plante și condițiile solului. Scopul *Agriculturii 4.0* este de a crește eficiența, de a reduce cantitatea de deșeuri și de a îmbunătăți recoltele, minimizând, în același timp, impactul asupra mediului [6-7].

Începând cu anul 2021 se desfășoară o nouă revoluție în agricultură, *Agricultura 5.0*. Denumită și *Agricultură Regenerativă* sau *Agricultura Inteligentă*, aceasta dezvoltă agricultura de precizie prin introducerea roboticii și a *Inteligenței Artificiale (IA)*. Noul model de agricultură va înlocui forța de muncă umană, facilitând luarea deciziilor în mod autonom [8]. Acest concept, se află încă în stadiile incipiente de dezvoltare și subliniază o abordare holistică a agriculturii care se concentrează pe sănătatea solului, conservarea biodiversității și îmbunătățirea serviciilor ecosistemice. *Agricultura 5.0* implică utilizarea tehnologiilor și practicilor de ultimă oră pentru a menține sănătatea solului și pentru a îmbunătăți starea generală și reziliența ecosistemelor agricole. Scopul final este de a crea sisteme durabile și regenerative care să aducă beneficii atât fermierilor, cât și mediului [9]. De menționat că, atât *Agricultura 4.0*, cât și *Agricultura 5.0* reprezintă progrese semnificative în tehnologia și practicile agricole și au potențialul de a transforma modul în care se vor produce alimente în anii următori. Această abordare progresivă implică utilizarea *Big Data*, *Inteligenței Artificiale* și a altor discipline ale științei informației cu

1.1.2. Concepte fundamentale de cunoaștere și înțelegere

Datele și cunoștințele posedă caracteristici diferite și funcționează diferit. În vederea obținerii rezultatului anticipat, este imperativ de a defini în mod precis noțiunile de dată și informație, entități caracterizate prin semnificații distincte. Astfel, **datele** reprezintă consecutivități de simboluri/hieroglife (cuvinte) într-un oarecare alfabet formal sau informal. Datele luate ca atare nu au nicio semnificație [11].

Datele organizate și utilizate într-un anumit context devin informație.

Informația este:

- (a) semnificația datelor,
- (b) plus noile semnificații ale datelor care pot fi obținute cu ajutorul cunoștințelor din semnificațiile cunoscute eventual obținute ale datelor [12].
- (c) Fiecare dintre elementele noi, în raport cu cunoștințele prealabile, cuprinse în semnificația unui simbol sau a unui grup de simboluri (text scris, mesaj vorbit, imagini plastice, indicație a unui instrument etc.) [13].

Apariția de noi fapte în baza de cunoștințe și/sau stabilirea de noi conexiuni între acestea implică schimbări în procesul de luare a deciziilor la rezolvarea problemelor. Datele sunt inutile fără analize. Posesia unei mulțimi de date nu duce la cunoaștere. Capacitatea de a extrage și de a controla acele date este ceea ce duce la perspicacitate și înțelepciune [13].

Cunoașterea este o stare reală de fapt, bazată pe fapte și argumente raționale, justificate logic sau faptic și verificate empiric sau practic. Știința și metoda ei științifică urmăresc obținerea de cunoștințe despre structura obiectelor și fenomenelor, despre relațiile lor esențiale [14].

Atât știința informației, cât și alte științe prelucrează nu numai date, ci și cunoștințe. În teoria inteligenței artificiale și a sistemelor expert, cunoașterea este un set de enunțuri despre proprietățile obiectelor, modele de procese și fenomene, precum și regulile pentru concluzia logică a declarațiilor altora și regulile de utilizare a acestora pentru a lua decizii. Principala diferență dintre cunoștințe și date este structura și activitatea lor: apariția unor fapte noi în baza de cunoștințe sau stabilirea de noi conexiuni între ele poate deveni o sursă de schimbare în procesul decizional [15].

Analiza datelor (*data analytics*) este descoperirea de modele și tendințe culese din date. *Analza datelor* este modul în care sunt înțelese datele și descoperirea tendințelor semnificative. *Modelul* este ceea ce un program încearcă să identifice într-un set de date, folosindu-se care încearcă să identifice un model similar într-un alt set de date. Datele sunt procesate pentru a identifica tipare și pentru a modela programul pentru realizarea sarcinilor [15].

În urma analizei informațiilor se obține *perspicacitatea* (*insights*) care poate fi definită ca capacitatea de pătrundere în esența lucrurilor, profunzime, surprindere și înțelegere a ceea ce scapă majorității [17-18]. În tehnologia informației *perspicacitatea de date* se referă la înțelegerea în interior a unui anumit fenomen (ce face de fapt acest model). Perspicacitatea folosește tehnologia de *Învățare Automata* (în engleză: *Machine Learning* – ML) și *Învățare Profundă* (*Deep Learning* - DL), componente ale Inteligenței Artificiale pentru procesarea

seturilor de date. În IA se utilizează termenul *Perspicacitate Artificială* (*Artificial Insights – AIns*) cu referință la capacitatea mașinilor de pătrundere profundă a fenomenelor. De exemplu, un *model de învățare automată* care estimează probabilitatea ca un lot de plante să fie mai sensibil la factori de stres va dezvălui care este impactul unor factorii care determină dezvoltarea plantelor, permițând biologilor sau agricultorilor să adapteze strategiile la noile condiții. Aceasta facilitează o perspectivă mai profundă asupra problemei investigate, furnizând înțelepciune esențială cu privire la componentele fundamentale ale acesteia și evidențiind intervențiile necesare pentru soluționare. Cu toate acestea, aceste informații nu pot fi dobândite în absența analizelor, iar analiza este inutilă fără date concludente.

Cunoștințele reprezintă culegeri de date, informații, adevăruri și principii învățate, acumulate în timp. Informațiile despre structura obiectelor și fenomenelor, despre relațiile lor esențiale care sunt reținute și înțelese care pot fi utilizate în luarea deciziilor, formează judecăți și opinii care devin cunoștințe [14,15].

Bază de cunoștințe (*Knowledge Base - KB*) semnifică o tehnologie folosită pentru a stoca informații complexe, atât structurate cât și nestructurate utilizate de către un sistem informatic. Cunoașterea este o stare reală de lucruri, bazată pe fapte și argumente raționale, justificată logic sau de fapt și verificată empiric sau practic [19]. Actualmente, tehnologiile moderne avansate atestă o expansiune vertiginoasă, fiind utilizate pe larg în diverse activități umane, oferind un suport solid în luarea deciziilor pentru monitorizarea sistemelor agricole.

Inteligența reprezintă capacitatea de a înțelege cu ușurință și cu acuratețe, eficient și în profunzime, de a sesiza ceea ce este esențial, de a rezolva situații sau probleme noi pe baza experienței acumulate anterior [20]. De asemenea, este facultatea de a descoperi proprietățile obiectelor și fenomenelor înconjurătoare, cât și a relațiilor dintre acestea, dublată de posibilitatea de a rezolva probleme noi [21-22].

Asemeni ființelor biologice înzestrate cu inteligență naturală mașinile pot fi dotate cu inteligență artificială, creată de oameni. Termenul *Inteligența Artificială* (abreviat IA) se referă la capacitatea mașinilor de a imita funcțiile umane, cum ar fi raționamentul, învățarea, planificarea și creativitatea. Prin intermediul IA, sistemele tehnice pot percepe mediul în care se află, procesează aceste informații și rezolvă probleme pentru a atinge un anumit obiectiv. John McCarthy a definit IA ca fiind știința și ingineria care se ocupă cu crearea mașinilor inteligente, în special a programelor de calculator inteligente. Acest domeniu este strâns legat de utilizarea calculatoarelor pentru a înțelege inteligența umană, dar IA nu este limitată la metodele observabile biologice [23].

Similar domeniului Inteligența Artificială se folosește conceptul *Inteligența Computațională* (IC) care este definită ca modelare a inteligenței biologice, spre deosebire de IA care se bazează pe noțiunea de cunoștințe [24].

Cel mai înalt nivel al inteligenței artificiale este ***Înțelepciunea artificială*** care poate fi descrisă ca inteligență artificială care atinge cel mai înalt nivel de luare a deciziilor atunci când se confruntă cu cele mai complexe situații provocatoare. Termenul de înțelepciune artificială este folosit atunci când *inteligența* se bazează mai mult decât pe colectarea și interpretarea de date întâmplătoare, ci prin proiectare îmbogățită cu strategii inteligente și de conștiință pe care oamenii înțelepți le-ar folosi [25].

Tehnologiile inteligente, cum ar fi inteligența artificială, învățarea automată și roboții autonomi, aduc cu sine o serie de avantaje, printre care:

- a) ***Eficiență crescută***: Tehnologiile inteligente pot automatiza și optimiza procese complexe, ceea ce duce la o creștere a eficienței și productivității. Ele pot procesa și analiza cantități mari de date într-un timp scurt, oferind rezultate mai rapide și mai precise decât metodele tradiționale.
- b) ***Luarea deciziilor bazate pe date***: Tehnologiile inteligente pot furniza informații și analize bazate pe date pentru a sprijini luarea deciziilor. Ele pot identifica modele și tendințe ascunse în datele colectate, oferind astfel o bază solidă pentru luarea deciziilor informate în diverse domenii, cum ar fi sănătatea, agricultura sau business-ul.
- c) ***Automatizare și autonomie***: Tehnologiile inteligente permit automatizarea sarcinilor și proceselor, eliminând necesitatea intervenției umane constante. De exemplu, roboții autonomi pot efectua sarcini repetitive sau periculoase în locul oamenilor, ceea ce duce la creșterea siguranței și reducerea erorilor umane.

1.1.3. Structura sistemelor de inteligență artificială

Un sistem de inteligență artificială, sau sistem bazat pe cunoștințe, este format din două componente fundamentale:

- (1) ***baza de cunoștințe*** care conține cunoștințele specifice domeniului problemei și, eventual, cunoștințe generale, și
- (2) ***motorul de inferență*** care utilizează cunoștințele pentru rezolvarea problemei. Metoda de inferență (inferența) reprezintă procedura de obținere la un moment dat, a noi elemente (fapte) implicate în mod direct de elementele particulare reprezentării. Un motor de inferență este o componentă a unui sistem de inteligență artificială sau a unui sistem bazat pe reguli care are capacitatea de a extrage concluzii și de a face deducții logice pe baza informațiilor disponibile.

Motorul de inferență utilizează reguli logice și cunoștințe predefinite pentru a evalua și combina diferite premise și fapte și pentru a genera rezultate noi [19].

Acesta este utilizat în diverse domenii, inclusiv în sistemele expert, în sistemele de recomandare și în procesarea limbajului natural, contribuind la luarea deciziilor și la rezolvarea problemelor complexe. Prin aplicarea regulilor și prin manipularea datelor, motorul de inferență poate genera concluzii sau acțiuni pe baza inferențelor logice realizate [26].

Din punct de vedere al inteligenței artificiale cunoștințele pot fi văzute ca mulțimea de fapte și principii acumulate de oameni sau, mai general, nivelul cunoașterii umane la un anumit moment dat. Cunoștințele includ idei, concepte, teorii, abstractizări, limbaje, reguli, locuri, obiceiuri. Cunoștințele dintr-un sistem inteligent, care se referă de obicei la un anumit domeniu, descriu universul problemei sau al clasei de probleme de rezolvat la nivel simbolic și formează conținutul bazei de cunoștințe.

Scopul final al inteligenței artificiale (IA) este de a dezvolta tehnologii care sunt cele mai capabile să servească umanitatea. Acest lucru necesită progrese care depășesc componentele de bază ale inteligenței generale. Termenul *inteligență* nu reprezintă cel mai bine nevoile tehnologice ale societății în avans, deoarece *înțelepciunea*, mai degrabă decât *inteligența*, este asociată cu o mai mare bunăstare, fericire, sănătate și poate chiar longevitatea individului și a societății. Astfel, nevoia viitoare în tehnologie este de *înțelepciune artificială* (en.: *Artificial Wisdom* - AW) [27].

Conform definiției Comisiei Europene există două tipuri de inteligență artificială:

- **Software:** asistenți virtuali, programe informatice de analiză a imaginilor, motoare de căutare, sisteme de recunoaștere vocală și facială;
- **IA încorporată:** roboți, automobile autonome, drone, Internetul Lucrurilor (Internet of Things – IoT) [28].

Inteligența Artificială explorează datele din resursele disponibile fără nicio intervenție umană. Apoi este procesat de algoritmul *ETL* (*Extract, Transform, Load*) pentru a explora baza de date mare și găsește automat relația dintre conținut și soluția necesară la problemă [29]. Sistemul inteligent învață din activitatea zi de zi pentru a fi actuală permanent și identifică evenimentele viitoare prin clasificare, clasare și regresie.

Printre tehnologiile inteligente utilizate sunt *Data Mining* (DM), *Text Mining* (TM), *Machine Learning* (ML), *Deep Learning* (DL).

Data Mining reprezintă procesul netrivial de extragere a datelor, informațiilor și modelelor dintr-o bază de date. *DM* implică extragerea de informații implicite și anterior necunoscute, care sunt interesante și potențial utile dintr-o cantitate mare de date. Aceste

informații pot fi prezentate sub forma de modele și șabloane de cunoaștere. Scopul final este descoperirea cunoștințelor în bazele de date care reprezintă procesul netrivial de identificare a modelelor de date valide, noi, potențial utile și în cele din urmă ușor de înțeles [30-35].

Termenul DM este adesea utilizat ca sinonim pentru conceptul de *descoperire a cunoștințelor în baze de date (Knowledge Discovery in Databases - KDD)* [33,34].

Inteligența artificială conține mai multe subdomenii printre care *Învățarea Automată (Machine Learning -ML)* și *Învățarea profundă (Deep Learning - DL)* [36-37].

- ML antrenează o mașină cum să învețe [38].
- DL este un tip de învățare automată care instruește un computer să execute sarcini asemănătoare celor umane, precum recunoașterea vorbirii, clasificarea imaginilor, detectarea obiectelor și chiar descrierea conținutului. Un exemplu din viața reală sunt ChatGPT, Google Bard, asistenții virtuali, cum ar fi Siri și Cortana, care sunt bazați, parțial, pe învățarea profundă.

Un rol important în analiza seturilor mari de date biologice și complexe o are învățarea automată. Deși există multe aplicații pentru metodele de învățare automată, aplicațiile acestora la datele biologice au fost în predicția genelor, adnotarea funcțională, biologia sistemelor, analiza datelor cu microarray, analiza căilor etc. [39-41].

Studii recente examinează provocările și limitările asociate utilizării învățării automate (ML) și învățării profunde (DL) în identificarea bolilor plantelor, inclusiv problemele legate de disponibilitatea datelor, calitatea imaginilor și distincția între plantele sănătoase și cele bolnave [42]. Acestea oferă perspective valoroase cercetătorilor, practicienilor și profesioniștilor din industrie cu privire la detectarea bolilor plantelor, prezentând soluții pentru aceste provocări și limitări, și furnizând o înțelegere cuprinzătoare a stadiului actual al cercetărilor în acest domeniu [43-44]. De asemenea, se evidențiază beneficiile și limitările metodelor ML și DL și se propun soluții potențiale pentru depășirea provocărilor implementării lor [39,45-46].

În cadrul procesului de analiză a datelor, adesea ne confruntăm cu problema disponibilității unui set de date redus, care nu furnizează suficiente informații pentru obținerea unui rezultat precis. Seturile de date mici pot avea un impact semnificativ asupra preciziei predicțiilor privind seceta sau productivitatea plantelor. Deoarece volumul de date este limitat, devine crucial să se trateze această problemă cu atenție și să se utilizeze tehnici și modele adecvate pentru a obține rezultate relevante [47]. În ciuda tuturor eforturilor depuse, este notabil faptul că, în cazul seturilor de date de dimensiuni reduse, previziunile pot manifesta o incertitudine mai accentuată. Este imperativ să se evidențieze asupra acestui nivel de

incertitudine în procesul de interpretare a rezultatelor obținute și să se extindă, pe cât posibil, dimensiunea setului de date în scopul reducerii acestei incertitudini.

Printre aspectele importante de avut în vedere este utilizarea tehnicilor de *Învățare prin Transfer* (*Transfer Learning* - TL). Tehnicile TL pot fi o abordare valoroasă atunci când datele sunt limitate. Se pot utiliza modele pre-antrenate și să se ajusteze pentru a se potrivi noilor date, ceea ce poate îmbunătăți performanța. În ultimii ani, datorită dezvoltării modelelor de învățare profundă și a creșterii disponibilității seturilor de date masive, a devenit mai prominentă Învățarea prin Transfer (TL). Abordarea TL presupune utilizarea unui set de tehnici care extind această abordare prin capitalizarea pe cunoștințele dobândite din domenii sau sarcini conexe pentru a instrui un model cu capacități de generalizare îmbunătățite [48].

TL este o tehnică în învățarea automată (*Machine Learning*) care implică utilizarea cunoștințelor și experienței dobândite într-un domeniu pentru a îmbunătăți performanța unui model sau algoritmi într-un alt domeniu sau sarcină similară [49]. Această tehnică se bazează pe premisa că informațiile și caracteristicile învățate dintr-un domeniu pot fi utile și aplicabile în alt domeniu, în special atunci când setul de date din domeniul țintă este limitat sau insuficient pentru a antrena un model eficient de la zero.

Principalele avantaje ale Transfer Learning:

- ***Reducerea necesității de date etichetate ample:*** În loc de a colecta și eticheta un set mare de date pentru fiecare sarcină specifică, se pot utiliza modele pre-antrenate care au fost deja instruite pe seturi mari de date. Acest lucru economisește timp și resurse.
- ***Îmbunătățirea performanței:*** Modelele pre-antrenate aduc cu ele cunoștințe din domeniul sursă, ceea ce poate duce la o inițializare mai bună și o convergență mai rapidă în timpul antrenării pe datele din domeniul țintă.
- ***Generalizare mai bună:*** Modelele pre-antrenate sunt adesea învățate pe date diverse, ceea ce le poate ajuta să generalizeze mai bine pe datele dintr-un domeniu țintă.

Există două abordări principale în *Transfer Learning*:

- ***Fine-tuning* (rafinarea):** În această abordare, se iau modele pre-antrenate (cum ar fi rețele neurale profunde) și se ajustează parametrii acestora pe baza datelor din domeniul țintă. Parametrii pre-antrenați sunt înghețați sau ajustați într-o anumită măsură pentru a se potrivi noilor date. Acesta este un mod comun de a aplica Transfer Learning în învățarea profundă.

- *Extragerea de caracteristici*: Se pot extrage caracteristicile latente învățate din modelul pre-trenat, fără a fi necesară ajustarea întregului model și utilizarea într-un alt model specific sarcinii din domeniul țintă. Acesta este frecvent utilizat atunci când setul de date din domeniul țintă este mic sau când nu sunt necesare toate caracteristicile învățate din modelul sursă.

Transfer Learning poate fi aplicat într-o varietate de domenii, inclusiv în analiza de imagini, prelucrarea limbajului natural și recunoașterea de vorbire. Este o tehnică puternică și versatilă care a revoluționat capacitatea de a construi modele de învățare automată performante într-o gamă largă de aplicații [49-50].

Modelarea biostatistică a datelor în domeniul fiziologiei plantelor implică utilizarea metodelor și tehnicilor statistice pentru a analiza și interpreta datele experimentale legate de funcționarea și răspunsurile plantelor la diferite stimuli și condiții de mediu. Aceste modele pot ajuta la înțelegerea proceselor fiziologice la nivel molecular, celular și de organism integru și pot oferi informații valoroase despre adaptarea plantelor la factorii de mediu și la schimbările climatice.

1.1.4. Învățarea automată și statistica

Granița dintre învățarea automată și statistică poate fi dificil de stabilit. Unele metode sunt comune ambelor domenii și oricare poate fi folosit pentru predicție și inferență. Cu toate acestea, învățarea automată și statisticile au focus, predicții sau inferențe diferite [51-52]. În general, metodele statistice clasice se bazează pe ipoteze despre sistemele de generare a datelor. Statisticile pot furniza inferențe explicite prin adaptarea unui model de probabilitate specificat atunci când sunt colectate suficiente date din studii bine concepute. Învățarea automată este preocupată de problema creării și aplicării algoritmilor care se îmbunătățesc pe parcursul timpului. Multe metode de învățare automată pot deriva modele pentru recunoașterea modelelor, clasificarea și predicția din datele existente și nu se bazează pe ipoteze stricte despre sistemele de generare a datelor, ceea ce le face mai eficiente în unele aplicații complicate.

Învățarea automată a fost utilizată pe scară largă în studiile biologice pentru predicție și descoperire. Odată cu disponibilitatea tot mai mare și a diferitelor tipuri de date omice, aplicarea metodelor de învățare automată, în special a abordărilor de învățare profundă, a devenit mai frecventă [44]. Un domeniu de oportunitate pentru abordările de învățare automată este predicția caracteristicilor genomice, în special a celor care sunt greu de prezis folosind abordările actuale. În biologie, algoritmi de învățare profundă procesează date în moduri în care oamenii nu pot, detectând caracteristici care altfel ar fi imposibil de surprins [53].

1.2. Integrarea Datelor în Cercetarea Avansată

1.1.1. Procesarea inteligentă a datelor (PID)

PID reprezintă un domeniu de cercetare în plină dezvoltare, care a fost catalizat de progresele recente în tehnologia informației și în special în domeniul inteligenței artificiale [54]. PID este o metodă de a extrage informații valoroase din datele colectate, prin utilizarea tehnologiilor inteligente de analiză, precum *Machine Learning*, inteligența artificială sau rețele neuronale. Aceste tehnologii sunt utilizate cu succes într-o gamă largă de domenii, inclusiv biologie și agricultură.

În biologie, PID poate fi utilizată pentru analiza datelor genetice, pentru a identifica mutații sau variații ale genelor, și pentru a dezvolta modele predictive ale bolilor sau caracteristicilor organismelor. De exemplu, PID poate fi utilizată pentru a identifica markeri genetici asociați cu rezistența la secetă sau îngheț, pentru a identifica specii de plante cu anumite caracteristici sau calități.

Deosebit de importantă este utilizarea IA în sectoarele economice fundamentale pentru dezvoltarea societății umane, în special, în agricultură. Aplicațiile inteligenței artificiale utilizate în agricultură se împart în trei categorii principale:

- **Roboți agricoli** – Fermierii utilizează roboți autonomi dezvoltați și programați de către companii TI pentru a îndeplini sarcini agricole, cum ar fi tratarea plantelor, aplicarea stimulatorilor și fertilizanților, eliminarea buruienilor, recoltarea culturilor în cantități mai mari și mai rapid decât oamenii.
- **Monitorizarea culturilor și a solului** – Cercetătorii și producătorii agricoli folosesc tehnologiile inteligente, managementul asistat de calculator și algoritmi de învățare profundă pentru colectarea și procesarea datelor colectate de drone și senzori sau tehnologii software pentru monitorizarea stării culturilor și a solului.

Analiză predictivă – Sunt dezvoltate modelele de învățare automată pentru a monitoriza și a prezice diferite efecte ale mediului asupra randamentului culturilor, cum ar fi cantitatea de precipitații, temperatura aerului, viteza vântului, calitatea aerului și alte fenomene [56].

În agricultură, PID poate fi utilizată pentru a analiza datele meteorologice, pentru a dezvolta modele de predicție a recoltelor și a condițiilor de creștere a plantelor, sau pentru a monitoriza și a gestiona eficient utilizarea resurselor, precum apa sau îngrășămintele. De exemplu, PID poate fi utilizată pentru a prezice când să fie plantate culturile, pentru a optimiza schema de irigare sau pentru a identifica zonele cu soluri fertile. De asemenea, pot fi selectate soiurile rezistente la îngheț pentru a fi plantate mai devreme în scopul creșterii și dezvoltării plantelor atunci când resursele de umiditate sunt suficiente.

Procesarea inteligentă a datelor poate fi utilizată în domeniul fiziologiei plantelor pentru a analiza și interpreta datele colectate în ceea ce privește comportamentul și performanța plantelor. Există o varietate de tehnologii utilizate în domeniul fiziologiei plantelor pentru a colecta date, cum ar fi senzori pentru măsurarea umidității solului, temperaturii și umidității aerului, precum și pentru măsurarea fluorescenței clorofilei și a unor procese metabolice ale plantelor (fotosinteza, respirația, fermentația, absorbția și transportul nutrienților, metabolismul lipidic și sinteza proteinelor, reacții de detoxifiere, ciclul Krebs și lanțul de transport al electronilor).

În domeniul fiziologiei plantelor se utilizează tehnici de analiză și modelare a datelor pentru a înțelege mai bine comportamentul plantelor și a identifica factorii care influențează creșterea și dezvoltarea acestora. În acest sens, procesarea inteligentă a datelor în domeniul fiziologiei plantelor poate oferi o abordare inovatoare pentru cercetarea și îmbunătățirea eficienței sistemelor agricole și a producției de plante [57].

Procesarea inteligentă a acestor date poate ajuta la identificarea modelelor și tendințelor din datele colectate, ceea ce poate ajuta la îmbunătățirea practicilor agricole și la optimizarea creșterii plantelor. De exemplu, algoritmi de învățare automată pot fi folosiți pentru a identifica relații între date utilizate pentru a anticipa comportamentul plantelor în funcție de anumite condiții, cum ar fi temperatura, umiditatea, nutriție și nivelul de lumină. Aceste informații pot fi utilizate pentru a lua decizii în timp real referitoare la irigarea, fertilizarea și alte practici agricole. Procesarea inteligentă a datelor poate ajuta, de asemenea, la identificarea plantelor care prezintă un risc mai mare de stres hidric sau de alte probleme, permițând astfel intervenții rapide și îmbunătățirea performanței plantelor [58].

Analiza inteligentă a datelor în domeniul fiziologiei plantelor reprezintă o zonă de cercetare în continuă dezvoltare, care oferă numeroase oportunități și provocări.

Oportunitățile acoperă următoarele aspecte:

1. *Îmbunătățirea randamentului culturilor:* analiza inteligentă a datelor poate ajuta la identificarea factorilor care afectează randamentul culturilor și la dezvoltarea de strategii pentru îmbunătățirea producției agricole.
2. *Îmbunătățirea sănătății plantelor:* prin monitorizarea datelor legate de starea plantelor, se pot identifica problemele de sănătate într-un stadiu incipient, ceea ce poate duce la o intervenție timpurie și la prevenirea răspândirii bolilor.
3. *Reducerea costurilor:* prin utilizarea analizei inteligente a datelor, se pot identifica metode de reducere a costurilor în ceea ce privește utilizarea de pesticide, erbicide și alte chimicale.

Provocările includ:

1. *Colectarea și stocarea datelor*: pentru a realiza analiza inteligentă a datelor, este necesară o colectare precisă și continuă a datelor din diferite surse, precum și o stocare adecvată a acestora.
2. *Analiza datelor*: analiza datelor poate fi dificilă și consumatoare de timp, în special în cazul datelor cu un volum mare sau cu o complexitate ridicată.
3. *Asigurarea preciziei*: este important să se asigure precizia datelor colectate, precum și a algoritmilor utilizați pentru analiza acestora, deoarece erorile pot duce la decizii incorecte și la pierderi financiare.
4. *Integritatea datelor*: este important să se protejeze integritatea datelor, pentru a evita orice tentativă de manipulare sau utilizare incorectă a acestora.

Aceste oportunități și provocări au fost studiate și abordate în lucrări diverse științifice, printre care:

1. *Intelligent Data Analysis for Plant Physiology* de Biruk Gebremariam și Mulugeta F. Dinssa, publicată în revista *Frontiers in Plant Science*. Lucrarea abordează utilizarea analizei inteligente a datelor pentru a optimiza producția agricolă și a îmbunătăți sănătatea plantelor, precum și provocările și soluțiile asociate cu acest proces [57].
2. *Smart Agriculture Using IoT and Big Data Analytics* de Shruti Devasenapathy și Sridhar Krishnan, publicată în revista *Journal of Big Data*. În lucrare se examinează cum analiza inteligentă a datelor poate fi utilizată pentru a îmbunătăți producția agricolă prin monitorizarea și analizarea datelor legate de sol, apă, plante și condiții meteorologice [58].
3. *Artificial intelligence in agriculture: A systematic review* de Pedro Luiz de Freitas, publicată în revista *Computers and Electronics in Agriculture*. Lucrarea oferă o perspectivă amplă asupra utilizării inteligenței artificiale și a analizei inteligente a datelor în agricultură, precum și o discuție despre provocările și soluțiile asociate cu utilizarea tehnologiei în domeniul agriculturii [59].

1.1.2. Abordarea holistică în cercetare

Sistemele biologice și agricole reprezintă sisteme complexe. Modul de funcționare a acestora poate fi înțeles și utilizat eficient printr-o abordare holistică [60-62]. Abordarea **holistică** a complexității sistemelor biologice se bazează pe înțelegerea faptului că componentele care formează întregul organism sunt mai mult decât suma părților lor [63]. Cercetări holistice implică abordarea unui sistem sau a unei entități biologice ca un întreg, luând în considerare interacțiunile și dependențele dintre componentele sale [64]. În contrast cu abordarea reduționistă, care descompune sistemul în părți individuale pentru a le analiza separat, cercetările holistice se concentrează pe înțelegerea globală a sistemului și a modului în care

aceste părți se influențează reciproc. Aceasta poate implica integrarea datelor și informațiilor provenite din diferite discipline științifice, cum ar fi biologia moleculară, ecologia, fiziologia și informatica [65].

Datorită abordării holistice științele, în special științele vieții și a mediului au înregistrat cele mai importante evoluții [66]. Ca domeniu de studiu, științele vieții analizează interacțiunile dintre componentele sistemelor biologice și rețelele pe care le formează, precum și modul în care aceste interacțiuni descriu funcțiile și comportamentul sistemului respectiv [67].

Biologia sistemelor reprezintă analiza și modelarea computațională și matematică a sistemelor biologice complexe și integrează multe domenii științifice [68] - biologie, informatică, bioinformatică, biofizică, bioinginerie și altele [69]. Aceasta are ca obiect de studiu interacțiunile complexe în cadrul sistemelor biologice, folosind o abordare holistică pentru a **prezice** modul în care aceste sisteme evoluează și se adaptează la diverse condiții economice [70], sociale și de mediu [71-72] pentru dezvoltarea de soluții durabile la problemele globale cum ar fi alimentația, sănătatea și mediul [73-75].

În ultimul timp, tehnologiile inteligente au condus la o explozie în generarea de date, în special al celor omice, care contribuie la realizarea studiilor avansate ale fenomenelor biologice complexe. Aceasta determină reevaluarea metodologiilor de analiză a datelor. Pentru a înțelege în profunzime fenomenele biologice complexe, cum ar fi multe procese sau trăsături cantitative și calitative la plante/animale/oameni, este nevoie de procesarea unui volum enorm și de multiple tipuri de date (*big data*) care sunt generate din studii complexe [76-77].

Astăzi, **sistemele complexe** constituie subiect de cercetare pentru o mare varietate de științe și metode de practici profesionale. În domeniul *științelor vieții*, principalul obiectiv al studiilor de complexitate ar fi acela de a explica interacțiunile dintre componentele unui sistem. Astfel de sisteme sunt folosite pentru a procesa modele în informatică, biologie, economie, fizică și multe alte domenii [78].

1.1.3. Tehnologiile Omics

Pentru rezolvarea problemelor din domeniul de cercetare și a înțelege în profunzime lucrurile este necesară studierea globală a componentelor unui sistem biologic la nivel molecular și informațional [78]. Aceste discipline poartă denumirea generică de *omice* (*omics*).

Tehnologiile omics reprezintă un domeniu interdisciplinar în bioinformatică și constau într-un ansamblu de abordări și tehnici utilizate în analiza și înțelegerea detaliată a interacțiunilor și funcțiilor de entități informaționale biologice în diverse straturi (*clusters*) ale vieții la diferite niveluri moleculare și funcționale ale organismelor, inclusiv plantele [79].

Studiile omics au revoluționat modul în care cercetăm și înțelegem sistemele biologice, permițând abordări sistemice și o mai bună înțelegere a proceselor biologice [78]. Cu toate acestea, integrarea datelor omice multidimensionale în context biologic semnificativ este o provocare, dat fiind volumul mare de date și diferențele dintre tipurile de date omice [80-81].

Domeniul omics are o gamă largă de aplicații în cercetarea agricolă, inclusiv îmbunătățirea calității produselor alimentare, protejarea mediului și remedierea acestuia, și îmbunătățirea sănătății umane [82-83]. Tehnologiile omics pot oferi informații la scară largă despre componentele biologice, iar cercetările holistice integrează aceste date pentru a obține o înțelegere mai completă a sistemelor biologice [84].

De asemenea, deosebit de importantă este utilizarea datelor omice în știința alimentației și nutriției, unde analiza omică multiplă ajută la înțelegerea funcționalității componentelor alimentare [85].

Cercetările omics și *cercetările holistice* se completează reciproc, deoarece cele omics oferă date și informații la scară largă despre componentele biologice, în timp ce cercetările holistice integrează aceste date pentru a obține o înțelegere mai completă a sistemelor biologice. De exemplu, studiile de genomică/genomics pot identifica genele implicate într-un anumit proces biologic, în timp ce cercetările holistice pot explora cum interacționează aceste gene cu alte componente moleculare și cum afectează fenotipul organismului. Cercetările omics și holistice în fenomica plantelor reprezintă o metodă puternică și complexă de investigare a sistemelor biologice ale plantelor la nivel molecular și sistemic. Această abordare aduce o înțelegere mai profundă a organismelor vegetale și a interacțiunilor acestora cu mediul lor [86], deschizând noi perspective în domeniul agriculturii durabile și al conservării resurselor naturale [87].

O altă ramură în care s-au utilizat datele omice este știința alimentației și a nutriției [88]. Cercetători au implementat analiza omică multiplă pentru a înțelege în mod holistic funcționalitatea componentelor alimentare [89]. Cu toate acestea, analiza datelor de omică integrată necesită o cantitate imensă de muncă și abilități ridicate de manipulare a datelor. În acest scop, a fost realizată o bază de date cu date nutriționale omice *Nutrigenomics*, pentru a ajuta oamenii de știință din domeniu să își analizeze mai eficient propriile *date omice*. Baza de date deține două seturi majore de date [89].

Tehnologiile omics, pot oferi informații detaliate despre expresia genelor, proteinele produse, metabolismul și compoziția chimică a plantelor în diverse condiții și în diferite stadii de dezvoltare. Aceasta implică analiza comprehensivă a genelor, proteinelor, metabolitilor și altor componente moleculare implicate în dezvoltarea, adaptarea și răspunsul plantelor la mediu.

Aceste date pot oferi o imagine de ansamblu asupra modului în care plantele reacționează la factori de mediu, cum ar fi lumină, temperatură, apă, substanțe chimice și stres biotic sau abiotic.

Cercetările omics au contribuții semnificative în Fenomica plantelor - o ramură a cercetării omics care se concentrează în mod specific pe analiza detaliată, studiarea diversității și variației fenotipurilor (caracteristicilor observabile) la plante la nivel molecular și informațional, adică caracteristicile observabile și măsurabile, care pot oferi informații importante despre comportamentul și răspunsul acestora la diferite condiții. Prin aplicarea cercetărilor fenomice în studiul plantelor, se poate obține o înțelegere mai profundă a biologiei plantelor și a modului în care acestea interacționează cu mediul înconjurător.

Tehnologiile omics și studiile omice au adus contribuții semnificative în înțelegerea sistemelor biologice, permițând o analiză detaliată a componentelor moleculare și funcționale ale organismelor și deschizând noi perspective în cercetarea agricolă, alimentație, sănătate și nutriție, scopul final fiind înțelegerea holistică a biologiei sistemelor [85]. Utilizarea combinată a tehnicilor omice crește eficiența și acuratețea analizei. Bioinformatica poate fi integrată pentru a procesa date cu metode analitice avansate [90-92].

1.1.4. Abordarea holistică în procesarea inteligentă a datelor în domeniul fiziologiei plantelor

În sistemele bio-agricole, studiul fiziologiei plantelor este important pentru că ea permite de mai bine să înțeleagă modul în care plantele răspund la condițiile de mediu, cum ar fi disponibilitatea nutrienților, a apei și a luminii și cum pot crește în mod durabil [93]. Studiul fiziologiei plantelor în sistemele bio-agricole poate include investigarea indicilor morfologici de creștere și dezvoltare, dar și fiziologici: ai fotosintezei, transpirației, respirației și a modului, în care plantele răspund la stres, cum ar fi seceta sau excesul de umiditate [94]. Cunoștințele asupra fiziologiei plantelor pot ajuta agricultorii să selecteze culturile mai bune adaptate mediului înconjurător, să îmbunătățească sănătatea solurilor și să maximizeze randamentul pentru minimizarea utilizării îngrășămintelor și a pesticidelor [95].

Abordarea holistică în procesarea inteligentă a datelor în domeniul fiziologiei plantelor se referă la utilizarea tehnologiilor informatice pentru a analiza și înțelege procesele fiziologice ale plantelor într-un mod cuprinzător și integrat. Această abordare poate implica utilizarea unor tehnologii precum *Machine Learning*, analiză de imagini și senzori IoT (*Internet of Things – Internetul Lucrurilor*) pentru a colecta și procesa date despre plante [96-97]. Aceste date pot include informații despre creșterea plantelor, schimburile de gaze, absorbția de nutrienți și alte procese fiziologice. De asemenea, pot fi dobândite cunoștințe valoroase referitor la plantele invazive [98].

În scopul dezvoltării de modele predictive sau analize de înaltă precizie în domeniul fiziologiei plantelor pe baza datelor disponibile pot fi aplicate și tehnici de *Transfer Learning* [99]. În acest context, TL poate ajuta la adaptarea modelelor sau a cunoștințelor pre-învățate într-un domeniu sau specie specifică de plante pentru a îmbunătăți înțelegerea sau predicția fenomenelor legate de creșterea, dezvoltarea și starea plantelor [100]:

- *Predicția Stării de Sănătate a Plantelor:* Transfer Learning poate fi utilizat pentru antrenarea modelelor de învățare profundă, identificarea speciilor de plante, determinarea stării de sănătate a plantelor, recunoașterea bolilor pe baza imaginilor [101]. Cu ajutorul imaginilor, Transfer Learning poate fi utilizat pentru a construi modele care să clasifice speciile de plante, ceea ce este util în proiecte de conservare a biodiversității. Se pot folosi modele pre-antrenate pentru clasificarea plantelor și apoi se pot adapta pentru speciile sau condițiile specifice de studiu. De asemenea, datele spectrale obținute de la plante pot fi analizate pentru a determina nivelurile de clorofilă, absorbția de apă sau alte caracteristici legate de sănătatea plantelor. Imaginile sau datele obținute de la senzori pot fi analizate pentru a monitoriza evoluția plantelor și pentru a detecta eventuale probleme în timp real [102].
- *Predicția randamentului culturilor:* Utilizând date despre condițiile de mediu, tipurile de sol, variabile meteorologice și date anterioare despre recolte, Transfer Learning poate fi aplicat pentru a dezvolta modele de predicție a randamentului culturilor. Prin antrenarea modelelor pe datele istorice privind randamentele culturilor și condițiile de mediu, se poate realiza o estimare a randamentelor viitoare în funcție de factorii de mediu actuali.
- *Optimizarea utilizării resurselor de apă și nutrienți:* Prin aplicarea Transfer Learning, modelele pot fi antrenate să interpreteze date despre umiditatea solului, condițiile meteo și alte variabile pentru a dezvolta modele de gestionare a irigației și a nutriției plantelor, folosind cunoștințele anterioare despre necesitățile de apă și nutrienți ale plantelor.
- *Analiza datelor transcriptomice și proteomice:* În domeniul geneticii și biologiei moleculare a plantelor, Transfer Learning poate ajuta la identificarea genelor sau a proteinelor cu rol important în diverse procese fiziologice, utilizând cunoștințele dobândite în domenii similare.
- *Monitorizarea schimbărilor climatice și adaptarea la noi medii sau variații de mediu:*
Pentru a evalua impactul schimbărilor climatice asupra plantelor, TL poate fi aplicat pentru a analiza datele de satelit și imagini aeriene pentru a identifica modificările în vegetație și sănătatea plantelor. Datele rezultate pot fi folosite pentru a investiga modul în

care plantele se adaptează la condiții climatice în schimbare, utilizând date și modele de la specii sau regiuni similare. TL poate fi folosit pentru a ajuta plantele să se adapteze la noi condiții de mediu sau la variații de mediu neașteptate. Acestea sunt doar câteva exemple și există multe alte aplicații potențiale ale TL în fiziologia plantelor. Această tehnică poate fi deosebit de utilă atunci când seturile de date disponibile sunt limitate sau când se dorește o adaptare rapidă la condiții variabile de mediu [103-104].

Odată colectate și procesate, aceste date pot fi utilizate pentru a dezvolta modele predictive ale proceselor fiziologice ale plantelor și pentru a identifica factorii care afectează aceste procese. Acest lucru poate ajuta la dezvoltarea de tehnologii agricole mai eficiente și sustenabile, precum și la îmbunătățirea înțelegerii noastre asupra modului în care plantele funcționează și interacționează cu mediul înconjurător [105-111]. De asemenea, au fost dezvoltate modele matematice pentru a înțelege procesele fiziologice ale plantelor, cum ar fi fotosinteza și absorbția de nutrienți [112-113].

Au fost realizate cercetări care au dus la diverse lucrări, cu accent pe îmbunătățirea producției agricole și eficienței utilizării resurselor. Exemplele includ utilizarea senzorilor și a tehnologiilor de învățare automată pentru monitorizarea plantelor și luarea deciziilor referitoare la irigare și fertilizare, utilizarea imaginilor satelitare și a analizei imaginilor pentru evaluarea stării culturilor și identificarea problemelor de sănătate a plantelor, utilizarea roboticii și a automatizării pentru operațiuni agricole și utilizarea senzorilor și a tehnologiilor de monitorizare pentru evaluarea calității solului și identificarea problemelor de poluare sau degradare a solului [114-119]. De asemenea, s-a evidențiat utilizarea datelor de teledetecție pentru monitorizarea plantelor și identificarea factorilor care le influențează creșterea și dezvoltarea [44]. În plus, învățarea automată și învățarea profundă au fost utilizate pentru identificarea și cuantificarea stresului la plante [120-122].

Rețele neuronale artificiale au fost elaborate pentru a analiza imagini ale frunzelor plantelor de tutun, pentru a identifica bolile și stresul fiziologic. Fiind instruită să detecteze diferențele subtile în textură și culori ale frunzelor bolnave sau stresate, comparativ cu frunzele sănătoase, rețeaua neurală poate detecta cu precizie bolile și stresul fiziologic al plantelor, făcând posibilă monitorizarea lor timpurie și intervenția promptă [123]. De asemenea, au fost dezvoltate și modele matematice pentru a înțelege mai bine procesele fiziologice ale plantelor, cum ar fi procesul de fotosinteză, absorbția de nutrienți sau transpirația.

În ultimii ani, au fost realizate numeroase cercetări în domeniul procesării inteligente a datelor în domeniul fiziologiei plantelor. Aceste cercetări se concentrează asupra utilizării

tehnologiilor de învățare automată și a învățării profunde în domeniul agriculturii și al fenotipării plantelor. Au fost analizate diverse aspecte legate de monitorizarea, diagnosticarea și analiza plantelor și a culturilor agricole folosind tehnici de procesare a datelor și modele de învățare automată [124-133]. Lucrările acoperă diferite aspecte ale fenotipării plantelor, cum ar fi identificarea diferitor trăsături caracteristice, extragerea și clasificarea imaginilor, utilizarea viziunii stereo multi-vizuale pentru obținerea de reprezentări tridimensionale ale plantelor, analiza creșterii și dezvoltării plantelor prin viziune artificială bazată pe învățarea profundă și predicția randamentului culturilor agricole [102,134-137].

De asemenea, sunt abordate aspecte precum diagnosticarea bolilor plantelor folosind rețele neuronale convoluționale și clasificarea deficiențelor de nutrienți ale plantelor prin analiza imaginilor. Un alt aspect important este utilizarea tehnologiilor IoT pentru monitorizarea în timp real și prezicerea parametrilor de mediu care afectează producția de culturi [138-140]. Lucrările analizează avantajele și limitările utilizării tehnologiilor de învățare automată în agricultură și fenotiparea plantelor și propun direcții viitoare de cercetare pentru îmbunătățirea eficienței și acurateței acestor metode. În general, aceste lucrări subliniază potențialul și importanța tehnologiilor de învățare automată în dezvoltarea agriculturii de precizie și îmbunătățirea practicilor agricole [141-142].

Odată colectate și procesate, aceste date pot fi utilizate pentru a dezvolta modele predictive ale proceselor fiziologice ale plantelor și pentru a identifica factorii care afectează aceste procese [143-144]. Acest lucru poate ajuta la dezvoltarea de tehnologii agricole mai eficiente și sustenabile, precum și la îmbunătățirea înțelegerii modului în care plantele funcționează și interacționează cu mediul înconjurător.

1.3. Infrastructuri de Date și Cercetări Avansate

1.1.5. Baze de date

O provocare în cercetare este dimensionalitatea datelor: datele omice sunt de înaltă rezoluție sau, altfel spus, de dimensionalitate înaltă.

Bazele de date joacă un rol crucial în biologie și agricultură, oferind o sursă valoroasă de informații pentru cercetători, fermieri și alți profesioniști implicați în aceste domenii. Aceste baze de date sunt utilizate pentru a stoca, organiza și analiza o gamă largă de informații legate de gene, proteine, plante, animale și alte aspecte relevante.

În biologie, bazele de date sunt folosite pentru a gestiona și analiza secvențele genetice, informațiile despre gene și proteine, structura proteinelor, interacțiunile dintre molecule și multe altele. De exemplu, bazele de date precum *GenBank*, *UniProt* și *Protein Data Bank* (PDB) conțin

milioane de secvențe genetice și informații despre proteine, care pot fi căutate și analizate pentru a înțelege mai bine funcțiile biologice și relațiile dintre molecule.

În agricultură, bazele de date sunt utilizate pentru a gestiona informații despre plante, animale, soluri, condiții meteorologice și multe alte aspecte relevante pentru producția agricolă. Aceste baze de date pot ajuta la identificarea și selectarea culturilor rezistente la boli, la monitorizarea fertilității solului, la optimizarea utilizării resurselor și la analiza datelor de la ferme pentru a lua decizii mai informate. De exemplu, baza de date *The Plant List* furnizează informații despre nomenclatura plantelor [145], iar bazele de date meteorologice permit monitorizarea condițiilor climatice în diferite regiuni.

Pe lângă aceste exemple, există numeroase alte baze de date specifice care acoperă diverse aspecte ale biologiei și agriculturii. Acestea pot fi accesate de către cercetători, fermieri și alte persoane interesate pentru a obține informații, pentru a efectua analize și pentru a sprijini deciziile și descoperirile în aceste domenii. În continuare sunt descrise câteva baze de date relevante.

Baza de date globală privind speciile de plante invazive (GISD) [146] este o sursă gratuită de informații online, care poate fi căutată, despre speciile exotice și invazive care au un impact negativ asupra biodiversității își propune să crească gradul de conștientizare a populației cu privire la speciile invazive și să faciliteze activități eficiente de prevenire și management prin diseminarea cunoștințelor și experienței specialiștilor către un public larg la nivel mondial. Se concentrează pe speciile exotice invazive care amenință biodiversitatea nativă și zonele naturale și acoperă toate grupurile taxonomice, de la microorganisme la animale și plante [147-148].

Produsele cheie de cunoștințe includ Baza de date globală pentru specii invazive (GISD), Baza de date pentru biodiversitate și speciile invazive (IBIS) și Registrul Global al speciilor introduse și invazive (GRIIS) [149-151].

Baza de date GISD are patru componente informaționale cheie:

- *Clasificarea datelor de distribuție*, inclusiv starea biologică (proveniență, apariție, abundență și gradul de invazie) și tipul de introducere.
- *Clasificarea datelor de management*, inclusiv opțiunile de management (prevenire, control și eradicare), starea managementului și starea eradicării.
- *Clasificarea datelor de impact*, inclusiv mecanismele de impact și rezultatele impactului.
- *Clasificarea datelor căii*, inclusiv tipurile principale și subcategoriile de căi/vectori.

Utilizarea bazelor de date necesită cunoștințe speciale de utilizare, de aceea, au fost elaborate resurse accesibile pentru publicul larg, cum ar fi CABI Invasive Species Compendium

[148-149], care conține o descriere detaliată a speciilor invazive care amenință mijloacele de trai și mediul în întreaga lume.

În scopul conservării și utilizării durabile a biodiversității a fost creată *Rețeaua Interamericană De Informații Despre Biodiversitate* (IABIN) care este o rețea de baze de date dedicată adoptării și promovării standardelor și protocoalelor de ecoinformatică în toate țările din America [152]. Scopurile IABIN sunt:

- Îmbunătățirea infrastructurii pentru schimbul de informații despre biodiversitate.
- Consolidarea capacității tehnice de a face schimb de informații și expertiză privind biodiversitatea peste granițele politice și instituționale.
- Îmbunătățirea capacității de a stoca, utiliza și distribui informații științifice solide și actualizate despre biodiversitate.
- Elaborarea sau adaptarea produselor informaționale pentru luarea deciziilor în scopul elaborării politicilor publice de mediu și de dezvoltare durabilă în regiune [151,153].

Datele biologice sunt baza cercetării în științele vieții. Bioinformatica este știința care face posibilă colectarea, stocarea și adăugarea de valoare acestor date, astfel încât cercetătorii din multe domenii să le poată prelua și analiza eficient. Pentru realizarea cu succes a cercetărilor, cercetătorii de astăzi depind de accesul la seturi mari de date care acoperă diferite tipuri de date. În domeniul *big data* din bioinformatică sunt sistematizate la Institutul European de Bioinformatică parte din Laboratorul European de Biologie Moleculară (EMBL) [154], o organizație interguvernamentală de cercetare. Pe această platformă sunt disponibile servicii de bioinformatică care ajută oamenii de știință să realizeze potențialul *big data* în biologie, exploatând informații complexe pentru a face descoperiri care să beneficieze omenirea, menține o gamă largă și cuprinzătoare a resurselor de date moleculare din lume. De asemenea, oferă instruire în bioinformatică.

Serviciile web EMBL-EBI permit interogarea resursele date biologice pentru dezvoltarea de *conducte* de analiză a datelor sau integrarea datele publice cu propriile aplicații. Datele de la EMBL-EBI acoperă genomica, proteine, expresii, molecule mici, structuri de proteine, sisteme, ontologii și literatura științifică. menține cea mai cuprinzătoare gamă din lume de resurse de date moleculare disponibile gratuit și actualizate. EMBL-EBI găzduiește *Ensembl Genomes*, un serviciu care permite cercetătorilor să acceseze și să compare date la scară genomică de la specii relevante din punct de vedere agricol (figura 1.1).

Exemple de utilizare a **EMBL**: Înțelegerea genomului plantelor ajută la identificare speciilor care vor fi cele mai tolerante la condițiile de mediu (secetă, boli și dăunători), oferind totuși o nutriție optimă.

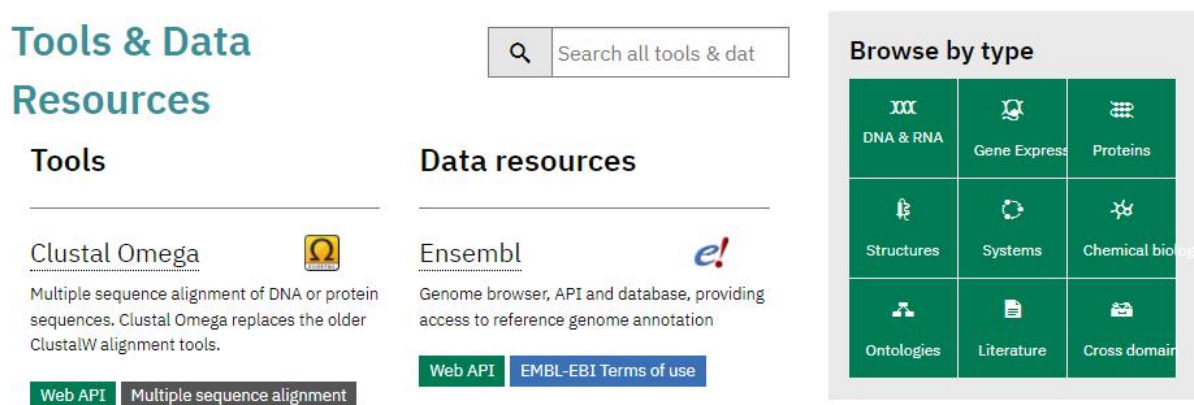


Figura 1.1. Resurse de date și instrumente EMBL-EBI

În domeniul agricol se desfășoară cercetări în diverse țări, utilizând tehnicile *Data Mining* aplicate asupra diferitor specii de plante agricole și implicit soia. Aceste cercetări se bazează pe soiurile și condițiile de sol și climatice din teritoriul respectiv. Una dintre aceste cercetări este realizată de către Jharna Majumdar, Sneha Naraseeyappa & Shilpa Ankalaki de la College of Engineering & Technology, Amravati, Maharashtra, India care au elaborat modelul *Soybean Productivity Modelling using Decision Tree Algorithms*. Acesta este un model de procesare a datelor, folosind sistemul *Waikato Environment for Knowledge Analysis* (WEKA) [153].

Pentru urmărirea dezvoltării plantelor de soia în timp real a fost elaborată aplicația SOYSIM (Soybean Growth Simulation Model) [155]. Modelul SoySim, elaborat de Tri D. Setiyono, Kenneth G. Cassman, James E. Specht, Albert Weiss, Achim Dobermann, Haishun Yang la Department of Agronomy and Horticulture, Institute of Agriculture and Natural Resources by University of Nebraska-Lincoln în 2010. Acest model este utilizat la fermele agricole din SUA, simulează creșterea soiei zilnic, simulează potențialul de producție de soia și utilizarea apei utilizarea nutrienților fără pierderi de producție din cauza factorilor abiotici și biotici. *SoySim* a obținut rezultate dezirabile chiar și cu parametri de intrare mult mai puțini specifici soiului în comparație cu alte modele. Pentru simularea potențialului de producție, modelul necesită date meteo zilnice pentru radiația solară și temperaturile maxime și minime. Pentru simularea utilizării apei și a recomandărilor de irigare, modelul necesită date meteo zilnice pentru radiația solară, temperatura maximă și minimă, precipitații, evapotranspirația de referință și umiditatea relativă a aerului [156].

1.1.6. Cercetări în cultivarea soia

Cultivarea soia este un subiect de cercetare important în domeniul agricol și a atras interesul cercetătorilor din întreaga lume. Soia este una dintre cele mai importante culturi din lume și are o gamă largă de utilizări, inclusiv ca sursă de hrană pentru animale și oameni, precum

și ca sursă de ulei și proteine. În ultimii ani, cercetătorii au abordat diferite aspecte ale cultivării soiei, inclusiv procesarea inteligentă a datelor pentru a spori randamentul culturilor și a reduce costurile.

Cercetările în domeniul cultivării soiei utilizează procesarea inteligentă a datelor pentru a identifica factorii cheie care influențează randamentul culturilor și pentru a dezvolta tehnologii agricole inovatoare care să maximizeze randamentul culturilor și să minimizeze costurile și impactul asupra mediului înconjurător [156]. Un exemplu de cercetare în acest domeniu este studiul efectelor utilizării diferitelor tipuri de îngrășăminte asupra randamentului soiei. Aceste studii folosesc diverse modele pentru a analiza datele colectate din terenuri agricole și a identifica combinații optime de îngrășăminte care pot îmbunătăți randamentul culturilor și pot reduce costurile pentru agricultorii din întreaga lume [157]. Un alt exemplu este utilizarea tehnologiilor de teledetecție pentru a monitoriza starea plantelor și a identifica factorii care influențează randamentul culturilor. Acest studiu a analizat datele colectate prin satelit și a utilizat algoritmi de învățare automată pentru a identifica factorii cheie care influențează randamentul soiei, cum ar fi nivelul de umiditate a solului, temperatura și nivelul de nutrienți.

Pentru studierea soiei au fost dezvoltate mai multe baze de date pentru analiza genomului de soia, cum ar fi SoyGD, SoyBase și SoyXpress. Aceste baze de date conțin o varietate de informații, cum ar fi secvențele genomului de soia, cromozomul artificial bacterian (BAC), etichetele de secvență exprimată (EST) și câteva instrumente utile, inclusiv browserele genomului, căutarea BLAST și căutarea căilor.

Baza de date SoyBase

SoyBase [158] este un depozit pentru genetică, genomică și resurse de date aferente pentru soia. Conține hărți actuale de secvențe genetice, fizice și genomice integrate cu trăsături calitative și cantitative. Conține informații genetice despre boabele de soia, cum ar fi hărți genetice și informații despre genetica mendeliană a soiei, date moleculare privind genele și secvențele de soia pe măsură ce acestea sunt disponibile. În 2010, a fost publicată secvența genomului de soia și ea și secvențele de gene de sprijin au fost integrate în browserul de secvențe *SoyBase*. Baza de date încorporează, de asemenea, informații despre mutanți și alte stocuri genetice de soia și servește ca punct de contact pentru comandarea tulpinilor din acele populații [158]. Modelele de expresie genetică sunt, de asemenea, disponibile la *SoyBase* prin paginile de expresie *SoyBase* și Atlasul Genelor Soia. Alte seturi de date de expresie / transcriptom / metilomice au fost și continuă să fie încorporate în browserul genomului *SoyBase*.

Baza de date a transportorilor de soia SoyDT

Transportorii (en.: *Transporter*) sunt o clasă de proteine membranare care facilitează schimbul de substanțe dizolvate, inclusiv diverse molecule și ioni de-a lungul membranei celulare, sunt componente vitale pentru supraviețuirea tuturor organismelor. Înțelegerea transportorilor de plante este importantă pentru a obține o perspectivă asupra proceselor celulare de bază, fiziologiei și mecanismelor moleculare, inclusiv absorbția nutrienților, semnalizarea, răspunsul la stresul extern și multe altele. În acest sens, s-a efectuat o analiză extinsă a transportorilor prevăzuți la soia și alte specii de plante. În plus, a fost dezvoltată o bază de date integrată pentru proteina transportoare de soia, *SoyTD* [159], care va facilita identificarea, clasificarea și caracterizarea extinsă a proteinelor transportoare prin integrarea expresiei, ontologiei genelor, domeniul și motivele conservate, organizarea structurii genelor, și caracteristicile de distribuție cromozomială. A fost efectuată o analiză cuprinzătoare pentru a identifica transportatorii foarte încrezători prin integrarea diferitelor instrumente de predicție. Aceasta a oferit perspective asupra expansiunii și duplicarea genelor transportoare în plantele terestre. *SoyTD* oferă o interfață cuprinzătoare pentru a studia funcția genetică și moleculară a transportorilor de soia, este disponibilă public la [160].

Baza de date Soybean-VCF2Genomes

Informațiile despre cel mai apropiat soi de soia vor permite crescătorilor să estimeze poziția germo-plasmatică relativă a eșantionului lor de interogare pentru a determina strategiile de ameliorare a soiei. Mai mult decât atât, schema *VCF2Genomes* poate fi extinsă la alte specii de plante în care secvențele întregii genom ale colecției de bază sunt disponibile public [161]. Cu toate acestea, aceste baze de date conțin doar adnotări generale pentru genomul de soia, în loc de cunoștințe care vizează în mod specific factorii de transcripție. Pentru a rezolva problema a fost dezvoltat *SoyDB* - o bază de date cuprinzătoare și integrată pentru factorii de transcripție din soia.

1.1.7. Depozite de date graf

Utilizarea depozitelor de date graf în domeniul fiziologiei plantelor este o direcție de cercetare relativ nouă, dar care prezintă deja rezultate promițătoare. Această abordare se bazează pe utilizarea rețelelor neuronale convoluționale și a datelor graf pentru analiza imaginilor plantelor, modelarea interacțiunilor și rețelelor metabolice ale plantelor, identificarea genelor implicate în dezvoltarea și răspunsul plantelor, clasificarea și identificarea plantelor pe baza caracteristicilor moleculare și morfologice, precum și înțelegerea rețelelor de semnalizare în

plante. *Rețelele Neuronale Convoluționale* (CNN) sunt un tip de arhitectură de rețea neuronală utilizată în principal în prelucrarea de imagini și recunoașterea de modele vizuale [162-165]. Ele au capacitatea de a extrage caracteristici relevante din imagini folosind straturi speciale numite straturi convoluționale. Cercetările realizate în acest domeniu evidențiază potențialul depozitelor de date graf în studiul fiziologiei plantelor și oferă exemple concrete de aplicare a acestora [166]. Exemplele includ utilizarea depozitelor de date graf în analiza imaginilor de plante, modelarea rețelelor metabolice, identificarea genelor reglementatoare, clasificarea plantelor și analiza datelor de fenotipare. Cercetătorii au abordat diverse aspecte legate de utilizarea datelor graf în domeniul fiziologiei plantelor, precum modelarea rețelelor metabolice, identificarea rețelelor de reglementare genică sau clasificarea plantelor. Aceste studii demonstrează implicarea cercetătorilor în dezvoltarea acestui domeniu și contribuția lor la avansarea cunoștințelor despre fiziologia plantelor. Câteva exemple sunt descrise mai jos.

Utilizarea rețelelor neuronale convoluționale (CNN) și a datelor graf în analiza imaginilor de plante [167-169]. Această metodă poate fi utilizată pentru a extrage caracteristici specifice ale plantelor, cum ar fi textura și forma frunzelor, și pentru a detecta și clasifica boli sau dăunători [170].

Datele graf modelează interacțiunile între plante și pentru a înțelege modul în care aceste interacțiuni influențează creșterea și dezvoltarea plantelor [171]. De exemplu, rețelele de interacțiune proteică pot fi utilizate pentru a identifica proteinele implicate în procesele de reglare a creșterii și dezvoltării plantelor [172].

Cu ajutorul grafurilor se pot modela rețelele metabolice ale plantelor și pentru a înțelege modul în care aceste rețele reglează procesele fiziologice. Această abordare poate fi utilizată pentru a identifica noi căi metabolice și pentru a descoperi noi compuși biochimici cu potențial terapeutic sau agronomic [173].

Rețelele de co-exprimare și a datelor graf pentru a identifica genele implicate în reglarea dezvoltării plantei și a răspunsului acesteia la stresul biotic și abiotic [174]. Această abordare poate fi utilizată pentru a descoperi noi gene cu potențial agronomic și pentru a dezvolta noi strategii de îmbunătățire a plantelor [175].

Datele graf și algoritmi de învățare automată clasifică și identifică plantele pe baza caracteristicilor lor moleculare și morfologice. Această abordare poate fi utilizată pentru a dezvolta tehnologii de monitorizare automată a plantelor și pentru a identifica speciile invazive sau periculoase [176].

Abordarea datelor graf pentru a modela rețelele de semnalizare în plante și pentru a înțelege modul în care acestea reglează procesele fiziologice [177]. Această abordare poate fi

utilizată pentru a identifica noi regulatoare și pentru a dezvolta noi strategii de îmbunătățire a plantelor prin modificarea semnalizării lor interne [174] și se concentrează pe utilizarea depozitului de date graf Neo4j în modelarea datelor de fenotipare a rădăcinilor plantelor, în vederea extragerii automate a trăsăturilor și analizei acestora [175].

O problemă esențială în agricultură este seceta. Aceasta este una serioasă și are impact asupra mediului, economiei și societății în general. Seceta se referă la o perioadă prelungită de lipsă a precipitațiilor sau la niveluri scăzute ale apei, care duce la uscarea solului și la scăderea disponibilității apei pentru diverse scopuri, cum ar fi agricultura, consumul uman și industrial.

Problema secetei implică câteva aspecte cheie legate de:

- a) *Impactul asupra agriculturii*: Seceta afectează producția agricolă, reducând recoltele și calitatea acestora. Culturile agricole au nevoie de o cantitate adecvată de apă pentru a crește și pentru a se dezvolta corespunzător. Lipsa apei poate duce la secarea plantelor, la pierderea recoltelor și la creșterea costurilor de irigație suplimentară.
- b) *Impactul asupra resurselor de apă dulce*: Seceta reduce nivelurile de apă în lacuri, râuri și bazine subterane, ducând la scăderea disponibilității apei dulci. Acest lucru afectează atât necesitățile umane de apă potabilă, cât și cele industriale și agricole. De asemenea, poate duce la intrarea în colaps a ecosistemelor acvatice și la scăderea biodiversității.
- c) *Impactul asupra economiei*: Seceta poate avea un impact semnificativ asupra economiilor, în special în zonele agricole. Reducerea recoltelor și a producției agricole poate duce la creșterea prețurilor alimentelor și la pierderi economice semnificative pentru fermieri. De asemenea, scăderea disponibilității apei poate afecta industria și serviciile care depind de aceasta.
- d) *Impactul asupra mediului*: Seceta poate afecta ecosistemele terestre și acvatice, ducând la scăderea biodiversității și la degradarea habitatelor naturale. Lipsa apei poate duce la moartea plantelor, la migrația forțată a animalelor și la pierderea speciilor.
- e) *Impactul asupra comunităților umane*: Seceta poate afecta direct viețile oamenilor, în special în comunitățile rurale și în țările în curs de dezvoltare. Lipsa apei potabile poate duce la probleme de sănătate și la creșterea incidenței bolilor legate de apă. De asemenea, seceta poate duce la migrația populațiilor, la conflicte sociale și la probleme socio-economice.

Pentru a aborda problema secetei, sunt necesare măsuri preventive și de gestionare adecvate, cum ar fi conservarea apei, tehnologii eficiente de irigație, gestionarea integrată a resurselor de apă, educație și conștientizare în privința utilizării responsabile a apei și adoptarea politicilor de adaptare la schimbările climatice. De asemenea, cercetarea și inovarea continuă

sunt esențiale pentru a dezvolta tehnologii și soluții sustenabile pentru a face față provocărilor asociate cu seceta.

Deși există acum mai multe studii științifice privind reducerea riscurilor de secetă decât în orice alt moment anterior, acestea joacă un rol crucial în fundamentarea diferitelor modele de predicție, algoritmi și formulări [176]. Cu toate acestea, aceste studii, care sunt destinate publicului larg, sunt destul de limitate și acoperă o arie geografică restrânsă care nu poate fi aplicate la alte regiuni [177]. Deși există proceduri puternice de prognoză numerică a vremii, rezultatele finale în ceea ce privește seceta, în special, sunt deosebit de importante pentru agricultură, gestionarea resurselor naturale, prevenirea dezastrelor și pregătirea acestora. Deciziile raționale pot fi luate pe baza unei interpretări adecvate a lucrărilor de modelare și prognoză, care susțin reducerea vulnerabilității la dezastre și gestionarea secetei cu riscuri și dezastre reduse [178].

1.4. Principalele lacune în implementarea tehnologiilor inteligente

Studiul și aplicarea tehnologiilor avansate în domeniul fiziologiei vegetale constituie domenii de cercetare în evoluție continuă. Implementarea agriculturii de precizie prezenta potențialul de a aduce multiple beneficii, însă și dificultăți și provocări semnificative. Această abordare se confruntă cu anumite aspecte care cere atenție pentru a realiza progrese semnificative:

1. *Costuri ridicate*: Implementarea tehnologiilor de precizie presupune investiții substanțiale în echipamente și software specializate, reprezentând o barieră în special pentru fermierii mici și mijlocii.
2. *Complexitate și limitări tehnologice*: Utilizarea tehnicilor de precizie necesită cunoștințe și abilități tehnice avansate. Fermierii trebuie să dobândească competența de a opera, întreține și interpreta corect datele generate de aceste tehnologii. Unele tehnologii și instrumente utilizate pentru colectarea și înregistrarea datelor fiziologice pot avea limitări tehnice inerente, care pot restricționa accesul la date sau impun constrângeri asupra volumului și complexității datelor.
3. *Acces la tehnologie*: lipsa infrastructurii necesare implementării agriculturii de precizie împiedică adoptarea acestor practici. Fermierii pot deveni dependenți de tehnologie, iar defecțiunile tehnice sau problemele de alimentare cu energie pot perturba activitățile agricole. În prezent, lipsa protocoalelor și normelor standardizate pentru studierea fiziologiei plantelor prin tehnologii inteligente împiedică compararea rezultatelor și reproductibilitatea experimentelor în diferite laboratoare sau instituții. Stabilirea de standarde și protocoale comune ar facilita colaborarea și schimbul de informații în

domeniul fiziologiei plantelor. *Reproductibilitatea* (R) reprezintă valoarea sub care diferența absolută dintre rezultatele unui singur test obținut în condiții de reproductibilitate (de exemplu, asupra materialelor identice obținute de operatori în laboratoare diferite, folosind metoda de test standard), este de așteptat să se mențină în limitele unei anumite probabilități (de obicei 95 %) [179].

4. *Reticența fermierilor și accesul la informații*: rezistența fermierilor la adoptarea de inovații, accesul insuficient la informații și formarea profesională adecvată în agricultura de precizie pot duce la disparități în performanță și rezultate. Formarea și educația continuă sunt necesare pentru ca fermierii să fie pregătiți în mod adecvat pentru a învăța și a-și dezvolta în mod constant abilitățile tehnologice, maximizând beneficiile tehnologiilor de precizie.
5. *Date insuficiente*: O provocare semnificativă în cercetarea fiziologiei plantelor este disponibilitatea datelor. În timp ce tehnologiile de monitorizare a instalațiilor, cum ar fi senzorii și dispozitivele de înregistrare, sunt din ce în ce mai folosite, datele colectate suferă adesea de limitări în ceea ce privește volumul, acoperirea spațială și intervalul temporal. Colectarea extinsă și sistematică a datelor este necesară pentru a înțelege mai bine procesele fiziologice ale plantelor și pentru a dezvolta modele predictive precise. Datele există sub formă fragmentată, acoperind regiuni specifice, specii de plante sau variații fenomice distincte. Extinderea domeniului de aplicare la alte specii sau soiuri poate da rezultate eronate.
6. *Calitatea datelor*: viabilitatea acestor metode se bazează în mare măsură pe calitatea datelor experimentale. Datele trebuie să fie exacte, cuprinzătoare și reprezentative pentru a produce rezultate valide și interpretabile. Calitatea slabă a datelor are un impact asupra agriculturii de precizie, care se bazează pe date precise și actualizate. Colectarea inexactă sau incompletă a datelor poate duce la decizii analitice eronate.
7. *Integrarea datelor*: fiziologia plantelor cuprinde un spectru larg, cuprinzând aspecte moleculare, celulare, tisulare și ale întregii plante. Integrarea datelor de la diferite niveluri și surse este imperativă pentru a obține o înțelegere holistică a fiziologiei plantelor. Dezvoltarea de metode și platforme pentru integrarea și analiza eficientă a datelor interdisciplinare complexe ar îmbunătăți înțelegerea proceselor fiziologice.
8. *Interpretarea datelor complexe*: Tehnologiile inteligente precum învățarea automată și analiza datelor oferă oportunități semnificative pentru analiza și interpretarea datelor fiziologice ale plantelor. Cu toate acestea, sunt necesare cercetări suplimentare pentru a dezvolta algoritmi și modele de învățare automată adaptate fiziologiei plantelor,

asigurând interpretarea corectă a datelor colectate. Analiza și interpretarea datelor fiziologice poate fi complicată din cauza varietății de variabile și a potențialelor interacțiuni.

9. *Confidențialitatea datelor*: Colectarea și stocarea datelor referitoare la câmpurile agricole și la practicile de cultivare ridică îngrijorări cu privire la securitate și confidențialitatea datelor, în special având în vedere creșterea atacurilor cibernetice.
10. *Generalizarea rezultatelor*: Unele cercetări pot fi limitate în capacitatea sa de a generaliza rezultatele în contexte mai largi. Rezultatele obținute într-un cadru experimental specific pot fi specifice speciei și pot varia în funcție de speciile de plante, soiuri, condițiile de mediu și alți factori specifici studiului.
11. *Implementare practică*: traducerea rezultatelor cercetării în aplicații practice reprezintă o altă provocare semnificativă. Utilizarea tehnologiilor inteligente în agricultură și industria alimentară necesită abordări practice și durabile care să permită integrarea perfectă în sistemele de producție existente. Dezvoltarea unor tehnologii ușor de utilizat, accesibile și eficiente din punct de vedere economic este esențială pentru implementarea lor pe scară largă.

Abordarea acestor probleme necesită un efort concertat din partea autorităților, organizațiilor agricole și a industriei tehnologice pentru a facilita adoptarea treptată și durabilă a agriculturii de precizie. Depășirea obstacolelor tehnice și economice este crucială. Cercetătorii trebuie să fie conștienți de aceste limitări și să le abordeze în mod corespunzător pentru a asigura valoarea și interpretarea corectă a rezultatelor obținute prin utilizarea tehnologiilor inteligente.

1.5. Concluzii la capitolul 1

Cercetările existente în utilizarea tehnologiilor inteligente în domeniul fiziologiei plantelor au adus în prim-plan o serie de beneficii și oportunități semnificative în domeniul agriculturii și a gestionării mediului înconjurător. O serie de tehnologii precum senzorii *IoT*, analiza datelor, inteligența artificială și roboții au fost integrate în studiile asupra fiziologiei plantelor, aducând avantaje majore pentru optimizarea creșterii și dezvoltării plantelor.

Printre principalele constatări ale cercetărilor se numără:

- a) *Monitorizarea precisă a condițiilor de mediu*: Tehnologiile inteligente permit o monitorizare continuă a factorilor de mediu precum lumină, umiditate, temperatură și niveluri de nutrienți. Acestea conduc la o înțelegere mai detaliată a nevoilor plantelor și ajută la ajustarea parametrilor pentru optimizarea creșterii.

- b) *Utilizarea eficientă a resurselor:* Integrarea senzorilor și a analizei datelor ajută la utilizarea mai eficientă a apei, fertilizanților și altor resurse, reducând astfel risipa de resurse și impactul asupra mediului.
- c) *Predictibilitate și intervenții proactive:* Tehnologiile inteligente permit obținerea unor modele de creștere și dezvoltare a plantelor bazate pe analize de date. Acestea permit fermierilor și cercetătorilor științifici să anticipateze unele probleme potențiale și să ia măsuri preventive pentru a minimiza impactul negativ asupra recoltelor.
- d) *Automatizarea proceselor:* Roboții și tehnologiile de automatizare pot fi utilizați pentru plantarea, recoltarea și alte operațiuni agricole. Acest lucru nu doar crește eficiența muncii, ci și reduce efortul uman necesar.
- e) *Avansarea cercetării științifice:* Utilizarea tehnologiilor inteligente permite obținerea de date mai detaliate și mai diverse cu privire la procesele fiziologice ale plantelor, contribuind la avansarea cunoștințelor în domeniile cercetate. Cercetările referitoare la utilizarea tehnologiilor inteligente în domeniul fiziologiei plantelor deschid perspective promițătoare pentru agricultură și sustenabilitate, permițând o gestionare mai precisă, eficientă și responsabilă a resurselor agricole și a ecosistemelor.

Abordarea holistică în procesarea inteligentă a datelor în domeniul fiziologiei plantelor poate aduce beneficii semnificative pentru agricultură și cercetările științifice, prin furnizarea de informații detaliate și relevante pentru optimizarea creșterii și producției plantelor.

Analiza literaturii existente demonstrează că procesarea inteligentă a datelor și depozitele de date sunt tehnologii importante în domeniul fiziologiei plantelor. Acestea pot fi utilizate cu succes pentru a colecta, analiza și interpreta datele în timp real, ceea ce poate ajuta la îmbunătățirea randamentului culturilor, reducerea costurilor și impactului asupra mediului și la optimizarea utilizării resurselor.

2. INTEGRAREA METODELOR BIOSTATISTICII ȘI A DEPOZITELOR DE DATE ÎN PROCESAREA DATELOR DIN DOMENIUL FIZIOLOGIEI PLANTELOR

2.1. Dezvoltarea bazei de cunoștințe

Fiziologia plantelor este o disciplină esențială în înțelegerea vieții plantelor și a ecosistemelor în care acestea trăiesc. Cu ajutorul tehnologiilor moderne și a analizei datelor, cercetătorii pot explora în profunzime procesele complexe care guvernează viața plantelor și pot contribui la dezvoltarea unor strategii mai eficiente pentru agricultură, conservarea biodiversității și gestionarea resurselor naturale. În ultimii ani, studiul fiziologiei plantelor a cunoscut o expansiune semnificativă, alimentată în mare parte de volumul enorm de date generat de cercetări avansate în domeniul biologiei moleculare, biologiei celulare și tehnologiilor de imagistică. Volumul crescut de date în domeniul fiziologiei plantelor a adus cu sine provocări și oportunități. Pe de o parte, cercetătorii au acces la un arsenal de date care pot fi utilizate pentru a înțelege mai bine modul în care plantele funcționează în detaliu. Pe de altă parte, gestionarea și analiza eficientă a acestor date devine o sarcină din ce în ce mai complexă. Pentru a face față acestor provocări, o combinație de tehnici de inteligență artificială, analiză de date și biologie computațională sunt folosite pentru a extrage informații semnificative din această "mare de date" și pentru a dezvolta modele predictive pentru comportamentul plantelor în diverse condiții.

Pentru a atinge obiectivele propuse, se propune dezvoltarea unei infrastructuri de cunoștințe pentru a fi utilizată într-un depozit de date. Acest proces implică integrarea metodelor biostatistice într-un cadru teoretic și practic coerent, adaptat în mod specific nevoilor și caracteristicilor domeniului fiziologiei plantelor. Depozitul de date va fi utilizat pentru stocarea și partajarea informațiilor experimentale în domeniul fiziologiei plantelor.

Dezvoltarea infrastructurii de cunoștințe necesită validarea și aplicarea acesteia într-un context practic. S-au utilizat date experimentale sau seturi de date existente pentru a evalua eficacitatea și utilitatea infrastructurii de cunoștințe în cercetările de fiziologie a plantelor. Au fost efectuate analize statistice relevante, iar concluziile privind contribuția și impactul infrastructurii de cunoștințe dezvoltate au fost trase.

La etapa actuală, caracterizată de progrese tehnologice fulminante și extinderea cunoașterii științifice, pentru dobândirea înțelepciunii prin înțelegerea profundă a fenomenelor devine imperativ să se investigheze și să se aprofundeze noțiunile și conceptele fundamentale ce stau la baza progresului tehnico-științific în diverse domenii.

Statistica este ramura științifică care se concentrează asupra procesului de colectare, organizare, analiză și interpretare numerică a datelor asociate fenomenelor de masă.

Biostatistica reprezintă disciplina care aplică metode statistice în contextul biologiei, având ca obiect de studiu analiza și interpretarea datelor biologice. Această abordare este esențială în înțelegerea și evaluarea rezultatelor provenite din experimente și studii în domeniul fiziologiei plantelor. Prin utilizarea adecvată a metodologiei statistice, se pot obține informații semnificative despre progresul fenomenelor fiziologice și despre interacțiunile dintre diversele variabile care influențează aceste fenomene.

Depozitele de date reprezintă o altă componentă esențială în dezvoltarea bazei de cunoștințe în domeniul fiziologiei plantelor. Acestea constau în colectarea și organizarea datelor într-un mod structurat și accesibil. Prin utilizarea depozitelor de date, cercetătorii pot împărtăși și accesa datele experimentale, facilitând astfel reproductibilitatea și colaborarea în cadrul comunității științifice. De asemenea, depozitele de date pot servi ca resurse valoroase pentru efectuarea analizelor statistice și dezvoltarea de modele predictive în domeniul fiziologiei plantelor.

Un *depozit de date* reprezintă un tip de sistem de gestionare a informațiilor conceput pentru a permite și susține activitățile de *Business Intelligence (BI)*, în special analiza datelor. Depozitele de date sunt dedicate exclusiv interogării și analizei și conțin cantități mari de date istorice. Datele dintr-un depozit de date provin în general dintr-o varietate largă de surse, precum fișierele jurnal ale aplicațiilor și aplicațiile de tranzacții. *Depozitul de date* centralizează și consolidează cantități mari de date provenite din multiple surse. Capacitățile sale analitice permit organizațiilor să obțină informații valoroase pentru afaceri din datele lor, în scopul îmbunătățirii procesului decizional. Pe parcursul timpului, depozitul de date construiește o înregistrare istorică care poate fi extrem de valoroasă pentru specialiștii în știința datelor și analiștii de afaceri [180].

O *bază de cunoștințe* este o colecție organizată de informații, date și cunoștințe structurate, accesibile și utilizabile într-un mod eficient. Aceste baze de cunoștințe sunt create pentru a stoca și gestiona informații relevante într-un anumit domeniu de expertiză sau pentru a rezolva anumite probleme [181]. Ele pot conține informații structurate într-o formă logică, cum ar fi date, fapte, reguli, relații, modele sau concepte. Bazele de cunoștințe, de asemenea, pot include și informații slab-structurate sau nestructurate, cum ar fi texte, documente sau fișiere media.

Bazele de cunoștințe pot fi utilizate pentru a realiza mai multe scopuri, cum ar fi:

- *Gestionarea informațiilor*: Ele pot ajuta la organizarea și stocarea eficientă a informațiilor într-un mod ușor de accesat și căutat. Astfel, utilizatorii pot găsi rapid informațiile de care au nevoie.
- *Luarea deciziilor*: Bazele de cunoștințe pot fi utilizate pentru a ajuta în procesul decizional prin furnizarea de informații relevante și sugestii. Ele pot fi programate să aplice reguli și algoritmi pentru a ajuta utilizatorii să ia decizii informate.
- *Automatizarea sarcinilor*: Prin utilizarea bazei de cunoștințe, anumite sarcini pot fi automatizate, deoarece sistemele pot accesa și utiliza informațiile stocate în bază pentru a efectua anumite operații sau procese.
- *Îmbunătățirea colaborării și comunicării*: O bază de cunoștințe centralizată poate facilita partajarea informațiilor între diferiți utilizatori și departamente, îmbunătățind astfel colaborarea și comunicarea în cadrul unei organizații.

Combinarea metodelor biostatistice și depozitelor de date graf poate fi extrem de utilă în cercetări de fiziologie a plantelor, deoarece permite analiza detaliată a datelor experimentale și identificarea modelelor sau relațiilor complexe între diferite variabile.

Metodele biostatistice sunt utilizate pentru a analiza și interpreta datele experimentale, identificând tendințe, efecte și diferențe semnificative. Ele pot fi utilizate pentru a realiza analize comparative între grupuri de plante expuse la diferite condiții sau tratamente, pentru a evalua eficacitatea unor anumite metode sau tratamente și pentru a identifica factorii de influență asupra performanței plantelor [182].

Depozitele de date graf sunt utilizate pentru a stoca, organiza și vizualiza datele experimentale. Ele pot conține informații despre diferite variabile, cum ar fi temperatura, umiditatea, nivelul de CO₂, luminozitatea și performanța plantelor. Utilizarea depozitelor de date graf poate facilita nu numai identificarea și vizualizarea relațiilor complexe între variabile, dar poate ajuta și la evidențierea tendințelor, oferind posibilitatea modelării datelor experimentale.

O componentă software esențială a depozitelor de date este motorul de inferență, care oferă funcționalități avansate de analiză și extragere de cunoștințe din grafurile stocate în baza de date. *Motorul de inferență Neo4j* este capabil să exploreze relațiile complexe dintre entități în graf și să deducă informații noi prin aplicarea regulilor și algoritmilor specifici. Acesta utilizează tehnici de raționament automat și logica deductivă pentru a găsi modele, structuri sau proprietăți ascunse în datele graf. Prin intermediul motorului de inferență *Neo4j*, este posibilă realizarea de inferențe și analize avansate, precum detectarea de tipare sau tendințe în rețele complexe, identificarea de relații nemenționate explicit în date, efectuarea de recomandări personalizate și generarea de noi cunoștințe din datele existente. Astfel, motorul de inferență *Neo4j* adaugă

valoare depozitelor de date prin extinderea capacităților de analiză și descoperire a informațiilor, permițând utilizatorilor să obțină o înțelegere mai profundă și să obțină rezultate mai relevante din datele grafice stocate în baza de date *Neo4j*.

Prin combinarea metodelor biostatistice și depozitelor de date graf se pot obține o înțelegere mai profundă a modului în care diferiți factori influențează procesele fiziologice ale plantelor. Aceasta poate include identificarea relațiilor cauzale între variabile, evaluarea eficacității unor anumite produse (îngrășăminte, substanțe biostimulatoare, pesticide etc.) și dezvoltarea de modele predictive pentru comportamentul plantelor în diferite condiții de existență.

2.2. Modelarea datelor

Modelarea datelor în domeniul fiziologiei plantelor implică reprezentarea structurată a informațiilor și a relațiilor dintre diferitele aspecte ale proceselor fiziologice din plante. Un limbaj formal este esențial pentru a asigura precizie, coerență și interoperabilitate în acest context. Limbajul formal se referă la un sistem de simboluri și reguli riguroase folosit pentru a comunica idei într-un mod precis și fără ambiguități. Acesta este utilizat într-o varietate de domenii, cum ar fi matematica, logică, știința calculatoarelor, lingvistica formală și altele, pentru a exprima concepte și relații într-un mod clar și coerent [183]. Un astfel de limbaj formal ar putea fi reprezentat de semnături și ecuații matematice, notări și simboluri specifice domeniului [184].

Soluționarea nevoilor din domeniile de activitate umane (de exemplu: agricultură) se realizează prin problematizarea diferitor chestiuni. Problema este o chestiune în care, pe baza unor ipoteze, se cere să se determine, prin calcule sau prin raționamente, anumite date care constituie soluția problemei. „Problema poate avea o singură soluție, mai multe soluții, o infinitate de soluții sau nicio soluție” [14,55].

În cercetarea biologică apar o serie de probleme a căror rezolvare necesită antrenarea noilor tehnologii, cum ar fi sistemele de procesare a datelor ne-relationale și inteligența artificială. Sarcina inițială, din punct de vedere conceptual, este de a descrie comportamentul de-a lungul timpului al unui sistem biologic. Dificultățile întâmpinate sunt recunoașterea variabilelor de stare, determinarea relațiilor cauză-efect. Ne confruntăm apoi cu problemele constatării validității diverselor ipoteze pe baza comparațiilor rezultatelor teoretice și experimentale (analiza statistică). Analizele de sensibilitate ale rezultatelor pot fi utilizate pentru a ghida experimentarea. În toate acestea, există noi oportunități de cercetare oferite de tehnologiile inteligente. Începând cu aceste procese descriptive, care pot fi deterministe sau stocastice, apelăm la procesele de control. Cum reacționează plantele și părțile acestora la

tratarea cu anumite preparate sau tratarea termică. Obiectivul este acela de a îmbunătăți preparatele pentru sporirea productivității și adaptarea plantelor la noile condiții climatice. O altă problemă este determinarea gradului de control atunci când mecanismele de bază sunt mai puțin înțelese. Spre exemplu, anumite preparate influențează anumite soiuri și anumite părți ale plantei într-o măsură mai mare decât altele.

În luarea deciziilor, trebuie să rezolvăm probleme de natură eterogenă, rezolvarea individuală a cărora necesită resurse semnificative și nu va oferi soluția potrivită în ansamblu, holistic. În același timp, eficiența și eficacitatea unui preparat nu vor putea fi stabilite dacă se analizează doar anumite aspecte, spre exemplu, doar parametrii morfologici sau activitatea unor enzime, fără a corela aceste date cu rezistența genotipurilor la secetă și cu productivitatea plantelor [185].

Pentru a realiza obiectivele trasate și a lua deciziile care se impun, a fost elaborat un sistem de procesare inteligentă a informației bazat pe *familii de probleme* (FP) [55,185]. Modelarea și programarea pe familii de probleme este o nouă paradigmă de dezvoltare a produselor program. Această paradigmă are avantajul de a standardiza arhitectura și tehnologia de proiectare, condițiile de exploatare și sporirea productivității elaborării sistemelor informaționale.

Familia de probleme reprezintă un concept din domeniul informaticii și al științei calculatoarelor care se referă la un grup de probleme sau sarcini similare sau legate între ele, care împărtășesc anumite caracteristici și trăsături comune. Aceste probleme sunt adesea rezolvate utilizând aceeași abordare, metodologie sau tehnică, și pot fi grupate într-o categorie sau clasă în funcție de aceste asemănări. O familie de probleme poate varia în complexitate și domeniu de aplicare. Aceasta poate cuprinde de la probleme relativ simple până la probleme mai complexe. Familia de probleme oferă un cadru util pentru a organiza și clasifica provocările din domeniul informaticii, permițând dezvoltatorilor să creeze soluții mai generale și reutilizabile pentru problemele similare din aceeași categorie.

Dezvoltarea aplicațiilor orientate pe familii de probleme constă în proiectarea, dezvoltarea și implementarea unor soluții software modulare și reutilizabile, având ca scop abordarea eficientă a unui spectru divers de probleme similare sau interconectate, în cadrul unui domeniu dat. Această abordare impune structurarea aplicației în componente funcționale și module interoperabile, pentru a facilita gestionarea și extinderea ulterioară.

Abordarea OFP oferă utilizatorilor un set de instrumente sau funcționalități versatile pentru a face față acestor probleme într-un mod eficient și coerent. Această abordare poate aduce

beneficii semnificative în ceea ce privește eficiența dezvoltării software, managementul proiectelor și experiența utilizatorilor.

În lumea complexă a învățării automate, abordarea tradițională presupunea antrenarea modelelor de la zero pentru fiecare problemă specifică. Cu toate acestea, o revoluție în domeniu a adus cu sine o paradigmă nouă și eficientă - transferul de învățare (Transfer Learning – TL). Această abordare remarcabilă a schimbat modul în care abordăm o diversitate de probleme, permițându-ne să extragem cunoștințe dobândite în rezolvarea unor familii de probleme similare și să le aplicăm cu succes în contexte specifice. Astfel, TL se dovedește a fi un instrument inestimabil în accelerarea și îmbunătățirea performanței modelelor de învățare automată într-o varietate de domenii. În această lucrare, sunt explorate implicațiile și beneficiile acestei tehnici, analizând modul în care familii de probleme pot beneficia de această abordare, evidențiind avantajele aduse de adaptarea cunoștințelor între contexte similare.

TL este o tehnică în învățarea automată care implică utilizarea cunoștințelor dobândite în rezolvarea unei probleme pentru a rezolva o altă problemă similară. În cazul transferului de învățare pe familii de probleme, ideea este aplicarea cunoștințelor și experiența acumulate în rezolvarea unui set de probleme în cadrul unui domeniu mai larg de probleme similare:

1. *Selectarea unei familii de probleme:* Se alege un set de probleme care au caracteristici și structuri similare. De exemplu, detectarea bolilor la plante, predicția randamentului culturilor, optimizarea utilizării resurselor, controlul dăunătorilor.
2. *Alegerea modelului de bază:* Folosește un model de învățare profundă pre-antrenat pe o problemă similară. Acesta va fi baza care va fi folosită pentru a începe antrenarea pe problema specifică cercetată.
3. *Fine-tuning (reantrenarea):* Modelul de bază se ajustează folosind datele specifice problemei. Poate fi necesar de adăugat un strat final pentru a se potrivi cu numărul de clase sau cu cerințele specifice ale problemei.
4. *Antrenarea pe setul de date specific:* Modelul se antrenează pe setul de date nou, folosind tehnici de optimizare și validare.
5. *Evaluarea performanțelor:* Evaluează performanța modelului pe un set de date de testare. Acest lucru va ajuta la evaluarea potrivirii modelului pentru problema specifică studiată.
6. *Adaptarea și optimizarea:* Dacă performanțele nu sunt satisfăcătoare, se pot face ajustări suplimentare în procesul de *fine-tuning* sau se utilizează alte tehnici de optimizare.

Beneficiile utilizării transferului de învățare pe familii de probleme includ:

1. *Reducerea costurilor de antrenare*: Pornind de la un model pre-antrenat, economisești timp și resurse.
2. *Îmbunătățirea generalizării*: Datorită antrenamentului pe un set mai larg de date, modelul are tendința de a generaliza mai bine pentru problemele specifice.
3. *Posibilitatea de a lucra cu seturi de date mici*: Dacă ai un set de date limitat, transferul de învățare te poate ajuta să obții performanțe bune.

Totuși, dacă diferențele dintre probleme sunt prea mari, transferul de învățare poate fi inefficient. Este posibil să fie necesare ajustări suplimentare în structura modelului sau în tehnici de fine-tuning.

Dezvoltarea aplicațiilor orientate pe FP urmează anumiți pași cheie:

4. *Identificarea Contextului și Definirea Domeniului*: Delimitează domeniul de aplicare și identifică o gamă de probleme sau sarcini relevante ce fac parte dintr-o familie specifică de probleme.
5. *Modelarea Cerințelor*: Modelează cerințele individuale ale fiecărei probleme din familie, evidențiind caracteristicile unice și cele comune. Utilizează tehnici formale pentru a specifica cerințele într-un limbaj precis și detaliat.
6. *Proiectarea Modulară*: Dezvoltă o arhitectură modulară pentru aplicație, împărțind funcționalitățile în module independente. Fiecare modul ar trebui să se concentreze pe abordarea unei anumite probleme sau a unui subset de probleme.
7. *Identificarea Interconexiunilor*: Identifică relațiile și interacțiunile dintre module, astfel încât să poți asigura o coerență și integrare eficientă între componente pentru a rezolva probleme multiple.
8. *Design Reutilizabil*: Adoptă principii de design reutilizabil pentru a dezvolta componente software generice și flexibile, care pot fi utilizate în contexte variate din familia de probleme.
9. *Implementare și Testare*: Dezvoltă modulele conform specificațiilor și cerințelor stabilite. Efectuează teste riguroase pentru fiecare componentă și pentru interacțiunile dintre ele, pentru a asigura funcționarea corectă și eficientă a aplicației.
10. *Interfață Utilizator Unificată*: Dezvoltă o interfață utilizator coerentă, care să ofere o experiență consistentă utilizatorilor pentru abordarea diferitelor probleme din familie.
11. *Gestionarea Datelor*: Proiectează o strategie eficientă de gestionare a datelor, asigurând sincronizarea și partajarea adecvată a informațiilor între modulele aplicației.

12. *Documentație Riguroasă*: Furnizează documentație detaliată, inclusiv specificații formale, diagrame și ghiduri pentru dezvoltatori și utilizatori, pentru a asigura înțelegerea corectă a funcționalităților și a interacțiunilor aplicației.
13. *Evoluție și Îmbunătățiri Continue*: Monitorizează performanța aplicației în timp, colectează feedback de la utilizatori și implementează actualizări și îmbunătățiri în mod regulat pentru a menține aplicația relevantă și eficientă în contextul evoluției cerințelor și tehnologiilor.

Pentru rezolvarea problemelor abordate în cercetarea curentă a fost dezvoltat un limbaj formal pentru modelarea datelor biologice.

Aplicația *OFP* reprezintă o aplicație generică (*abstractă*) în raport cu aplicațiile $A_1, A_2, \dots, A_n$, iar aplicațiile $A_1, A_2, \dots, A_n$ reprezintă n aplicații specifice în raport cu aplicația generică A .

În cazul, când mulțimea problemelor $P_1, P_2, \dots, P_n$ poate fi privită ca o familie de probleme, poate fi elaborată o singură aplicație A , care realizează pe calculator familia de probleme.

Domeniul de cercetare „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor” (abreviat *MBDFP*) integrează o gamă de tipuri de entități.

Limbajul formal *MBDFP*, care descrie soluțiile unor probleme din aceste domeniu de cercetare, solicitate de cercetători științifici specializați în biologia plantelor și de către agricultori, este următorul:

$\langle \text{limbajul MBDFP} \rangle ::= \langle \text{entitate studiată} \rangle \mid \langle \text{relație între tipuri de entități} \rangle \mid \langle \text{definire noduri și relații dintre noduri} \rangle \mid \langle \text{relație între noduri} \rangle \mid \langle \text{familia de probleme MBDFP} \rangle$

$\langle \text{entitate studiată} \rangle ::= \langle \text{proprietățile obiectelor studiate} \rangle \mid \langle \text{tipuri de soluri} \rangle \mid \langle \text{fermă} \rangle \mid \langle \text{zonă geografică} \rangle \mid \langle \text{resursă de apă} \rangle \mid \langle \text{umiditatea solului} \rangle \mid \langle \text{cantitatea de precipitații} \rangle \mid \langle \text{temperatura aerului} \rangle \mid \langle \text{coeficient hidrotermic} \rangle \mid \langle \text{viteza vântului} \rangle \mid \langle \text{prognoză meteorologică} \rangle \mid \langle \text{producător} \rangle \mid \langle \text{pesticide și îngrășăminte} \rangle \mid \langle \text{cercetători și instituția de cercetare} \rangle \mid \langle \text{echipament agricol} \rangle \mid \langle \text{vânzare și tranzacție} \rangle$

$\langle \text{relație între tipuri de entități} \rangle ::= \langle \text{relație între culturi și ferme} \rangle \mid \langle \text{relație între culturi și tipuri de soluri} \rangle \mid \langle \text{relație între ferme și culturi} \rangle \mid \langle \text{relație între ferme și resurse de apă} \rangle \mid \langle \text{relație între ferme și zone geografice} \rangle \mid \langle \text{relație între ferme și resurse de apă} \rangle \mid \langle \text{relație între zone geografice și resurse de apă} \rangle \mid \langle \text{relație între zone geografice și cantitatea de precipitații} \rangle \mid \langle \text{relație între zone geografice și temperatura aerului} \rangle \mid \langle \text{relație între zone geografice și viteza vântului} \rangle \mid \langle \text{relație între culturi și pesticide/îngrășăminte} \rangle \mid \langle \text{relație între soiuri și calitatea recoltei} \rangle \mid \langle \text{relație între ferme și producători de produse pentru plante} \rangle \mid \langle \text{relație între culturi și cercetători/instituții de cercetare} \rangle \mid \langle \text{relație între ferme și echipamente agricole} \rangle \mid$

<relație între ferme și vânzări/tranzacții> | <relație între culturi și prognoze meteorologice>
 <definire nod și relație dintre noduri> ::= <nod cultură> | <nod tip de soi> | <nod fermă> |
 <nod zonă geografică> | <nod resursă de apă> | <nod umiditatea solului> | <noduri de
 cantitatea de precipitații> | <nod temperatura aerului> | <nod de viteză a vântului>
 <relație între noduri> ::= | <cultivat în> | <potrivit pentru> | <se află în> | <disponibil
 între nodurile de ferme și resurse de apă disponibile> | <disponibil între nodurile de
 zone geografice și resurse de apă disponibile> | <înregistrează între nodurile de zone
 geografice și cantitatea de precipitații> | <înregistrează între nodurile de zone
 geografice și temperatura aerului pentru> | <înregistrează între nodurile de zone
 geografice și viteza vântului>
 <familia de probleme MBDFP> ::= <analiza statistică: conținutul de azot și fosfor în
 plante> | <analiza statistică: influența unor substanțe biologice active asupra creșterii
 și dezvoltării plantelor de soia cultivate pe medii nutritive apoase knop> | **Ipoteza
 nulă**
 <analiza statistică: conținutul de azot și fosfor în plante> ::= <soi> | <tratament> |
 <conținutul de azot (N_2)> | <variația conținutului de azot (N_2)> | <influența
 tratamentului asupra conținutului de N_2 > | <conținutul de proteină> | <variația
 conținutului de proteine> | <conținutul de fosfor (P_2)> | <variația conținutului de
 fosfor (P_2)> | <conținutul de pentaoxid de fosfor (P_2O_5)> | <variația conținutului de
 pentaoxid de fosfor (P_2O_5)> | <influența tratamentului asupra lungimii rădăcinii> |
 <volumul rădăcinii (ml)> | <biomasa rădăcinii proaspete> | <influența tratamentului
 asupra biomasei proaspete a rădăcinii> | <biomasa rădăcinii uscate> | <scurgerea
 electroliților> | <lungimea rădăcinii> | <biomasa rădăcinii proaspete>
 <analiza statistică ::= influența unor substanțe biologice active asupra creșterii și dezvoltării
 plantelor de soia cultivate pe medii nutritive apoase knop>
 <soi> ::= <Clavera> | <Dorința> | <Horboveanca> | <Colina>
 <tratament> ::= <Knop> | <Knop + Heteroauxină> | <Knop + Reglalg> | <Knop +
 Biovit>
 <cercetare MBDFP> ::= <familie probleme 1> | <familie probleme 2> | ... |
 <familie probleme n>
 <familie probleme 1> ::= **Analiză statistică influență** <tratament>
 [Conținut | Variație] <indicator calitate> <soi>

14. <tratament> ::= <substanță biologic activă> |

15. <substanță biologic activă> + <substanță biologic activă>

<substanță biologic activă> ::= **Heteroauxină** | **Reglalg** | **Biovit** | **Martor**

<indicator calitate> ::= **Conținut** <nutrient> | **Lungime rădăcină** | **Biomasă rădăcină proaspătă** | **Biomasă rădăcină uscată** | **Variație biomasă rădăcină uscată** | **Biomasă tulpină proaspătă** | **Variație biomasă tulpină proaspătă** | **Volum rădăcină** | **Variație volum rădăcină** | **Variație conținut apă** | **Variație masă făină proaspătă** | **Variație conținut ulei** | **Volum rădăcină** | **Variație volum rădăcină** | **Nr. total boabe** | **Variație Nr. total boabe** | ...

<nutrient> ::= **Azot** | **Proteină** | **Fosfor** | **Pentaoxid fosfor** | **Natriu** (sodiu, Na) | **Potasiu** (K) | **Clor** (Cl) | **Calciu** (Ca) | **Magneziu** (Mg) | **Fier seric** (Fe) | **Fosfat** | **Bicarbonat** | **Apă** | **Ulei** | **Masa făină proaspătă** | **Masă făină uscată** | ...

<soi> ::= **Clavera** | **Dorința** | **Horboveanca** | **Colina**

<nod> ::= <substanță biologic activă> | <soi> | <fermă> | <zonă geografică> | <resursă apă> | <cantitate precipitații> | <temperatură aer> | <coeficient hidrotermic> | <viteză vânt> | <prognoză meteorologică> | <producător> | <pesticid> | <îngrășământ> | <cercetător> | <instituție cercetare> | <echipament agricol>

<relație între noduri> ::= **Cultivat în** (<nod>) | **Potrivit pentru** (<nod>, <resursă apă>) | **Se află în** (<nod>, <nod>) | **Disponibil între** (<fermă>, <resursă apă>) | **Disponibil între** (<zonă geografică>, <resursă apă>) | **Înregistrează** (<zonă geografică>, <cantitatea de precipitații>) | **Înregistrează** (<zonă geografică>, <temperatura aerului>) | **Înregistrează** (<zonă geografică>, <viteza vântului>)

2.3. Modelarea biostatistică a datelor din domeniul fiziologiei plantelor

Utilizarea metodelor biostatistice și depozitelor de date graf în cercetarea fiziologiei plantelor poate aduce numeroase beneficii și poate contribui la înțelegerea mai profundă a proceselor și mecanismelor fiziologice ale plantelor. Aceste metode sunt extrem de eficiente pentru:

1. *Analiza datelor experimentale*: Metodele biostatistice pot fi utilizate pentru a analiza și interpreta datele experimentale obținute din studiile fiziologice. Aceasta poate implica utilizarea testelor statistice pentru a evalua diferențele semnificative între grupuri de plante expuse la diferite tratamente sau condiții. De exemplu, se pot utiliza testele *t Student* sau analiza varianței (ANOVA) pentru a compara nivelurile de performanță sau răspunsurile fiziologice ale plantelor în diferite condiții.

2. *Identificarea modelelor și relațiilor:* Prin utilizarea metodelor biostatistice, cercetătorii pot identifica modele sau relații complexe între diferite variabile fiziologice. Acestea pot include relații cauzale între factori de mediu și răspunsurile fiziologice ale plantelor sau modele matematice care descriu schimbările în timp ale unor procese fiziologice specifice. Depozitele de date graf pot fi utilizate pentru a vizualiza aceste modele și relații, facilitând înțelegerea și interpretarea rezultatelor.
3. *Optimizarea și predictibilitatea:* Utilizând metodele biostatistice și depozitele de date graf, cercetătorii pot dezvolta modele și algoritmi pentru a optimiza procesele fiziologice ale plantelor sau pentru a prezice comportamentul acestora în diferite condiții. De exemplu, se pot utiliza metodele de regresie pentru a dezvolta modele predictive pentru creșterea plantelor sau pentru a identifica combinații optime de factori de mediu (cum ar fi lumină, temperatură, umiditate) pentru a obține producții maxime.
4. *Analiza datelor de teren sau de observație:* Metodele biostatistice pot fi aplicate și în analiza datelor de teren sau de observație, cum ar fi studiile ecologice sau monitorizarea plantelor în habitatul lor natural. Aceasta poate implica analiza distribuției spațiale a plantelor, evaluarea influenței factorilor de mediu asupra distribuției și abundenței speciilor sau identificarea tendințelor și schimbărilor în timp ale parametrilor fiziologici.
5. Domeniul statisticii include studierea și utilizarea teoriilor și tehnicilor de analiză a datelor care rezultă din procese sau fenomene naturale. Statistica include unele dintre cele mai de bază metode și tehnici ale metodei științifice:
 - Formarea ipotezelor;
 - Proiectarea experimentelor și a studiilor observaționale derivate din date;
 - Prezentare generală a datelor;
 - *Realizarea de inferențe de date* (de exemplu, testarea ipotezelor).

O ramură a aplicațiilor statistice este *biostatistica* care este axată pe științele vieții. În domeniul agricol *statisticile agricole* joacă un rol important în dezvoltarea durabilă a economiei [186-188]. Biostatistica ne permite să înțelegem dacă diferențele pe care le observăm în sistemele biologice sunt aleatorii (rezultatele se datorează întâmplării) sau reflectă diferențe semnificative care trebuie luate în considerare. În același mod, se pot crea modele pentru a prezice comportamentul unei variabile, de exemplu, folosind corelații. Pe lângă studiul asocierii dintre variabile, tehnicile statistice pot fi utilizate și pentru determinarea diferențelor dintre grupuri: un grup de referință (martor) și grupuri independente asupra cărora s-au aplicat anumiți factori, și permit evaluarea efectelor unei variabile independente asupra unei variabile

dependente. Pentru a afla gradul de influență asupra parametrilor cercetați se realizează teste statistice care reprezintă metode de decizie care ne ajută la validarea sau invalidarea cu un anumit grad de siguranță a unei ipoteze statistice. Aceste teste verifică veridicitatea unor ipoteze (inferențe statistice) [189].

În teză sunt expuse modalități de analiză a datelor din domeniul fiziologiei plantelor cultivate pe diferite medii nutritive și suplinite cu preparate care au ca scop sporirea calității și productivității plantelor de soia și adaptarea la diverse condiții de mediu pentru fiecare genotip studiat, de asemenea analiza influenței biostimulatorilor asupra plantelor și rezultatul obținut.

Obiectivele cercetărilor sunt găsirea interdependențelor netriviiale între datele genetice și fiziologice în funcție de diverși factori cum ar fi influența unor compuși cu activitate biologică, nutriție, umiditate, temperatură, precum și impactul acestora asupra agriculturii Republicii Moldova. Sunt analizați parametri morfologici a plantelor: lungimea și volumul rădăcinii, lungimea hipocotilului (partea plantei de la rădăcină până la cotiledon care este o frunză cu rezerve nutritive a embrionului plantelor cu semințe cu rolul de a hrăni planta imediat după încolțire) și epicotilului (partea plantei de la cotiledon până la prima frunză), lungimea internodurilor, înălțimea totală a plantei (fig. 2.1). Parametrii biometrici au fost corelați cu indicii cantitativi ai biomasei proaspete și uscate a rădăcinii și a tulpinii.

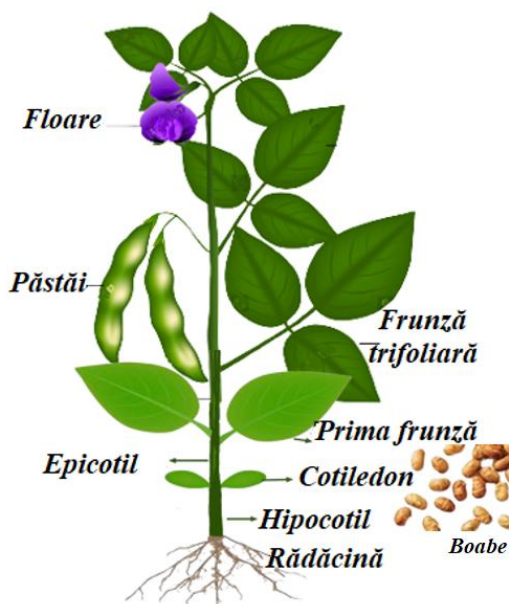


Figura 2.1. Structura plantei

Există mai multe metode și tehnici statistice utilizate în modelarea biostatistică a datelor în domeniul fiziologiei plantelor. Selectarea tipului de test statistic depinde de natura datelor și a

ipotezelor de cercetare pe care le dorim să le testăm [189-190]. Există diferite tipuri de teste statistice, iar alegerea potrivită depinde de următoarele aspecte:

- *Tipul datelor*: Dacă datele sunt continue sau discrete, selectăm un test statistic care este adecvat pentru tipul de date pe care îl avem. De exemplu, testele t și testele *ANOVA* sunt utilizate pentru date continue, în timp ce testele Chi-pătrat sunt utilizate pentru date nominale sau ordonate.
- *Distribuția datelor*: În unele situații, testele statistice presupun că datele urmează o anumită distribuție, cum ar fi distribuția normală. În funcție de acest lucru, vom selecta un test statistic care este robust la distribuția datelor noastre sau care face anumite asumții despre aceasta.
- *Scopul testului*: Există diferite ipoteze de cercetare pe care le putem testa, cum ar fi diferența între medii, asociații între variabile, diferențe între proporții etc. Alegerea testului statistic va fi determinată de ipoteza specifică pe care dorim să o testăm.
- *Dimensiunea și structura eșantionului*: Unele teste statistice sunt mai potrivite pentru eșantioane mici sau mari, în timp ce altele sunt concepute pentru date cuplate sau necuplate. De exemplu, testele t de student pentru eșantioane independente sunt utilizate atunci când avem două grupuri distincte, în timp ce testele t de student pentru eșantioane cuplate sunt utilizate atunci când avem aceeași populație măsurată în două condiții diferite.
- *Natura ipotezei*: În funcție de ipoteza nulă și alternativă, se selectează un test statistic corespunzător. De exemplu, testul t de student este utilizat pentru a testa dacă există o diferență semnificativă între medii, în timp ce testul χ^2 este utilizat pentru a testa independența între două variabile nominale.

Datele experimentale au fost determinate prin eșantioane mici și având de determinat diferența semnificativă între mediile grupurilor suplinate cu biostimulatori și grupurile de control (martor), s-a utilizat testul t (*Student*). Cu ajutorul acestui test se testează dacă o medie a unui set de date obținute în urma unor măsurări pe un eșantion de plante este compatibilă cu o valoare teoretică prestabilită (t_{teor}) [191]. În dependență de tipurile eșantioanelor studiate se alege tipul de test:

- Test independent – atunci când se compară un lot (grup) experimental asupra căruia s-au aplicat anumiți factori cu lotul martor.
- Test dependent – se analizează datele înainte de aplicarea unor factori și după aplicarea acestora, în scopul determinării efectelor acestor factori asupra plantelor.

După numărul de cozi există două tipuri de teste t :

- Testul t cu o coadă (one-tailed t-test);
- Testul t cu două cozi (two-tailed t-test).

Testul t cu o coadă și testul t cu două cozi sunt două tipuri diferite de teste statistice utilizate pentru a evalua diferențe semnificative între medii în două grupuri de date.

Testul t cu o coadă:

Acest test este folosit atunci când avem o ipoteză specifică cu privire la direcția diferenței între grupuri. Exemplu: Vrem să vedem dacă un anumit tratament face ca rezultatele să fie mai bune, dar nu ne interesează dacă ar putea face rezultatele mai proaste.

Testul t cu două cozi:

Acest test este folosit atunci când ne interesează dacă există o diferență semnificativă între grupuri, fără a avea o direcție specifică în minte. Exemplu: Vrem să vedem dacă un tratament are vreun efect semnificativ asupra rezultatelor, indiferent dacă le îmbunătățește sau le înrăutățește.

Deoarece avem eșantioane independente s-a folosit testul t pentru eșantioane independente și o coadă, deoarece vrem să vedem dacă utilizarea preparatelor îmbunătățește creșterea plantelor [192]. Atunci când există dispersii diferite ($\sigma_1^2 \neq \sigma_2^2$), valorile comparate nu alcătuiesc o pereche de valori corelate. Se analizează două grupe: un grup experimental (plantele sau mediul de dezvoltare este suplimentat cu anumite preparate) și un grup de control (martor). Pentru probe independente utilizăm formula pentru testul t :

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; \text{ unde:} \quad (2.1)$$

$\bar{x}_1$ - este media aritmetică a primului eșantion;

$\bar{x}_2$ - media aritmetică a celui de-al doilea eșantion;

σ_1^2 - dispersia primului eșantion;

σ_2^2 - dispersia celui de-al doilea eșantion;

n_1 - volumul primului eșantion;

n_2 - volumul celui de-al doilea eșantion.

Această statistică are o distribuție t . Distribuția " t " sau distribuția *Student* este o distribuție de probabilitate utilizată în inferența statistică și în testele de ipoteze. Aceasta este folosită atunci când se estimează parametrii unei populații pe baza unei eșantioane mici. Distribuția " t " este utilizată în testele de ipoteze când se compară două medii de eșantioane independente sau atunci când se dorește determinarea intervalului de încredere pentru media unei populații. Distribuția " t " este similară cu distribuția normală, dar are cozi mai groase, ceea ce

înseamnă că are o probabilitate mai mare pentru valorile extreme. Forma exactă a distribuției " t " depinde de numărul de grade de libertate, care este determinat de mărimea eșantionului utilizat în estimarea parametrilor. Pe măsură ce numărul de grade de libertate crește, distribuția " t " se apropie tot mai mult de distribuția normală. Aceasta este esențială în multe analize statistice și este utilizată într-o varietate de domenii, inclusiv în cercetarea științifică, analiza financiară și testarea medicamentelor [193-196].

Aplicarea testului t

Deoarece, seturile de date sunt independente, se testează dacă diferența dintre mediile celor două eșantioane este semnificativă din punct de vedere statistic.

După identificarea datelor se stabilesc ipotezele statistice [187,194]. Astfel, se analizează *ipoteza nulă* pentru a determina dacă mediile eșantioanelor sunt egale și nu există diferențe între eșantioane:

$$H_0: \mu_1 - \mu_2 = 0 \quad (2.2)$$

sau *ipoteza alternativă* prin care se constată diferențe semnificative:

$$H_a: \mu_1 - \mu_2 \neq 0 \quad (2.3)$$

În cazul când avem dispersii diferite ($\sigma^2 \neq \sigma^2$) și număr de repetiții este egal ($n_1 = n_2$), se va calcula media diferențelor.

Etapele testării:

1. Se calculează diferența mediilor ($\bar{d} = x_1 - x_2$); (2.4)
2. Se calculează eroarea mediei diferențelor, folosind formula: $s_{\bar{d}}$

$$S_{\bar{d}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (2.5)$$

3. Determinăm diferențele semnificative (t_{exp}), raportând diferența mediilor la eroarea diferențelor conform formulei (1);
4. Determinăm numărul de grade de libertate pentru testul independent:

$$df = n_1 + n_2 - 2 \quad (2.6)$$

Gradele de libertate reprezintă numărul comparațiilor independente între elementele unei mărimi de observație sau numărul valorilor care pot fi alese aleatoriu în cadrul unei specificări și sunt utilizate pentru a determina rezultatul diferitelor scenarii dintr-un sistem.

5. Se determină valoarea critică t_{crit} pentru nivelul de semnificație ales α (de regulă 0.05 și 0.01) și pentru gradul de libertate (df - degree of freedom).
6. Dacă $t_{exp} > t_{teor}$, ipoteza nulă privind lipsa diferențelor semnificative se respinge, altfel se acceptă, diferențele se află în cadrul intervalului de încredere.

7. Rezultatul poate fi confirmat și prin utilizarea analizei *cele mai mici diferențe seminifcative* (*lsd* - *Least Significant Difference*), calculând produsul dintre media diferențelor și eroarea diferențelor:

$$\text{In}[*] := \text{lsd} = t_{0.05} * s_d^- \quad (2.7)$$

$$\text{Out}[*] = \text{lsd} = t_{0.01} * s_d^- \quad (2.8)$$

O alternativă pentru punctul de respingere a ipotezei nule (H_0) este valoarea p [197-199] ., care oferă cel mai mic prag de semnificație sub care ipoteza alternativă (H_a) se adevărește că există suficiente dovezi pentru a respinge ipoteza nulă (H_0). Este o măsură a semnificației statistice. Rezultatul p al testului, furnizat ca un număr între 0 și 1, și reprezintă *probabilitatea de eroare* dacă respingem ipoteza H_0 . Dacă p este mai mic decât pragul de semnificație α ales (de obicei $\alpha = 0,05$), se respinge ipoteza H_0 și admitem ca adevărată ipoteza H_a [197]. Interpretarea valorilor p se face la majoritatea testelor statistice astfel că:

- $p < 0.05$ (5%), legătura statistică este semnificativă (S, nivel de încredere 95%).
- $p < 0.01$ (1%), legătura statistică este semnificativă (S, nivel de încredere 99%).
- $p < 0.001$ (0.1%), legătura statistică este înalt semnificativă (HS, nivel de încredere 99.9%).
- $p > 0.05$, legătura statistică este nesemnificativă (NS).

Valoarea p nu oferă informații despre:

8. Șansa de beneficiu pentru o plantă sau un soi luat individual;
9. Procentul de plante care vor avea un beneficiu în urma suplinirii cu un anumit preparat;
10. Gradul de beneficiu așteptat pentru o anumită plantă.

2.4. Modelarea datelor experimentale

Pentru realizarea obiectivelor cercetării au fost elaborate 9 modele de analiză biostatistică a datelor obținute experimental în cadrul cercetărilor desfășurate în cadrul Departamentului Biologie și Ecologie a Universității de Stat din Moldova. Aceste modele sunt realizate în *Wolfram Mathematica*, dezvoltată de *Wolfram Research*, este un sistem avansat de calcul simbolic și numeric, furnizând o platformă integrată pentru dezvoltarea aplicațiilor matematice. Abordarea simbolică permite manipularea expresiilor matematice simbolice fără a le evalua numeric. *Wolfram Mathematica* dispune de propriul limbaj de programare, numit limbajul *Wolfram*, care permite crearea de algoritmi personalizați și dezvoltarea de aplicații matematice complexe. Datorită funcționalității sale extinse și a capacității sale de calcul, *Wolfram Mathematica* este considerată una dintre cele mai puternice și versatile unelte de calcul matematic disponibile în prezent [200]. În plus, *Wolfram Mathematica* poate fi utilizată și pentru

analiza biostatistică a datelor în biologie și agricultură. Aceasta furnizează funcții și pachete specializate pentru a efectua analize statistice și pentru a genera vizualizări relevante pentru datele din domeniul biologic și agricol [201].

În teză sunt analizate unele modele de analiză biostatistică a datelor (modelele sunt prezentate integral în anexe). Aceste modele au ca obiective asistența cercetărilor pentru determinarea gradelor de influență a preparatelor *Reglalg* (preparat de natură algală) și *Biovit* (preparat de natură humică), precum și a fitohormonului *Heteroauxina* asupra creșterii și dezvoltării plantelor de soia, (*Glycine max* (L.) Merrill.) pe fond de administrare a preparatelor obținute din surse naturale.

Componentele modelului au fost dezvoltate în *Wolfram Mathematica*.

Pentru a determina gradul de influență asupra parametrilor cercetați în baza unui eșantion mic, s-a utilizat testul *t*. Remarcăm, că seturile de date sunt independente cu dispersii diferite ($\sigma_1^2 \neq \sigma_2^2$) și număr de egal de repetiții ($n_1 = n_2 = 3$), astfel, se va calcula diferența mediilor.

Pentru realizarea sarcinilor expuse au fost elaborate funcțiile necesare pentru fiecare soi și parametru. Funcțiile elaborate sunt prezentate și descrise în continuare. Aceste funcții au fost elaborate pentru analiza fiecărui parametru.

1) **Declararea variabilelor:** Pentru procesarea datelor la fiecare parametru se declară variabile pentru introducerea valorilor sub formă de liste:

cN2M = {5.35, 5.39, 5.37}, *cN2R* = {6, 6.05, 6.03}, *cN2B* = {6.17, 6.19, 6.26};

Generarea diagramei: Se creează diagrama care demonstrează media valorilor procesate cu eroarea standard în baza datelor obținute experimental:

Prin intermediul funcției **MeanAround** se calculează media valorilor analizate și eroarea standard care va fi indicată pe bară.

Funcția **BarChart** determină tipul de diagramă sub formă de coloane și are ca parametri de formatare stilul, etichete, legenda, aspectul ș. a.

```
ClaveraN2={MeanAround[{5.35, 5.39, 5.37}], MeanAround[{6.0, 6.05, 6.03}], MeanAround[{6.17, 6.19, 6.26}]}];
BarChart[ClaveraN2, ChartElementFunction->"GlassRectangle",
ChartStyle-> "Pastel", ImageSize->400, Chartlabels ->{"1", "2", "3"},
Chartlegends-> {"Martor", "Reglalg", "Biovit"}];
```

Variația parametrului studiat

Au fost elaborate funcțiile pentru analiza variației valorilor parametrului studiat. Aceasta se materializează într-un tabel, utilizând funcția *Grid* și funcțiile statistice pentru fiecare parametru și grup experimental: *MeanAround* (media cu eroarea standard), *Variance* (dispersia –

σ^2), *StandardDeviation* (abaterea standard - σ), *Coeficientul de variație*, exprimat în procente, determinat ca *raportul dintre abaterea standard și media valorilor eșantionului * 100*, intervalul de încredere determinat prin funcția *MeanCI[cN2M]*:

```
Grid[{{"Parametri", "x̄ (%)", "σ²", "σ", "CV(%)", "CI**"},
{"Martor", MeanAround[cN2M], Variance[cN2M], StandardDeviation[cN2M],
StandardDeviation[cN2M]/Mean[cN2M]*100, MeanCI[cN2M]}, {"Reglalg",
MeanAround[cN2R], Variance[cN2R], StandardDeviation[cN2R],
StandardDeviation[cN2R]/Mean[cN2R]*100, MeanCI[cN2R]}, {"P₂",
MeanAround[cN2B], Variance[cN2B], StandardDeviation[cN2B],
StandardDeviation[cN2B]/Mean[cN2B]*100, MeanCI[cN2B]}}/N]]
```

Influența tratamentului asupra parametrului studiat

Au fost elaborate funcțiile pentru analiza influenței unor factori asupra parametrului studiat. Se compară influența unui biostimulator cu eșantionul de control (martorul) pentru a afla:

- Diferența mediilor eșantioanelor ($\bar{d}$),
- Eroarea standard a diferențelor ($s_{\bar{d}}$),
- Rezultatul testului t ,
- Limita celei mai mici diferențe semnificative (**lsd**):

```
Grid[{{"Comparatie", "d̄", "s_d", "t", "lsd0.05", "lsd0.01"},
{"Martor vs Reglalg", medCN2MR = Abs[Mean[cN2M] - Mean[cN2R]], ercN2MR = √
0.12² + 0.15², t = (medCN2MR/ercN2MR)*, lsd0.05 = t0.05 * ercN2MR*, lsd0.01 = t0.01 *
ercN2MR*},
{"Martor vs Biovit", medCN2MB = Abs[Mean[cN2M] - Mean[cN2B]], ercN2MB = √
0.12² + 0.27², t = (medCN2MB/ercN2MB)*, lsd0.05 = t0.05 * ercN2MB*, lsd0.01 = t0.01 *
ercN2MB*}]
```

Influența biostimulatorilor asupra parametrului studiat se materializează într-un tabel din care rezultă valoarea t_{exp} obținută în urma analizei datelor la utilizarea fiecărui biostimulator, atât la nivel pragului de semnificație de **0.05**, cât și de **0.01**. Astfel, ipoteza nulă *se confirmă* sau *se respinge*.

Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative. Dacă diferențele mediilor ($\bar{d}_{param}$) sunt mai mari decât valorile **lsd** obținute conform datelor obținute se confirmă *se respinge* ipoteza nulă. În caz contrar se confirmă ipoteza nulă.

Pentru a confirma ipoteza alternativă se utilizează suplimentar testul **p**. În acest sens se utilizează funcția *StudentTPValue* care primește ca parametri valoare t_{exp} și gradele de libertate (**df**):

$$StudentTPValue[t_{exp}, df] \rightarrow p$$

Dacă rezultatul acestui test este mai mic decât pragul de semnificație, *se confirmă* rezultatul obținut prin testul **t**.

2.5. Depozit de date în agricultură

Un depozit de date (*data warehouse - DW*), este o platformă sau un sistem care colectează, integrează și stochează date din diferite surse și formate într-un format structurat, ușor de accesat și de analizat. Scopul unui depozit de date este să ofere un mediu centralizat pentru stocarea și gestionarea informațiilor într-un mod optimizat pentru analiză și raportare [204].

Principalele caracteristici ale unui depozit de date includ:

- a) *Integrare a datelor*: Depozitele de date adună datele din surse multiple și diferite, cum ar fi bazele de date operaționale, fișiere, aplicații și alte surse de informații. Aceste date sunt integrate și transformate pentru a se potrivi într-un model de date comun în cadrul depozitului de date.
- b) *Structurare și organizare*: Datele sunt organizate într-un format structurat și coerent, utilizând scheme de bază de date, tabele, noduri și relații. Aceasta facilitează interogările și analizele ulterioare ale datelor.
- c) *Performanță și optimizare*: Depozitele de date sunt proiectate pentru a oferi performanță ridicată și răspuns rapid la interogări complexe. Se utilizează tehnici de optimizare, cum ar fi indexarea, partiționarea și agregarea datelor, pentru a asigura accesul eficient la informații.
- d) *Acces și analiză ușoară*: Un depozit de date permite utilizatorilor să acceseze și să analizeze datele într-un mod ușor și flexibil. Sistemele de interogare și instrumentele de analiză permit utilizatorilor să formuleze interogări complexe și să extragă informații relevante din date.
- e) *Istoricul și managementul modificărilor*: Depozitele de date păstrează istoricul datelor, permițând utilizatorilor să analizeze tendințele și evoluția în timp. De asemenea, se asigură că modificările datelor sunt gestionate în mod corespunzător și că versiunile anterioare ale datelor pot fi accesate și analizate.

Depozitele de date sunt utilizate în diverse domenii și industrii pentru a sprijini luarea deciziilor, analiza afacerilor, raportarea și analiza datelor. Acestea furnizează o sursă centralizată și fiabilă de informații, ceea ce facilitează procesele de analiză și îmbunătățește înțelegerea datelor pentru a lua decizii mai informate.

Un depozit de date în agricultură este o platformă sau un sistem centralizat care colectează, stochează și gestionează date relevante pentru domeniul agricol. Acesta poate include informații despre sol, climă, precipitații, tipuri de culturi, tehnici agricole, rezultate ale recoltelor, date privind dăunătorii și multe altele. Scopul principal al unui depozit de date în agricultură este

să ofere acces rapid și ușor la informații importante pentru a sprijini deciziile și operațiunile agricole eficiente [202].

Depozitele de date agricole pot include diferite tipuri de date, cum ar fi:

- a) *Date despre sol și condiții pedoclimatice*: Aceste date se referă la caracteristicile solului, inclusiv compoziția chimică, textura, *pH*-ul, capacitatea de reținere a apei etc. De asemenea, includ informații despre condițiile meteorologice specifice zonei, cum ar fi temperatura, precipitațiile, viteza vântului etc.
- b) *Date despre culturi și semințe*: Aceste date cuprind informații despre tipurile de culturi cultivate, semințe utilizate, tehnici de însămânțare, calendarul de plantare, răspunsul culturilor la diferiți factori și altele.
- c) *Date privind managementul culturilor*: Acestea includ informații despre practicile agricole utilizate, cum ar fi irigația, fertilizarea, tratamentele fitosanitare, lucrările mecanice etc. De asemenea, pot include informații despre rotirea culturilor, metodele de control al buruienilor, monitorizarea dăunătorilor și altele.
- d) *Date despre recolte și randamente*: Aceste date se referă la rezultatele recoltelor, inclusiv cantitatea și calitatea produselor agricole obținute. Ele pot include informații despre randamentul culturilor, pierderile în timpul recoltării, valoarea nutritivă a culturilor și altele.
- e) *Date privind prețurile și piețele agricole*: Aceste date includ informații despre fluctuațiile prețurilor la produsele agricole, cererea și oferta de pe piețele locale și internaționale, tendințe de consum etc. Aceste informații pot fi utile pentru fermieri în luarea deciziilor de comercializare a produselor agricole.

Un depozit de date în agricultură poate utiliza tehnologii precum senzori *IoT*, teledetecție, analiză geospațială și altele pentru a colecta și gestiona datele. Aceasta permite fermierilor, cercetătorilor, specialiștilor agricoli și decidenților să acceseze și să utilizeze informațiile necesare pentru a lua decizii informate și pentru a optimiza operațiunile agricole [203]. Un astfel de sistem poate aduce multiple beneficii și poate fi o resursă valoroasă pentru industria agricolă. Iată câteva motive pentru care este necesar un astfel de depozit de date:

- a) *Luarea deciziilor informate*: Agricultura modernă se bazează tot mai mult pe date și informații pentru a lua decizii eficiente. Un depozit de date agricole poate colecta, stoca și analiza informații despre sol, climă, precipitații, tipuri de culturi, tehnici agricole, rezultate ale recoltelor etc. Acesta poate ajuta fermierii și specialiștii agricoli să ia decizii

informate privind momentul optim de semănat, necesarul de apă și nutrienți, gestionarea dăunătorilor și multe altele.

- b) *Îmbunătățirea productivității*: Prin colectarea și analiza datelor agricole, un depozit de date poate oferi informații despre cele mai bune practici agricole și tehnici de creștere a productivității. Acesta poate identifica modele și tendințe în funcție de diferiți factori, precum solul, condițiile meteorologice sau sistemele de irigații, și poate oferi sfaturi personalizate pentru a maximiza randamentul culturilor.
- c) *Reducerea riscurilor și optimizarea resurselor*: Un depozit de date agricole poate ajuta la identificarea riscurilor asociate cu anumite boli, dăunători sau schimbări climatice. Cu ajutorul datelor, fermierii pot implementa strategii preventive și pot interveni rapid pentru a minimiza impactul negativ. De asemenea, prin analiza datelor, se pot identifica modele de utilizare eficientă a resurselor, cum ar fi apa, fertilizanții și pesticidele, ceea ce poate reduce costurile și impactul asupra mediului.
- d) *Progresul cercetării agricole*: Un depozit de date agricole poate deveni o sursă importantă de informații pentru cercetători și instituții de învățământ agricol. Aceste date pot fi utilizate pentru a dezvolta noi tehnologii, hibrizi de culturi rezistente sau metode inovatoare de creștere a plantelor. Accesul la un depozit centralizat de date poate accelera progresul în cercetarea agricolă și poate promova schimbul de cunoștințe între diferiți actori din industrie.
- e) *Planificarea strategică a agriculturii*: Un depozit de date agricole poate contribui la planificarea strategică a agriculturii la nivel regional sau național. Colectarea și analiza datelor agricole din diferite zone geografice pot ajuta la identificarea nevoilor și provocărilor specifice, la evaluarea tendințelor și la dezvoltarea politicii agricole adecvate.

Pentru a dezvolta un depozit de date în agricultură a fost ales sistem de gestionare a bazelor de date de tip graf *Neo4j*. Acest sistem reprezintă o platformă care se concentrează pe gestionarea grafurilor. Integrarea *Neo4j* în acest context poate aduce beneficii semnificative în gestionarea informațiilor complexe și în efectuarea analizelor și vizualizărilor în format graf. Grafurile sunt utile pentru a reprezenta și a vizualiza relațiile complexe între diferite entități [204]. Acestea permit o abordare holistică în gestionarea domeniului agricol și alte domenii legate cu acesta [205-206].

Neo4j este o bază de date graf orientată către obiecte care poate fi utilizată pentru a construi aplicații complexe de procesare a datelor, inclusiv în domeniul fiziologiei plantelor și agriculturii. Abordarea holistică în domeniul fiziologiei plantelor prin *Neo4j* poate implica

utilizarea acestei baze de date pentru a modela interacțiunile complexe dintre plante și factorii de mediu, precum temperatura, umiditatea, nivelul de nutrienți, etc. Această abordare poate oferi o viziune mai cuprinzătoare asupra modului în care plantele interacționează cu mediul lor și pot fi utilizate pentru a dezvolta modele predictive pentru creșterea și dezvoltarea plantelor [206].

În agricultură, *Neo4j* poate fi folosit pentru a modela relațiile între diverse factori, cum ar fi tipul de sol, clima, istoricul culturii și istoricul de tratamente aplicate culturilor, pentru a ajuta la optimizarea producției și la luarea de decizii bazate pe date în agricultură. Aceste modele pot ajuta agricultorii să ia decizii mai bine informate și să îmbunătățească randamentul culturilor [206].

Neo4j poate fi, de asemenea, utilizat pentru a analiza datele din senzori *IoT* care sunt amplasați în teren, și pot fi utilizate pentru a monitoriza diferiți parametri cum ar fi temperatura, umiditatea, nivelul de nutrienți, etc. Aceste date pot fi apoi integrate în baza de date și analizate pentru a înțelege mai bine procesele fiziologice ale plantelor și pentru a optimiza practicile agricole.

Aceste tehnologii sunt holistice deoarece utilizează conexiuni simple pentru generarea unui comportament inteligent, bazat pe cunoaștere. Pentru a obține un răspuns eficient se pot utiliza analiza grafurilor, organizând datele într-o bază de date de tip graf. O bază de date graf utilizează structuri de tip graf pentru interogări semantice cu noduri, relații și proprietăți pentru a reprezenta și stoca date. Bazele de date graf prezintă datele așa cum sunt vizualizate conceptual. Acest lucru se realizează prin transferul datelor în noduri și relațiile acestora în margini [207].

Tipul de baze de date graf, în special *Neo4j*, se utilizează cu succes în cercetările biologice și medicale, dar în agricultură și implicit, în studierea dezvoltării plantelor de soia, nu sunt reflectate. Conform companiei de consultanță și dezvoltare *IT Gartner*, „până în anul 2025, tehnologiile graf vor fi utilizate în 80% de date și inovații de analiză, în sus de la 10% în 2021, facilitând luarea rapidă a deciziilor ce se impun” [208].

Pentru crearea unui depozit de date cu *Neo4j* în agricultură, există câteva aspecte importante de luat în considerare:

- *Modelarea datelor*: Înainte de a începe construirea depozitului de date, este necesară realizarea unei scheme a grafului care să reprezinte entitățile și relațiile din domeniul agriculturii. Aceasta implică definirea nodurilor pentru culturi, specii de plante, animale, soluri, ferme, cercetători, experimente și altele, precum și stabilirea relațiilor între acestea.
- *Inserarea și importul datelor*: Pentru a putea utiliza datele în *Neo4j*, acestea trebuie importate în platformă. Importul poate fi realizat folosind diverse metode, cum ar fi

importul din fișiere CSV sau utilizarea limbajului de interogare *Cypher* pentru a introduce direct datele în baza de date.

- *Crearea de indici și restricții*: *Neo4j* oferă suport pentru crearea de indici și restricții, care contribuie la îmbunătățirea performanței și integrității datelor. De exemplu, se pot crea indici pentru anumite proprietăți pentru a facilita căutarea și accesarea eficientă a datelor.
- *Interogări și analize*: Prin utilizarea limbajului de interogare *Cypher*, este posibil să se efectueze interogări complexe pentru a extrage informații specifice din depozitul de date. *Neo4j* oferă funcționalități puternice de analiză și vizualizare grafică, permițând identificarea de modele, efectuarea de analize de rețea și vizualizarea datelor într-un mod intuitiv.
- *Integrarea cu alte sisteme*: În funcție de nevoile specifice, depozitul de date *Neo4j* poate fi integrat cu alte sisteme și aplicații utilizate în domeniul agriculturii. Aceasta poate implica integrarea cu software-uri de analiză sau de gestionare a fermelor, facilitând utilizarea mai extinsă a datelor și extinderea funcționalității.

Este important de menționat că dezvoltarea eficientă și gestionarea unui depozit de date cu *Neo4j* în domeniul agriculturii care să asigure răspunsuri comprehensive și soluționarea problemelor specifice domeniului studiat necesită expertiză în domeniul bazelor de date graf și în domeniul specific al agriculturii. De asemenea, se recomandă adoptarea unor bune practici de proiectare a depozitului și adaptarea soluțiilor la cerințele și nevoile proiectului în cauză, precum și la nevoile utilizatorilor finali. În continuare, vor fi prezentate câteva dintre cele mai bune practici recomandate pentru crearea unui depozit de date în agricultură utilizând platforma *Neo4j*:

- *Planificarea schemei grafului*: Înainte de a începe construirea depozitului de date, este esențial să se planifice cu atenție schema grafului. Acest proces implică identificarea entităților cheie și a relațiilor dintre acestea în contextul agriculturii și se va crea un model grafului care să reflecte aceste aspecte. Se va lua în considerare modul în care datele vor fi interconectate și cum se vor utiliza relațiile pentru a extrage informații relevante.
- *Utilizarea indexării*: *Neo4j* oferă capacități puternice de indexare, creându-se indecși cu eficiență înaltă pentru proprietățile cheie. Este recomandat să se utilizeze indexarea pentru proprietățile frecvent utilizate în operațiile de căutare și filtrare. Această practică poate accelera semnificativ interogările și accesul la datele dorite și va permite o căutare mai rapidă a datelor.

- *Evitarea relațiilor extinse*: Se va evita crearea unor lanțuri lungi de relații în depozitul de date. Acestea pot afecta performanța sistemului și pot conduce la interogări ineficiente și la performanțe scăzute. În schimb, se vor identifica relațiile semnificative și crearea unor modele graf bine structurate, care să permită interogări rapide și eficiente.
- *Utilizarea tranzacțiilor*: Pentru a asigura coerența, consistența și integritatea datelor și pentru a preveni conflictele, se recomandă utilizarea tranzacțiilor atunci când se modifică datele în depozitul *Neo4j*.
- *Optimizarea interogărilor*: Pentru a obține rezultate cât mai rapide și eficiente, se recomandă optimizarea interogărilor. Aceasta poate implica utilizarea indecșilor, se va lua în considerare structura grafului de date și se vor utiliza clauzele de filtrare și limitare pentru a restrânge volumul de date procesate în timpul interogărilor.
- *Monitorizarea performanței*: Este important să se monitorizeze în mod regulat performanța depozitului de date *Neo4j* pentru a identifica posibilele probleme și pentru a îmbunătăți performanța și eficiența sistemului. Se vor utiliza instrumentele de monitorizare disponibile și se vor ajusta configurările și strategiile de indexare în funcție de nevoi.
- *Asigurarea securității datelor*: agricultura poate implica date sensibile și confidențiale, cum ar fi informații despre culturi, specii de plante sau cercetări. Este esențial să se implementeze măsuri adecvate pentru a proteja datele împotriva accesului neautorizat și a pierderii datelor.

La elaborarea unui depozit de date este important să se evalueze nevoile specifice ale proiectului și să se adapteze structura depozitului pentru a reflecta în mod adecvat diversele aspecte și interacțiuni din industria agricolă.

Un depozit de date poate fi constituit din mai multe și diverse modele sau grafuri. Acest lucru poate fi realizat prin utilizarea unei abordări cunoscute în programarea orientată pe obiect sub numele de *polimorfism*. În modelarea depozitelor de date putem vorbi de „*polimorfismul modelului de date*”. Acest lucru permite utilizarea mai multor modele de date în cadrul aceluiași depozit pentru a gestiona diferite tipuri de informații și pentru a satisface diverse cerințe. Există mai multe motive pentru a utiliza mai multe modele de date în cadrul unui depozit. Unele dintre aceste motive includ:

- *Diversitatea datelor*: Dacă depozitul de date conține o varietate mare de informații, cum ar fi date structurate, semi-structurate sau nestructurate, poate fi necesar să se utilizeze diferite modele de date pentru a stoca și gestiona aceste tipuri diverse de date în mod eficient.

- *Cerințe de acces și interogare*: Diferite modele de date pot oferi funcționalități și capacități diferite pentru accesul și interogarea datelor. Prin utilizarea mai multor modele de date, se poate asigura că fiecare tip de informație este stocat și gestionat într-un mod care să permită accesul și interogarea optimă.
- *Performanță și scalabilitate*: Unele modele de date pot fi mai eficiente și scalabile în gestionarea anumitor tipuri de date sau operațiuni. Prin utilizarea modelelor de date potrivite pentru fiecare scenariu specific, se poate obține o performanță mai bună și o scalabilitate mai mare a depozitului de date.

Este important să se evalueze și să se planifice cu atenție utilizarea mai multor modele de date în cadrul unui depozit, deoarece poate implica o complexitate suplimentară în proiectare, dezvoltare și administrare. De asemenea, este important să se asigure coerența și integritatea datelor între diferitele modele de date și să se stabilească legături și relații între ele, acolo unde este necesar, pentru a obține o viziune integrată și coerentă asupra informațiilor stocate în depozit.

Bazele de date de tip graf, în special *Neo4j*, reprezintă un instrument puternic pentru abordarea holistică a rezolvării problemelor și dezvoltarea durabilă. *Neo4j* este un sistem de gestionare a bazelor de date de tip graf care permite modelarea și stocarea datelor sub formă de noduri și relații. Aceste baze de date pot fi utilizate pentru a rezolva atât probleme structurate, cât și slab structurate sau nestructurate [209].

Neo4j oferă funcționalități avansate, inclusiv un motor de stocare pentru gestionarea datelor grafului și limbajul de interogare *Cypher* pentru interogarea și actualizarea datelor. Optimizatorul de interogări îmbunătățește performanța interogărilor complexe, iar *Neo4j* asigură scalabilitate, securitate și integrare cu alte tehnologii și limbaje de programare. Modulele de vizualizare și analiză facilitează înțelegerea datelor grafului, iar replicarea și redundanța asigură disponibilitatea și durabilitatea datelor.

Utilizarea sistemului graf *Neo4j* poate sprijini sarcinile de predicție a fenomenelor. De exemplu, datele și relațiile stocate într-o bază de date *Neo4j* pot fi utilizate ca intrare pentru modele de învățare automată care fac predicții bazate pe aceste date. Limbajul de interogare *Cypher* poate fi, de asemenea, folosit pentru a identifica modele și tendințe în date, care pot fi apoi utilizate ca intrare pentru un model de predicție. Aceste capacități permit reprezentarea relațiilor complexe între diferite elemente ale unei probleme și evaluarea impactului lor asupra întregului sistem.

În cadrul dezvoltării durabile, bazele de date de tip graf pot fi utilizate pentru a analiza și înțelege impactul diferitelor activități și decizii asupra sistemului studiat. Aceste baze de date

permit analiza interdependențelor dintre factori și evaluarea impactului asupra economiei, mediului sau societății în ansamblu. De asemenea, ele pot fi utilizate pentru a identifica interdependențele și relațiile complexe dintre diferitele probleme ale societății, cum ar fi problema alimentară și schimbările climatice. Prin identificarea soluțiilor potențiale și găsirea unor soluții integrate care să abordeze toate aspectele relevante ale unei probleme, bazele de date graf contribuie la abordarea holistică a rezolvării problemelor și dezvoltarea durabilă.

Utilizarea bazelor de date grafice oferă decidenților posibilitatea de a identifica mai precis consecințele diferitelor opțiuni și de a lua decizii mai informate și durabile. Această abordare poate contribui la dezvoltarea de strategii de dezvoltare durabilă și la reducerea impactului negativ asupra mediului și societății.

Pentru a elabora un depozit de date eficient în domeniul agriculturii, este important să se determine tipurile de entități cheie, entitățile (nodurile) și relațiile dintre acestea. Această etapă este esențială pentru a crea o schemă a grafului coerentă și pentru a facilita accesul și analiza datelor relevante. În continuare sunt redate unele entități și relații luate în considerare în elaborarea depozitului de date în domeniul agriculturii. Entitățile și relațiile sunt analizate ținând cont de specificul proiectului și se identifică entitățile și relațiile relevante pentru a obține o reprezentare coerentă și comprehensivă a datelor agricole. Modelele realizate utilizează pentru instruire datele din perioada anilor 2002 – 2018. Pentru testare au fost verificate datele din anii 2019 – 2022 [210].

Determinarea tipurilor de entități (definirea nodurilor)

Aceste tipuri de entități și relații pot fi modelate în *Neo4j* pentru a construi un depozit de date graf care să permită stocarea și interogarea eficientă a informațiilor despre culturi, soluri, ferme, zone geografice, resursele și condițiile meteorologice asociate. În *Neo4j*, entitățile pot fi reprezentate ca noduri, iar relațiile pot fi reprezentate ca muchii între noduri. Atributele specifice ale fiecărei entități pot fi asociate nodurilor respective, pentru a stoca informațiile relevante.

În elaborarea unui depozit de date graf în *Neo4j* pentru a gestiona informațiile despre culturi, tipuri de soluri, ferme, zone geografice, resurse de apă, umiditatea solului, cantitatea de precipitații, temperatura aerului (media și mediana) și viteza vântului, putem defini următoarele tipuri de entități și relații:

- *Culturi*: permite reprezentarea diferitelor tipuri de plante și culturi agricole, precum cereale, legume, fructe etc. Aceasta poate include informații despre proprietățile și caracteristicile specifice ale fiecărei culturi. Fiecare cultură poate avea atribute precum nume, descriere, cerințe specifice de sol, cerințe de apă, etc.

- *Soiuri*: Acest tip de entități descriu soiurile de plante de fiecare specie (figura 2.3). Aceste entități au proprietăți precum nume, descriere, cerințe specifice, culoarea frunzelor, ș. a.

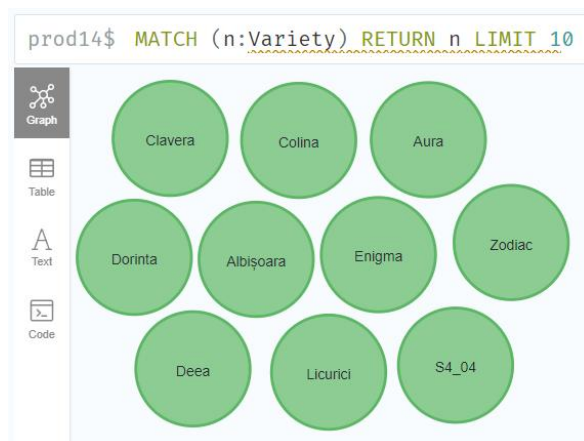


Figura 2.2. Noduri de tip Soiuri

- *Tipuri de soluri*: reprezintă diferitele categorii de soluri utilizate în agricultură, precum sol argilos, sol nisipos, solul negru etc. Aceasta poate cuprinde informații despre compoziția solului, fertilitatea și alte caracteristici relevante. Fiecare tip de sol poate avea atribute precum nume, descriere, textură, compoziție, pH, etc.
- *Ferme*: reprezintă unitățile agricole și include informații despre diferitele ferme sau exploatații agricole existente. Aceasta poate cuprinde detalii despre dimensiunea fermei, localizare, proprietar, echipamente utilizate și altele. unde se desfășoară activități agricole. Fiecare fermă poate avea atribute precum nume, locație, suprafață, etc.
- *Zone geografice*: reprezintă regiunile geografice în care se află fermele și culturile. Fiecare zonă geografică poate avea atribute precum nume, țară, regiune, coordonate geografice, etc.
- *Resurse de apă*: reprezintă sursele de apă disponibile pentru irigarea culturilor. Fiecare resursă de apă poate avea atribute precum tipul resursei (râu, lac, sondă, bazin etc.), capacitatea de stocare, calitatea apei, etc.
- *Umiditatea solului*: reprezintă nivelul de umiditate a solului pentru o anumită zonă și cultură. Acesta poate fi măsurat în procente și poate varia în funcție de condițiile meteorologice și de gestionarea apei.
- *Cantitatea de precipitații*: reprezintă volumul de apă care cade pe o anumită zonă în formă de precipitații (ploaie, zăpadă etc.). Acesta poate fi măsurat în unități de lungime (milimetri) sau volum (litri), în funcție de preferințe.
- *Temperatura aerului*: reprezintă temperatura aerului într-o anumită zonă și perioadă de timp. Aceasta poate fi măsurată în grade Celsius și poate avea ca valori cantitatea de precipitații pe

lună, medii și mediane pe an o cultură specifică poate fi asociată cu un anumit tip de sol adecvat pentru a evidenția variațiile.

- *Coeficient hidrotermic (CHT)* este un indicator relativ (fără unitate de măsură) a stării de hidratare a teritoriului [212-213]. Se determină prin raportul cantității de precipitații (R) pentru perioada cu temperatura medie zilnică a aerului mai mare de 10°C și al sumei temperaturilor medii zilnice pentru această perioadă, micșorată de 10 ori (caracterizează suficient de exact evaporarea) conform formulei:

$$CHT = r/0.1 * \Sigma t^{\circ} > 10^{\circ}, \text{ unde: } \quad (2.9)$$

CHT - Coeficientul hidrotermic,

R - Cantitatea de precipitații pe unitatea de timp studiată,

Σt° - Suma temperaturilor medii.

Tipul de secetă este definit prin proprietatea *drought* de tip numeric, deoarece limbajul *Cypher* și algoritmi de învățare automată din biblioteca *GDS* nu procesează date de tip șir de caractere. Această proprietate reprezintă clasa din care face parte anul respectiv. Unele noduri sunt arătate în figura 2.4. Astfel, conform Serviciului Hidrometeorologic de Stat din Moldova pe baza analizei detaliate după ani a coeficientului hidrotermic (CHT), a stabilit că valorile [213]:

- $CHT \geq 1.0$ caracterizează o umiditate suficientă, codificată prin valoarea **0**,
- $CHT \geq 0.7$ indică o climă secetoasă, codificată prin valoarea **1**,
- $CHT = 0.6$ indică o secetă ușoară, codificată prin valoarea **2**,
- $CHT \leq 0.5$ indică o secetă puternică și foarte puternică, codificată prin valoarea **3**.

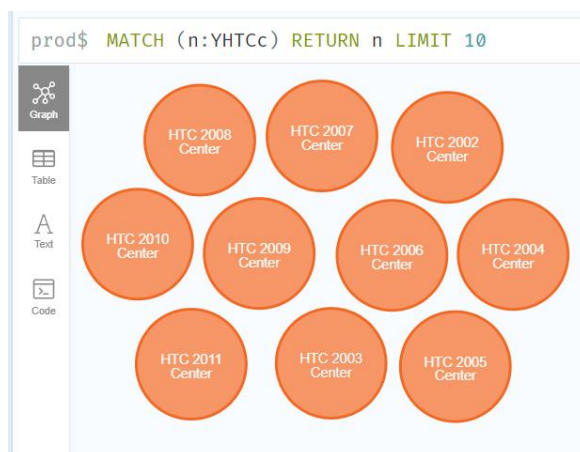


Figura 2.3. Nodurile de tip Coeficientul hidrotermic

- *Viteza vântului*: reprezintă viteza și direcția vântului într-o anumită zonă și perioadă de timp. Aceasta poate fi măsurată în metri pe secundă sau kilometri pe oră.

- De asemenea, pot fi create tipuri de entități ca *Proгноze meteorologice*: Entitatea prognoze meteorologice poate include informații despre condițiile meteorologice prognozate, relevante pentru agricultură, cum ar fi temperatură, precipitații, umiditate, viteza vântului etc. Aceasta poate fi utilă în planificarea și gestionarea activităților agricole.
- *Producători*: Entitatea producători poate reprezenta diferiți producători agricoli sau companii care sunt implicate în procesul de producție și distribuție a produselor agricole. Aceasta poate include informații despre nume, locație, dimensiunea exploatației, culturi principale cultivate și altele.
- *Pesticide și îngrășăminte*: În contextul agriculturii, entitățile pesticide și îngrășăminte pot fi importante pentru monitorizarea și gestionarea utilizării acestor substanțe în culturi. Aceste entități pot conține informații despre tipul de pesticide/îngrășăminte, dozele utilizate, perioadele de aplicare și altele.
- *Cercetători și instituții de cercetare*: Entitatea cercetători poate reprezenta indivizi sau echipe de cercetători implicați în studii și experimente în domeniul agriculturii. Aceasta poate include informații despre specializare, instituție de cercetare, publicații relevante și colaborări.
- *Echipamente agricole*: Această entitate poate reprezenta diferite tipuri de echipamente utilizate în agricultură, cum ar fi tractoare, combine agricole, irigatoare etc. Aceasta poate include informații despre specificațiile tehnice, starea și utilizarea echipamentelor.
- *Vânzări și tranzacții*: Entitatea vânzări și tranzacții poate reprezenta înregistrările despre vânzările de produse agricole sau inputuri agricole, precum semințe, îngrășăminte, pesticide etc. Aceasta poate conține informații despre cantități, prețuri, data tranzacțiilor și părțile implicate.

Determinarea relațiilor între tipuri de entități:

- *Relația între Culturi și Ferme*: evidențiază asocieri specifice culturilor cu fermele în care sunt cultivate, Precum și între ferme și culturile agricole pe care le cultivă. Această relație indică faptul că o anumită fermă poate cultiva una sau mai multe culturi, iar o cultură poate fi cultivată într-o sau mai multe ferme.
- *Relația între Culturi și Tipuri de soluri*: reprezintă asocierea culturilor cu tipurile de soluri potrivite pentru ele. Această relație indică faptul că o anumită cultură poate avea preferințe specifice în ceea ce privește tipul de sol în care se dezvoltă cel mai bine. O cultură specifică poate fi asociată cu un anumit tip de sol adecvat.
- *Relația între ferme și culturi*: Această relație arată ce culturi sunt cultivate în cadrul fiecărei ferme. Aceasta poate evidențiază asocieri specifice între ferme și culturile agricole pe care le cultivă.

- *Relația între Ferme și Zone geografice*: reprezintă asocierea fermei cu zona geografică în care se află. Această relație indică faptul că o fermă poate fi localizată într-o anumită zonă geografică.
- *Relația între Ferme și Resurse de apă*: reprezintă asocierea fermei cu resursele de apă disponibile pentru irigare. Această relație indică faptul că o fermă poate avea acces la o anumită sursă de apă pentru a satisface necesitățile de irigare ale culturilor.
- *Zone geografice și Resurse de apă*: reprezintă asocierea zonei geografice cu resursele de apă disponibile în acea zonă. Această relație indică faptul că o anumită zonă geografică poate avea acces la diferite surse de apă, cum ar fi râuri, lacuri sau fântâni.
- *Relația între Zone geografice și Cantitatea de precipitații*: reprezintă asocierea zonei geografice cu cantitatea de precipitații înregistrată în acea zonă. Această relație indică faptul că o anumită zonă geografică poate înregistra o anumită cantitate de precipitații într-un anumit interval de timp.
- *Relația între Zone geografice și Temperatura aerului*: reprezintă asocierea zonei geografice cu temperaturile aerului înregistrate în acea zonă. Această relație indică faptul că o anumită zonă geografică poate înregistra anumite temperaturi medii și mediane în funcție de sezon și de condițiile meteorologice locale.
- *Relația între Zone geografice și Viteza vântului*: reprezintă asocierea zonei geografice cu viteza vântului înregistrată în acea zonă. Această relație indică faptul că o anumită zonă geografică poate înregistra anumite valori ale vitezei vântului în funcție de teren, altitudine și condițiile atmosferice.
- *Relația între culturi și pesticide/îngrășăminte*: Această relație poate evidenția ce pesticide sau îngrășăminte sunt utilizate în cadrul anumitor culturi. De exemplu, se pot identifica culturile care necesită aplicarea anumitor tipuri de pesticide sau îngrășăminte.
- *Relația între soiuri și calitatea recoltei*: descrie conținutul de elemente și substanțe în boabe/frunze, ș. a.
- *Relația între ferme și producători de produse pentru plante*: Această relație arată ce producători furnizează produsele agricole către anumite ferme. Aceasta poate include informații despre contracte, furnizori preferați și parteneriate.
- *Relația între culturi și cercetători/instituții de cercetare*: Această relație evidențiază implicarea cercetătorilor și a instituțiilor de cercetare în studiile și experimentele legate de anumite culturi. Acest tip de relație poate facilita colaborarea și schimbul de informații între cercetători și comunitatea agricolă.

- *Relația între ferme și echipamente agricole:* Această relație arată ce echipamente agricole sunt utilizate în cadrul fiecărei ferme. Poate include informații despre echipamentele deținute, închiriate sau partajate între ferme.
- *Relația între ferme și vânzări/tranzacții:* Această relație evidențiază tranzacțiile comerciale între ferme și alte entități, cum ar fi distribuitori, procesatori sau consumatori finali. Aceasta poate cuprinde informații despre produsele vândute, cantități, prețuri și altele.
- *Relația între culturi și prognoze meteorologice:* Această relație poate arăta modul în care condițiile meteorologice afectează diversele culturi agricole. Poate fi utilă în analiza impactului schimbărilor climatice asupra randamentului culturilor și în adoptarea unor strategii de gestionare a riscurilor.

Aceste relații permit conectarea și interogarea informațiilor între entități în cadrul depozitului de date graf. De exemplu, putem obține toate culturile cultivate într-o fermă specifică, identifica tipurile de soluri potrivite pentru o anumită cultură sau găsi zonele geografice care înregistrează temperaturi medii ridicate.

Este important de menționat că structura și modelele de date într-un depozit de date graf pot varia în funcție de nevoile și contextul specific al aplicației sau domeniului de lucru. Exemplele de entități și relații prezentate mai sus sunt doar o orientare generală și pot fi adaptate în funcție de cerințele concrete ale utilizatorilor.

În cadrul depozitului de date în agricultură, pe lângă entitățile și relațiile menționate anterior, există și alte entități și relații importante de luat în considerare. Aceste tipuri de entități și relații pot contribui la obținerea unei reprezentări mai complete și mai relevante a datelor agricole în cadrul depozitului de date. Este important să se adapteze modelul graf în funcție de nevoile și specificul proiectului în agricultură, astfel încât să se asigure o gestionare eficientă și o analiză pertinentă a datelor.

În cadrul prezentei tezei, au fost abordate o serie de experimente bazate pe diferite modele de date pentru a investiga și analiza relațiile complexe dintre entități într-un context specific. Aceste experimente au avut ca scop obținerea de rezultate relevante și înțelegerea aspectelor cheie ale problemelor abordate.

Una dintre metodele comune pentru a prezice productivitatea culturilor agricole este analiza datelor meteorologice și agricole anterioare. Această metodă se bazează pe faptul că există o corelație între condițiile meteorologice, cum ar fi temperatura și umiditatea, și productivitatea culturilor agricole. Modelele realizate pot lua în considerare mai mulți factori, cum ar fi calitatea solului, managementul culturilor, nivelul de irigare și fertilizare, precum și condițiile meteorologice, pentru a oferi o predicție mai precisă a productivității. Aceste date pot

fi apoi procesate prin algoritmi de învățare automată și pot fi utilizate pentru a estima productivitatea culturilor.

Scopul unui model de predicție este evaluarea riscurilor și la luarea deciziilor ce se impun, estimarea producției viitoare, minimizarea impactului schimbărilor climatice asupra mediului și a comunităților umane, gestionarea resurselor naturale, precum și dezvoltarea economică durabilă.

Obiectivele modelului de predicție:

1. *Prognostizarea producției viitoare*: Unul dintre principalele obiective ale unui model de predicție a productivității culturilor agricole este de a estima producția viitoare a recoltei în funcție de condițiile meteorologice și a altor variabile.
2. *Identificarea factorilor de influență*: Modelul poate fi utilizat pentru a identifica factorii de influență care afectează productivitatea culturii de soia și pentru a determina care dintre acești factori au cea mai mare influență asupra producției în funcție de factori precum condițiile meteorologice, solul, tipul și cantitatea de input-uri agricole (de exemplu, fertilizanți, biostimulatori, erbicide și pesticide), precum și alte variabile cheie, cum ar fi nivelul de umiditate, temperatură, luminozitate sau nivelul de infestare cu dăunători.
3. *Optimizarea practicilor agricole*: Modelul poate fi utilizat pentru a identifica cele mai bune practici agricole și pentru a optimiza utilizarea input-urilor agricole, precum și pentru a ajuta la luarea deciziilor în timp real în timpul sezonului de cultură.
4. *Îmbunătățirea rentabilității*: Prin furnizarea de estimări precise ale producției viitoare, modelul poate ajuta la creșterea rentabilității prin reducerea costurilor și creșterea producției.
5. *Reducerea riscului și creșterea rezilienței*: Modelul poate ajuta la reducerea riscului asociat cu culturile agricole prin identificarea condițiilor care pot afecta productivitatea și prin furnizarea de informații în timp util pentru a permite agricultorilor să ia măsuri preventive.

Predicția legăturilor este o sarcină în teoria grafurilor și învățarea automată în care scopul este de a prezice legăturile sau relațiile lipsă între nodurile dintr-un graf. În contextul predicției secetei, *Neo4j* poate fi folosit pentru a reprezenta relațiile dintre diverși factori care contribuie la secetă, cum ar fi temperatura, precipitațiile, umiditatea solului, coeficientul hidrotermic, etc. Există mai mulți algoritmi de învățare automată care pot fi utilizați pentru a crea un model de predicție în *Neo4j*. Acești algoritmi pot fi implementați în SGBD *Neo4j* folosind bibliotecile *APOC (Awesome Procedures On Cypher)* și *GDS (Graph Data Science)*, care oferă o varietate de algoritmi de învățare automată și grafuri. Utilizarea algoritmilor de analiză de date, cum ar fi algoritmi de regresie sau de învățare automată, pentru a analiza relațiile dintre date și a determina care dintre acestea sunt cele mai importante pentru prezicerea secetei. Pentru a realiza un model de predicție este important să se aleagă algoritmul potrivit și să se testeze diferite

metode pentru a se vedea care este mai precisă și mai eficientă. Modelarea unui graf de învățare automată de predicție se desfășoară în mai multe etape:

- 1) *Preprocesarea grafului* se referă la procesul de pregătire și prelucrarea datelor brute pentru a le face mai ușor de analizat și de utilizat în algoritmul de învățare automată cum ar fi curățarea și transformarea unui set de date graf înainte de a fi utilizare. Scopul preprocesării este de a oferi caracteristici bune pentru algoritmul de învățare.
- 2) *Analiza datelor*: Odată ce datele au fost modelate se pot dezvolta modele matematice, modele graf, și algoritmi de analiză a datelor din biblioteca *GDS* pentru a analiza datele și a identifica tendințe, modele sau relații relevante. Aceste modele pot fi utilizate pentru a identifica factorii care afectează productivitatea culturilor și a animalelor și pentru a realiza predicții despre producția viitoare. De exemplu, se pot utiliza algoritmi de clusterizare pentru a grupa zonele cu condiții meteorologice similare și utiliza algoritmi de regresie pentru a identifica factorii care au cel mai mare impact asupra productivității.
- 3) *Realizarea predicțiilor* [214]: Pe baza modelelor validate, se pot realiza predicții cu privire la producția viitoare a productivității culturilor agricole în baza noastră de date graf. De exemplu, putem utiliza algoritmi de învățare automată (*ML*). Analiza complexă a grafurilor prin metodele de învățare automată se poate realiza prin crearea de conductele de procesare a grafurilor (în engleză, *Graph Processing Pipelines*) care reprezintă un ansamblu de procese și tehnologii utilizate pentru a efectua operațiuni de analiză și procesare pe grafuri, adică pe modele de date care constau din noduri și relații interconectate. În biblioteca *GDS* din SGBD *Neo4j*, conductele oferă un flux de lucru *end-to-end*, de la extragerea de caracteristici la instruire și aplicarea modelelor de învățare automată. Aceste conducte sunt utilizate în mod obișnuit în domeniul analizei datelor și al învățării automate, pentru a efectua o serie de operații, cum ar fi:
 - a) Sunt extrase date de la noduri și relații
 - b) Se calculează metrici și proprietăți ale grafurilor, cum ar fi centralitatea grafurilor și gradul nodurilor
 - c) Se realizează analize și modelare pe grafuri, cum ar fi clustering-ul și detectarea comunităților
 - d) Se instruiesc și se testează modele de învățare automată pentru a efectua clasificări și predicții pe grafuri.

Metrici. Un instrument esențial în evaluarea performanței grafurilor și înțelegerii rezultatelor îl reprezintă *Metricile*. Ele oferă măsurători obiective și cantitative care permit evaluarea și comparația diferitelor aspecte ale unei activități sau unui sistem [215]. Aceste

metrici sunt concepute pentru a măsura și evalua diferențele dintre rezultatele prezise și valorile de referință, fie că este vorba de clasificarea unui nod într-un anumit grup sau comunitate, de prezicerea unei legături între două noduri sau de alte proprietăți în graf.

Este important de menționat că nu există o metrică universală care să se potrivească tuturor scenariilor de predicție în grafuri. Alegerea și utilizarea metricilor adecvate depind de contextul specific al problemei și de obiectivele de analiză.

Metricile de predicție în grafuri sunt măsurători utilizate pentru a evalua performanța algoritmilor de predicție sau de clasificare aplicate asupra datelor dintr-un graf. Aceste metrici sunt folosite pentru a determina cât de bine algoritmul de predicție reușește să anticipeze sau să clasifice corect anumite proprietăți sau comportamente ale grafurilor.

Printre cele mai comune de metrici de predicție utilizate în analiza grafurilor sunt următoarele:

- 1) *Acuratețe (Accuracy)*: Aceasta măsoară proporția corectă a predicțiilor făcute de algoritm. Se calculează ca raportul dintre numărul de predicții corecte și numărul total de predicții.
- 2) *Sensibilitate (Recall)*: Aceasta măsoară capacitatea algoritmului de a identifica corect elementele pozitive (veritabile) din graf. Se calculează ca raportul dintre numărul de elemente pozitive corect identificate și numărul total de elemente pozitive din graf.
- 3) *Precizie (Precision)*: Aceasta măsoară cât de precis este algoritmul în identificarea elementelor pozitive. Se calculează ca raportul dintre numărul de elemente pozitive corect identificate și numărul total de elemente identificate ca fiind pozitive.
- 4) *Scorul FI (F1 Score)*: Aceasta este o măsură a echilibrului între precizie și sensibilitate. Se calculează ca media armonică între precizie și sensibilitate și oferă o măsură globală a performanței algoritmului de predicție.
- 5) *AUC-ROC (Area Under the Receiver Operating Characteristic Curve)*: Aceasta măsoară capacitatea algoritmului de a clasifica corect elementele pozitive și negative în funcție de un anumit scor de predicție. Este o măsură folosită în special în problemele de clasificare binară.

2.6. Concluzii la capitolul 2

1. A fost dezvoltat un limbaj formal pentru domeniul de cercetare „Modelarea Inteligentă a Datelor din Domeniul Fiziologiei Plantelor”, reflectând complexitatea proceselor biologice din funcționarea plantelor. Limbajul standardizează termenii și conceptele în modelarea datelor legate în domeniul fiziologiei plantelor, facilitând comunicarea și colaborarea între cercetători. Include și reguli matematice pentru reprezentarea proceselor, permițând simulări și prezicerea comportamentului plantelor în diverse condiții.

Modelarea datelor din domeniul fiziologiei plantelor prin utilizarea unui limbaj formal are o importanță și relevanță semnificativă în înțelegerea proceselor complexe care guvernează dezvoltarea, creșterea și răspunsurile plantelor la diverși stimuli. Aceste aspecte sunt esențiale pentru cercetători, biologi, agronomi și alte persoane implicate în studiul plantelor și a mediului înconjurător. Modelarea datelor în acest domeniu are o serie de avantaje, printre care:

- *Precizie în înțelegerea proceselor fiziologice:* Modelele de date formale permit cercetătorilor să reprezinte procesele fiziologice complexe ale plantelor într-o manieră logică și structurată, oferind oportunități de înțelegere mai profundă a interacțiunilor dintre diferitele componente și procese ale plantelor.
- *Previziuni și simulări:* Prin modelarea datelor, se pot crea simulări virtuale ale plantelor și a mediului lor de creștere, ce permit testarea diferitelor ipoteze și scenarii fără a efectua experimente costisitoare în realitate și predicția reacției diferitor genotipuri la acțiunea diferitor factori endogeni și exogeni, precum schimbări climatice sau utilizarea diferitelor tipuri de nutrienți.
- *Modelarea datelor ajută la înțelegerea modului în care plantele răspund la diferite stresuri,* cum ar fi seceta, boală sau infestările cu dăunători. Pot fi dezvoltate strategii mai eficiente pentru combaterea acestor probleme și protejarea recoltelor.
- *Explorarea interacțiunilor ecosistemice:* Modelarea datelor poate ajuta la înțelegerea interacțiunilor complexe dintre organismelor vii și mediul înconjurător și a efectelor pe care le au asupra ecosistemelor în ansamblu.

2. Limbajul elaborat poate servi drept bază pentru software specializat în simularea proceselor fiziologie din plante, ajutând la testarea ipotezelor și prognozarea răspunsurilor plantelor în scenarii diverse, cu implicații în agricultură și domeniile conexe. Continuând cercetarea, limbajul formal poate evolua, reflectând noile descoperiri. Prin colaborare interdisciplinară, acest limbaj poate adânci înțelegerea interacțiunilor plantelor cu mediu.

3. Abordarea holistică permite identificarea unor modele și relații complexe între variabilele studiate la plante și dezvoltarea unor strategii eficiente pentru gestionarea resurselor și a condițiilor de mediu. De asemenea, abordarea holisitcă este esențială pentru înțelegerea și prezicerea comportamentului plantelor în diverse condiții, aducând beneficii în cercetare, biologie și agricultură.

4. Integrarea tehnologiilor inteligente și biostatisticii în abordări holistice este din ce în ce mai importantă, aducând claritate și înțelegere asupra sistemelor complexe. Utilizarea depozitelor de date graf și metodelor biostatistice facilitează analizele complexe și optimizarea performanței plantelor în medii variabile.

5. Combinarea metodelor biostatistice cu utilizarea depozitelor de date sub formă de graf în cercetări de fiziologie a plantelor aduce multiple beneficii pentru înțelegerea proceselor fiziologice complexe și pentru optimizarea performanței și adaptării plantelor la factorii de mediu variabili. Utilizarea metodelor biostatistice, oferă posibilitatea de analiză a datelor experimentale și observaționale pentru a identifica modele și relații între variabilele ce țin de domeniul fiziologiei plantelor și domeniile conexe.

Depozitele de date graf oferă un cadru pentru organizarea și analizarea relațiilor complexe între diferitele entități și variabile. Prin reprezentarea datelor sub formă de graf, interconexiunile dintre componente devin mai vizibile. Prin intermediul algoritmilor graf sunt dezvăluite dependențele și interacțiunile complexe care influențează performanța plantelor. Prin analiza datelor și algoritmi de inteligență artificială, modele predictive pot anticipa răspunsul plantelor la stimuli și medii diferite.

3. APLICAREA INTELIGENTĂ A METODELOR BIOSTATISTICII ÎN CERCETĂRI DE FIZIOLOGIE A PLANTELOR

3.1. Model de analiză biostatistică a datelor privind conținutul de azot, conținutul de proteină, conținutul de fosfor și pentaoxid de fosfor în boabele de soia

În teză sunt expuse unele componente ale sistemului suport pentru studiul compoziției chimice a boabelor de soia (*Glycine max* (L.) Merrill.) pentru analiza unor compuși organici și anorganici cu rol hotărâtor în asigurarea calității boabelor de soia. Plantele au fost cultivate în condiții de câmp pe fond de administrare a preparatelor obținute din surse naturale. Obiectivele componentelor software au fost asistența cercetărilor pentru determinarea gradelor de influență a preparatelor *Reglalg* (preparat de natură algală) și *Biovit* (preparat de natură humică) asupra calității boabelor de soia [216].

În acest experiment seturile de date sunt independente cu dispersii diferite ($\sigma_1^2 \neq \sigma_2^2$) și număr de egal de repetiții ($n_1=n_2=3$), astfel, se va calcula diferența mediilor.

Pentru realizarea sarcinilor expuse, au fost elaborate funcții necesare pentru descifrarea unor parametri biochimici în funcție de genotipul (soiul) cercetat, descrise în capitolul 2.

Analiza statistică și rezultatele obținute sunt prezentate în continuare. Funcțiile elaborate sunt descrise pentru analiza conținutul de azot la soiul *Clavera*. În mod similar se procesează datele și pentru alte specii, soiuri și parametri (Anexa 9).

Conținutul de Azot (N_2)

Se declară variabile pentru introducerea valorilor sub formă de liste:

```
cN2M={5.35,5.39,5.37}, cN2R={6,6.05,6.03}, cN2B={6.17,6.19,6.26};
```

Se creează diagrama care ilustrează media conținutului de azot cu eroarea standard în baza datelor obținute experimental (figura 3.1):

```
ClaveraN2={MeanAround[{5.35,5.39,5.37}], MeanAround[{6,6.05,6.03}],  
MeanAround[{6.17,6.19,6.26}]};  
BarChart[ClaveraN2,ChartElementFunction->"GlassRectangle",  
ChartStyle->"Pastel",ImageSize->400,ChartLabels->{"1","2","3"},  
ChartLegends->{"Martor","Reglalg","Biovit"}]
```

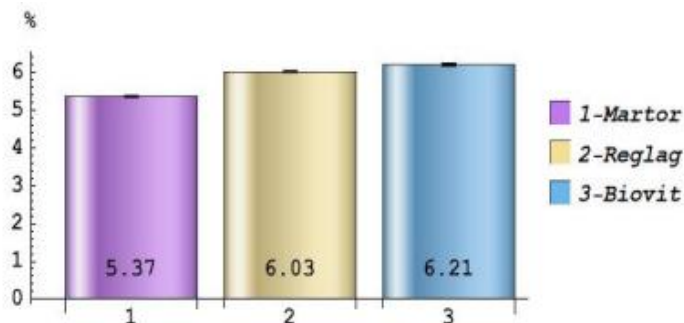


Figura 3.1. Conținutul de azot din boabele de soia (soiul *Clavera*, supus tratamentului foliar cu *Reglalg* și *Biovit*), (%)

Variația conținutului de azot (N₂)

Au fost elaborate funcțiile pentru analiza variației conținutului de azot. Aceasta se materializează în tabelul 3.1:

```
Grid[{"Parametri", "x̄(%)", "σ²", "σ", "CV(%)", "IÎ"},
{"Martor", MeanAround[cN2M], Variance[cN2M], StandardDeviation[cN2M],
StandardDeviation[cN2M]/Mean[cN2M]*100, MeanCI[cN2M]}, {"Reglalg",
MeanAround[cN2R], Variance[cN2R], StandardDeviation[cN2R],
StandardDeviation[cN2R]/Mean[cN2R]*100, MeanCI[cN2R]}, {"Biovit",
MeanAround[cN2B], Variance[cN2B], StandardDeviation[cN2B],
StandardDeviation[cN2B]/Mean[cN2B]*100, MeanCI[cN2B]}] //N]
```

Tabelul 3.1. Variația conținutului de azot (N₂)

Parametri	x (%)	σ ²	σ	CV(%)	CI
Martor	5.37 ± 0.012	0.0004	0.02	0.372	{5.32, 5.42}
Reglalg	6.027 ± 0.015	0.000633	0.0252	0.418	{5.96, 6.09}
Biovit	6.207 ± 0.027	0.00223	0.0473	0.761	{6.09, 6.32}

Rezultatele influenței tratamentului foliar asupra conținutului de azot (N₂)

Se cere să se determine influența biostimulatorilor asupra conținutului de azot din boabele de soia. Rezultatele sunt descrise în tabelul 3.2.

```
Grid[{"Comparatie", "d̄", "s̄d", "t", "lsd0.05", "lsd0.01"},
{"Martor vs Reglalg", medCN2MR = Abs[Mean[cN2M] - Mean[cN2R]],
ercN2MR = √((0.122² + 0.152²)/n), t = (medCN2MR / ercN2MR)*, lsd0.05 = t0.05 * ercN2MR*, lsd0.01 = t0.01 * ercN2MR*},
{"Martor vs Biovit", medCN2MB = Abs[Mean[cN2M] - Mean[cN2B]],
ercN2MB = √((0.122² + 0.272²)/n), t = (medCN2MB / ercN2MB)*, lsd0.05 = t0.05 * ercN2MB*, lsd0.01 = t0.01 * ercN2MB*}];
```

Se determină influența biostimulatorilor asupra conținutului de azot. Din tabelul 3.2 rezultă că valoarea t_{exp} obținută în urma analizei datelor este **59.2** la utilizarea *Reglalg*-ului și **49** la utilizarea *Biovit*-ului ceea ce este mai mare decât valoarea teoretică (t_{teor}), atât la nivel de semnificație de 0.05, cât și de 0.01. Astfel, **se respinge ipoteza nulă**.

Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative. Diferențele mediilor $\bar{d}_r = 0.657$, $\bar{d}_b = 0.837$ sunt mai mari decât valorile lsd obținute conform tabelului 3.2, prin urmare se confirmă rezultatul anterior, **se respinge ipoteza nulă**.

Tabelul 3.2. Rezultatele influenței tratamentului foliar asupra conținutului de azot din boabele de soia (soiul Clavera)

Comparație	$\bar{d}(\%)$	$s_{\bar{d}}$	t_{exp}	$lsd_{0.05}$	$lsd_{0.01}$
Martor vs Reglalg	0.66	0.01	59.2*	0.02*	0.04**
Martor vs Biovit	0.84	0.01	49.*	0.03*	0.06**

*Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

Pentru a confirma ipoteza alternativă se utilizează suplimentar testul p :

$StudentTPValue[59.2, 4] \rightarrow 2.1843 \cdot 10^{-6}$

$StudentTPValue[49, 4] \rightarrow 4.6273 \cdot 10^{-6}$

Rezultatul acestui test este mai mic decât pragul de semnificație. Astfel, **se confirmă** rezultatul obținut prin testul t .

Concluzie: Astfel, analiza statistică a datelor obținute cu referire la conținutul de azot, denotă faptul că preparatul *Reglalg* și preparatul *Biovit*, în dozele utilizate în studiul de față, majorează conținutul de azot din boabele de soia, rezultatele demonstrând influență semnificativă atât la nivel de veridicitate de 95%, cât și de 99%.

Conținutul de Proteină

Se declară variabile pentru introducerea valorilor sub formă de liste:

$cProtM = \{33.44, 33.69, 33.56\}$; $cProtR = \{37.5, 37.81, 37.69\}$; $cProtB = \{38.56, 38.69, 39.13\}$;

Se creează diagrama care demonstrează media conținutului de proteină cu eroarea standard în baza datelor obținute experimental (figura 3.2):

```
ClaveraProt={MeanAround[{33.44,33.69,33.56}], MeanAround[{37.5,
37.81, 37.69}], MeanAround[{38.56, 38.69, 39.13}]}];
BarChart[ClaveraProt, ChartElementFunction->"GlassRectangle",
ChartStyle->"Pastel", ImageSize->400, ChartLabels->{"1", "2", "3"},
ChartLegends-> {"Martor", "Reglag", "Biovit"}, PlotLabel-> "Conținut
Proteină (%)", LabelStyle->{FontSize->11, FontFamily-> "Courier",
Bold}, AxesLabel->{"", "%"}, AspectRatio->0.4, BarSpacing-> 0.2,
LabelingFunction -> Bottom]
```

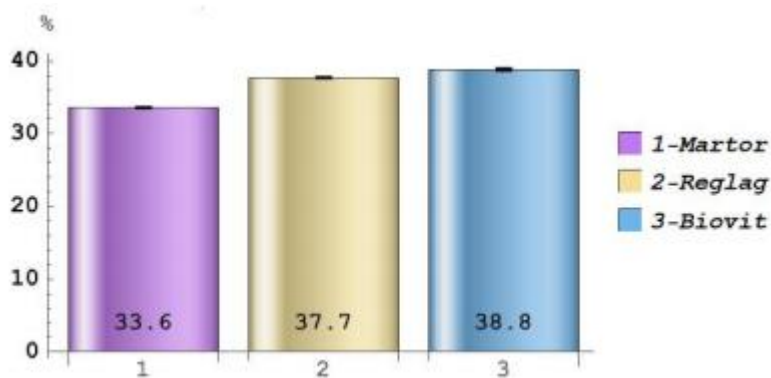


Figura 3.2. Conținutul de proteină din boabele de soia (soiul *Clavera*, supus tratamentului foliar cu *Reglag* și *Biovit*), (%)

Variația conținutului de proteine

Au fost elaborate funcțiile pentru analiza variației conținutului de proteină. Aceasta se materializează în tabelul 3.3.

$Grid[\{"Parametri", "x^{-}(\%)", "\sigma^2", "\sigma", "CV(\%)", "I\hat{I}"}],$

```

{"Martor", MeanAround[cProtM], Variance[cProtM],
StandardDeviation[cProtM], StandardDeviation[cProtM]/ Mean[cProtM] *
100, MeanCI[cProtM]},
{"Reglalg", MeanAround[cProtR], Variance[cProtR],
StandardDeviation[cProtR], StandardDeviation[cProtR]/Mean[cProtR] *
100, MeanCI[cProtR]},
{"Biovit", MeanAround[cProtB], Variance[cProtB],
StandardDeviation[cProtB], StandardDeviation[cProtB]/
Mean[cProtB]*100, MeanCI[cProtB]}} //N,
Frame->All, Spacings->{0.8, 1.5}, Alignment->{Center, Center},
ItemStyle-> {Automatic, Directive [18, Bold]}, Background->{{White,
{White}}, {None, {Green, Yellow}}}, Dividers->{1->True, 1-> True},
Frame->{{True},{True}},FrameStyle->Directive[Thickness[2.`],Blue]

```

Tabelul 3.3. Variația conținutului de proteine în boabele de soia (soiul *Clavera*)

Parametri	$\bar{x}$ (g)	σ^2	σ	CV(%)	CI
Martor	33.56 ± 0.07	0.0156	0.125	0.373	{33.3, 33.9}
Reglalg	37.67 ± 0.09	0.0244	0.156	0.415	{37.3, 38.1}
Biovit	38.79 ± 0.17	0.0892	0.2999	0.77	{38.1, 39.5}

Rezultatele influenței tratamentului asupra conținutului de proteină

Au fost elaborate funcțiile pentru analiza influenței tratamentului asupra conținutului de proteină. Rezultatele sunt prezentate în tabelul 3.4:

```

Grid [{{"Comparatie", "d̄", "sd̄", "t", "lsd0.05", "lsd0.01"},
{"Martor vs Reglalg", medCProtMR = Abs[Mean[CProtM] - Mean[CProtR]],
erCProtMR=√((0.07² + 0.09²)/n), t=(medMR / erCProtMR) *, lsd0.05 =t0.05 *
erCProtMR*, lsd0.01 =t0.01 * erCProtMR*},
{"Martor vs P2", medCProtMB = Abs[Mean[CProtM] - Mean[CProtB]],
erCProtMB=√((0.07² + 0.17²)/n), t=(medMB / erCProtMB) *, lsd0.05 =t0.05 *
erCProtMB*, lsd0.01 =t0.01 * erCProtMB*}]

```

Tabelul 3.4. Rezultatele influenței tratamentului asupra conținutului de proteină în boabele de soia (soiul *Clavera*)

Comparație	$\bar{d}$	$S_{\bar{d}}$	t_{exp}	$lsd_{0.05}$	$lsd_{0.01}$
Martor vs Reglalg	4.1033	0.07	62.3*	0.14*	0.24**
Martor vs Biovit	5.23	0.1	49.3*	0.22*	0.4**

* Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

Se determină influența biostimulatorilor asupra conținutului de proteine. Din tabelul 2.5 rezultă că valoarea t_{exp} obținută în urma analizei datelor este **62.3** la utilizarea *Reglalg*-ului și **49.3** la utilizarea *Biovit*-ului ceea ce este mai mare decât valoarea teoretică, atât la nivel de 0.05, cât și de 0.01. Astfel, **se respinge ipoteza nulă**.

Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative. Diferențele mediilor $\bar{d}_{reglalg} = 4.1033$, $\bar{d}_{biovit} = 5.23$ sunt mai mari decât valorile lsd

obținute conform tabelului 3.4, prin urmare se confirmă rezultatul anterior, *se respinge ipoteza nulă*.

Concluzie: Similar datelor cu referire la conținutul de azot, prelucrarea matematică a datelor confirmă că preparatul *Reglalg* și preparatul *Biovit*, în dozele utilizate în studiul de față, majorează semnificativ conținutul de proteine din boabele de soia, atât la nivel de veridicitate de 95%, cât și de 99%, sugerând faptul că compușii dați acționează asupra metabolismului proteic, între indicii cercetați (conținutul de azot și conținutul de proteină) existând o dependență strânsă.

Conținutul de fosfor (P_2)

Se declară variabile pentru introducerea valorilor sub formă de liste:

$cP2M = \{342.99, 342.9, 343.04\}$; $cP2R = \{536.72, 536.79, 536.86\}$; $cP2B = \{361.75, 361.8, 361.79\}$;

Se creează diagrama care demonstrează media conținutului de fosfor cu eroarea standard în baza datelor obținute experimental (figura 3.3):

```
ClaveraP2={MeanAround[{342.99,342.9,343.04}],
MeanAround[{536.72,536.79,536.86}],
MeanAround[{361.75,361.8,361.79}]}; BarChart[ClaveraP2,
ChartElementFunction->"GlassRectangle",ChartStyle->"Pastel",
ImageSize->400,Chartlabels->{"1","2","3"},Chartlegends->
{"Martor","Reglalg","Biovit"}]
```

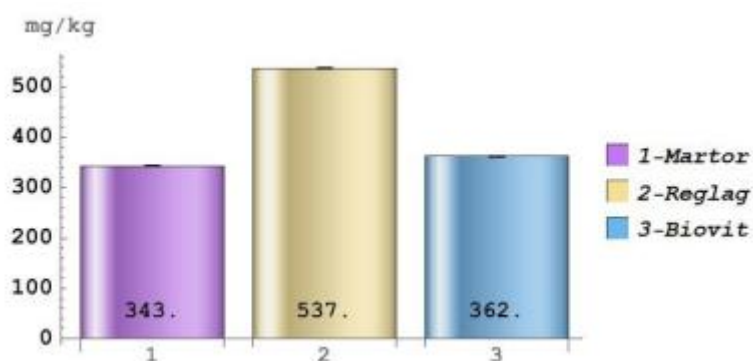


Figura 3.3. Conținutul de fosfor (P_2) din boabele de soia (soiul *Clavera*, supus tratamentului foliar cu *Reglalg* și *Biovit*), (mg/kg)

Variația conținutului de fosfor (P_2)

Au fost elaborate funcții pentru analiza variației conținutului de fosfor. Aceasta se materializează în tabelul 3.5:

```
Grid[{{"Parametri", "x̄(mg/kg)", "σ²", "σ", "CV(%)", "IÎ**"},
{"Martor", MeanAround[cP2M], Variance[cP2M], StandardDeviation[cP2M],
StandardDeviation[cP2M]/Mean[cP2M]*100, MeanCI[cP2M]}, {"Reglalg",
MeanAround[cP2R], Variance[cP2R], StandardDeviation[cP2R],
StandardDeviation[cP2R]/Mean[cP2R]*100, MeanCI[cP2R]}, {"P2",
MeanAround[cP2B], Variance[cP2B], StandardDeviation[cP2B],
StandardDeviation[cP2B]/Mean[cP2B]*100, MeanCI[cP2B]}} //N]
```

Tabelul 3.5. Variația conținutului de fosfor (P₂) din boabele de soia (soiul *Clavera*)

Parametri	$\bar{x}$ (mg/kg)	σ^2	σ	CV(%)	CI
Martor	342.98 ± 0.04	0.0050333	0.070946	0.020685	{342.8, 343.15}
Reglalg	536.79 ± 0.04	0.0049	0.07	0.01304	{536.62, 536.96}
Biovit	361.78 ± 0.015	0.0007	0.026458	0.0073131	{361.71, 361.85}

Rezultatele influenței tratamentului asupra conținutului de fosfor (P₂)

Au fost elaborate funcții pentru analiza influenței tratamentului asupra conținutului de fosfor. Rezultatele se materializează în tabelul 3.6.

```
Grid [{"Comparatie", "d", "sd", "t", "lsd0.05", "lsd0.01"}, {"Martor vs Reglalg", medCP2MR = Abs[Mean[cP2M] - Mean[cP2R]],  
ercP2MR = sqrt((0.042^2 + 0.042^2)/n), t = (medCP2MR/ercP2MR)*, lsd0.05 = t0.05 *  
ercP2MR*, lsd0.01 = t0.01 * ercP2MR*}, {"Martor vs Biovit", medCP2MB = Abs[Mean[cP2M] - Mean[cP2B]],  
ercP2MB = sqrt((0.042^2 + 0.0152^2)/n), t = (medCP2MB/ercP2MB)*, lsd0.05 = t0.05 *  
ercP2MB*, lsd0.01 = t0.01 * ercP2MB*}]
```

Se determină influența biostimulatorilor asupra conținutului de fosfor. Din tabelul 3.6 rezultă că valoarea t_{exp} obținută în urma analizei datelor este **3426.2** la utilizarea Reglalg-ului și **440.15** la utilizarea Biovit-ului ceea ce este mai mare decât valoarea teoretică, atât la nivel de 0.05, cât și de 0.01. Astfel, **se respinge ipoteza nulă**.

Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative (figura 3.6). Diferențele mediilor $\bar{d}_{reglalg} = 193.81$, $\bar{d}_{biovit} = 18.803$ sunt mai mari decât valorile lsd obținute conform tabelului 3.6 prin urmare se confirmă rezultatul anterior, **se respinge ipoteza nulă**.

Tabelul 3.6. Rezultatele influenței tratamentului asupra conținutului de fosfor (P₂) din boabele de soia (soiul *Clavera*)

Comparație	$\bar{d}$	s_d	T	$lsd_{0.05}$	$lsd_{0.01}$
Martor vs Reglalg	193.81	0.03	3426.2*	0.06*	0.12**
Martor vs Biovit	18.803	0.02	762.4*	0.05*	0.09**

* Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

Concluzie: Preparatul *Reglalg* și preparatul *Biovit*, în dozele utilizate în studiul de față, majorează semnificativ conținutul de fosfor din boabele de soia, atât la nivel de veridicitate de 95%, cât și la nivel de veridicitate de 99%, influență mai accentuată asupra parametrului cercetat atestându-se în cazul tratamentului foliar cu preparatul *Reglalg*.

Conținutul de pentaoxid de fosfor (P₂O₅)

Se declară variabile pentru introducerea valorilor sub formă de liste:

```
cP2O5M = {7854.47,7852.41,7855.62}; cP2O5R =
{12290.9,12292.5,12294.1}
cP2O5B = {8284.08,8285.22,8284.99};
```

Se creează diagrama care demonstrează media conținutului de pentaoxid de fosfor cu eroarea standard în baza datelor obținute experimental (figura 3.4):

```
ClaveraP2O5={MeanAround[{7854.47,7852.41,7855.62}],MeanAround[{12290.9,12292.5,12294.1}], MeanAround[{8284.08,8285.22,8284.99}]};
BarChart[ClaveraP2O5,ChartElementFunction->"GlassRectangle",
ChartStyle->"Pastel", ImageSize->450, Chartlabels->{"1", "2","3"},
Chartlegends-> {"Martor","Reglalg", "P2" }, Plotlabel-> "Conținut
Pentaoxid de Fosfor (mg/kg)", AxesLabel->{"","mg/kg"}]
```

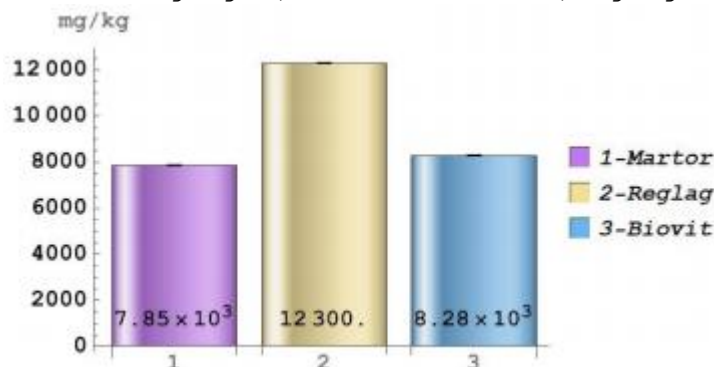


Figura 3.4. Conținutul de pentaoxid de fosfor (P_2O_5) din boabele de soia (soiul Clavera, supus tratamentului foliar cu Reglalg și Biovit), (%)

Variația conținutului de pentaoxid de fosfor (P_2O_5)

Au fost elaborate funcțiile pentru analiza variației conținutului de pentaoxid de fosfor.

Rezultatul se materializează în tabelul 3.7:

```
Grid[{{"Parametri", "x̄ (mg/kg)", "σ²", "σ", "CV(%)", "IÎ**"},
{"Martor", MeanAround[cP2O5M], Variance[cP2O5M],
StandardDeviation[cP2O5M],
StandardDeviation[cP2O5M]/Mean[cP2O5M]*100, MeanCI[cP2O5M]},
{"Reglalg", MeanAround[cP2O5R], Variance[cP2O5R],
StandardDeviation[cP2O5R],
StandardDeviation[cP2O5R]/Mean[cP2O5R]*100, MeanCI[cP2O5R]}, {"P2",
MeanAround[cP2O5B], Variance[cP2O5B], StandardDeviation[cP2O5B],
StandardDeviation[cP2O5B]/Mean[cP2O5B]*100, MeanCI[cP2O5B]}} //N,]]
```

Tabelul 3.7. Variația conținutului de pentaoxid de fosfor (P_2O_5) în boabele de soia (soiul Clavera)

Parametri	$\bar{x}$ (mg/kg)	σ^2	σ	CV(%)	IÎ
Martor	7854.2 ± 0.9	2.645	1.6264	0.020707	{7850.1, 7858.2}
Reglalg	$12.292.5 \pm 0.9$	2.56	1.6	0.013016	{12289, 12296}
Biovit	8284.76 ± 0.35	0.36343	0.60285	0.0072767	{8283.3, 8286.3}

Rezultatele influenței tratamentului asupra conținutului de pentaoxid de fosfor (P_2O_5)

Au fost elaborate funcțiile pentru analiza influenței tratamentului asupra conținutului de pentaoxid de fosfor. Rezultatele sunt prezentate în tabelul 3.8.

```
Grid [{"Comparatie", "d", "sd", "t", "lsd0.05", "lsd0.01"}, {"Martor vs Reglalg", medCP2O5MR = Abs[Mean[cP2O5M] - Mean[cP2O5R]], ercP2O5MR = sqrt((0.92^2 + 0.92^2)/n), t = (medCP2O5MR/ercP2O5MR)*, lsd0.05 = t0.05 * ercP2O5MR*, lsd0.01 = t0.01 * ercP2O5MR*}, {"Martor vs P2", medCP2O5MB = Abs[Mean[cP2O5M] - Mean[cP2O5B]], ercP2O5MB = sqrt((0.92^2 + 0.352^2)/n), t = (medCP2O5MB/ercP2O5MB)*, lsd0.05 = t0.05 * ercP2O5MB*, lsd0.01 = t0.01 * ercP2O5MB*}]
```

Tabelul 3.8. Rezultatele influenței tratamentului asupra conținutului de pentaoxid de fosfor (P₂O₅) în boabele de soia (soiul Clavera)

Comparație	$\bar{d}$	s_d	T	$lsd_{0.05}$	$lsd_{0.01}$
Martor vs Reglalg	4438.3	0.73	6039.8*	1.56*	2.76**
Martor vs Biovit	430.6	0.56	772.3*	1.19*	0.09**

* Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

Se determină influența biostimulatorilor asupra conținutului de pentaoxid de fosfor. Din tabelul 3.8 rezultă că valoarea t_{exp} obținută în urma analizei datelor este **6039.8** la utilizarea Reglalg-ului și **772.3** la utilizarea Biovit-ului ceea ce este mai mare decât valoarea teoretică, atât la nivel de veridicitate de 0.05 , cât și de 0.01 . Astfel, **se respinge ipoteza nulă**.

Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative.

Diferențele mediilor $\bar{d}_{reglalg} = 4438.3$, $\bar{d}_{biovit} = 430.6$ sunt mai mari decât valorile lsd obținute conform tabelului 3.8, prin urmare se confirmă rezultatul anterior, **se respinge ipoteza nulă**.

Concluzie: Preparatul *Reglalg* și preparatul *Biovit*, în dozele utilizate, majorează semnificativ conținutul de pentaoxid de fosfor din boabele de soia, atât la nivel de veridicitate de 95%, cât și la nivel de veridicitate de 99%, influență mai accentuată asupra parametrului cercetat atestându-se în cazul tratamentului foliar cu preparatul *Reglalg*.

În mod similar s-au procesat datele pentru soiurile *Dorința* și *Horboveanca*. Modelul este prezentat în *Anexa 11*.

În concluzie, analiza statistică a datelor denotă faptul că preparatele experimentate influențează substanțial compoziția chimică a boabelor de soia.

Preparatele utilizate influențează diferențiat parametrii studiați la diverse genotipuri de soia. Astfel, administrarea preparatelor *Reglalg* și *Biovit* majorează substanțial conținutul de azot, proteină, fosfor și pentaoxid de fosfor în cazul soiului *Clavera*, rezultatele demonstrând influență semnificativă atât la nivel veridicitate de 95%, cât și la nivel de 99%.

La soiul *Dorința* conținutului de azot indică o influență semnificativă doar la nivel de veridicitate de 95%, însă ambele preparate exercită o influență semnificativă asupra conținutului de proteină, fosfor și pentaoxid de fosfor atât la nivel de veridicitate de 95%, cât și la 99% (anexa 9).

În urma procesării datelor pentru soiul *Horboveanca* rezultă că asupra conținutului de azot (N_2) influențează semnificativ doar preparatul *Biovit*, iar preparatul *Reglalg* nu are influență asupra indicelui cercetat, rezultatele fiind în marja intervalului de încredere (anexa 11). Conținutul de proteină este influențat semnificativ de către preparatul *Biovit*, atât la nivelul de veridicitate de 95%, cât și 99%, în timp ce preparatul *Reglalg* nu modifică conținutul de proteină la genotipul dat.

Conținutul de fosfor (P_2) și pentaoxid de fosfor (P_2O_5) este influențat semnificativ de către ambele preparate atât la nivel de veridicitate de 95%, cât și la 99%, (anexa 11). Administrarea preparatelor de natură algală și humică necesită o abordare diferențiată, luându-se în calcul mai multe variabile, la baza aplicării acestora fiind particularitățile specifice ale reacției soiului la acțiunea compușilor biologic activi.

3.2. Model de analiză statistică a influenței unor substanțe biologic active asupra parametrilor de creștere și dezvoltare a plantelor de soia cultivate pe medii nutritive apoase Knop

Dezvoltarea plantelor pe mediu nutritiv apos se referă la creșterea plantelor într-un mediu în care rădăcinile acestora sunt în contact direct cu o soluție nutritivă lichidă. Rădăcinile plantelor sunt imerse în soluția nutritivă care conține toate substanțele esențiale pentru creșterea și dezvoltarea acestora. Această soluție nutritivă apoasă furnizează plantei toți nutrienții de care are nevoie, inclusiv săruri minerale, oligoelemente și alte substanțe esențiale [217]. Prin cultivarea plantelor într-un mediu nutritiv apos, se elimină necesitatea unui sol tradițional, iar planta poate beneficia de o absorbție mai echilibrată a nutrienților. Acest tip de cultură permite controlul mai precis al nivelului de substanțe nutritive, pH -ului și altor factori de mediu, ceea ce poate duce la o creștere mai rapidă și o producție mai mare a plantelor.

Metoda utilizată oferă oportunități de a studia, în condiții dirijate, efectul unor substanțe/compuși ce posedă activitate biologică asupra plantelor. Astfel, dezvoltarea plantelor pe mediu nutritiv apos este utilizată în diferite aplicații, inclusiv în culturile comerciale, în cercetare științifică și în practicile agricole la scară mică.

În rezultatul procesării datelor obținute în urma desfășurării experimentelor în laborator, utilizând metode biometrice, au fost obținute cunoștințele necesare realizării obiectivelor trasate.

În acest context au fost dezvoltate componente informatice în cadrul sistemului suport pentru studiul creșterii și dezvoltării plantelor. Obiectivele componentelor sistemului suport au fost asistența cercetărilor pentru determinarea gradului de influență a fitohormonului Heteroauxina (0.0001%), preparatelor Reglalg (1:300) și Biovit (1:500) asupra unor parametri morfofiziologici corelați cu creșterea și dezvoltarea plantelor.

Se cere să se determine ipoteza statistică referitoare la diferența mediilor dintre perechile independente ale eșantioanelor. Sunt analizați mai mulți parametri: lungimea rădăcinii și tulpinii, biomasa proaspătă și uscată a rădăcinii și tulpinii. Deoarece observațiile sunt două șiruri independente și dispersiile diferă ($\sigma^2 \neq \sigma^2$), se compară *valoarea diferenței mediilor* dintre perechile de observații ($\bar{d}$); rezultatele tratamentelor cu preparatele respective și un grup de control (*Martorul*) referitor la eficiența acestora.

Volumul eșantionului pentru acest model este 3 repetiții. Gradele de libertate sunt 4 ($n_1 = 3 + n_2 = 3 = 6 - 2$). Valorile teoretice pentru criteriul t sunt: $t_{0.5} = 2.13$; $t_{0.01} = 3.75$.

În continuare sunt expuse variabilele care conțin datele obținute pe cale experimentală, diagrama cu media și eroarea standard, precum și rezultatul influenței biostimulatorilor asupra diferitor părți ale plantei. Modelul complet este prezentat în *Anexa 12*.

Lungimea rădăcinii principale

Se declară variabile pentru introducerea valorilor sub formă de liste:

$RadM = \{7.3, 6.8, 8.2\}$; $RadH = \{10, 9.4, 11\}$; $RadR = \{7.4, 5.9, 4.9\}$; $RadB = \{9.6, 9.7, 10.2\}$.

Se creează diagrama care demonstrează media lungimea rădăcinii principale cu eroarea standard în baza datelor obținute experimental (figura 3.5):

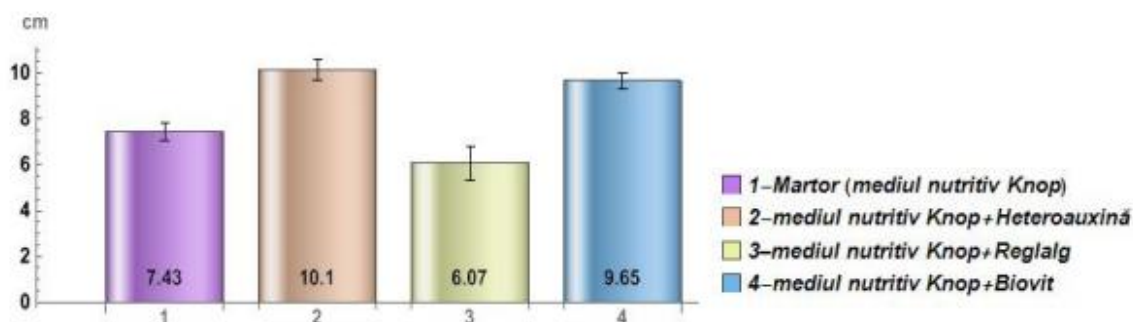


Figura 3.5. Lungimea rădăcinii principale a plantelor cultivate pe medii apoase suplinite cu substanțe biologic active

Rezultatele influenței tratamentului asupra volumului rădăcinii

Au fost elaborate funcții pentru analiza influenței tratamentului asupra volumului rădăcinii. Se compară influența unui preparat cu eșantionul de control (*Martorul*) pentru a afla

diferența mediilor eșantioanelor ($\bar{d}$), eroarea standard a diferențelor (s_d), rezultatul testului t , limita celei mai mici diferențe semnificative (lsd).

Se determină influența biostimulatorilor asupra lungimii rădăcinii. Din tabelul 3.9 rezultă că *Heteroauxina* și *Biovit*-ul influențează semnificativ dezvoltarea rădăcinii. Valoarea testului t_{exp} este 7.3, fiind mai mare decât valoarea t_{teor} pentru 4 grade de libertate ($t_{0.5} = 2.13$; $t_{0.01} = 3.75$). **Ipoteza nulă se respinge.** Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative (tabelul 3.9). Diferențele mediilor $\bar{d} = 2.7$, respectiv 2.22 sunt mai mari decât valorile lsd obținute, prin urmare **se confirmă rezultatul anterior.**

Reglalg-ul influențează creșterea lungimii rădăcinii principale doar la nivel de veridicitate de 95%.

Tabelul 3.9. Rezultatele influenței compușilor biologic activi asupra lungimii rădăcinii

Comparație	$\bar{d}$	s_d	t	$lsd_{0.05}$	$lsd_{0.01}$
Martor vs Heteroauxina	2.7	0.37	7.3*	0.79*	1.39**
Martor vs Reglalg	1.37	0.46	2.9*	0.99*	1.74
Martor vs Biovit	2.22	0.3	7.41*	0.64*	1.12**

* Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

Volumul rădăcinii (ml)

Se declară variabile pentru introducerea valorilor sub formă de liste:

$VolRadM = \{1., 1., 1.\}$; $VolRadH = \{1, 1.5, 1\}$; $VolRadR = \{1, 2, 1.5\}$;
 $VolRadB = \{1.4, 1.6, 1.52\}$

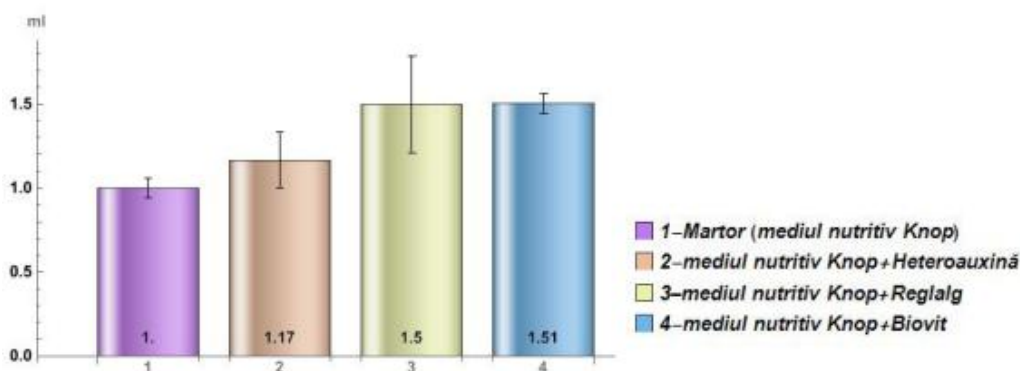


Figura 3.6. Volumul rădăcinii plantelor cultivate pe medii apoase suplinite cu substanțe biologic active

Rezultatele influenței tratamentului asupra volumului rădăcinii

Au fost elaborate funcții pentru analiza influenței tratamentului asupra volumului rădăcinii. Se compară influența unui preparat cu eșantionul de control (*Martorul*) și se află

diferența mediilor eșantioanelor ($\bar{d}$), eroarea standard a diferențelor (s_d), rezultatul testului t , limita celei mai mici diferențe semnificative (lsd).

Se determină influența biostimulatorilor asupra volumului rădăcinii. Din tabelul 3.10 rezultă că valoarea t_{exp} obținută în urma analizei datelor este **0.28** la utilizarea Heteroauxinei, **0.83** – la utilizarea *Reglalg*-ului și **0.88** la utilizarea *Biovit*-ului ceea ce este mai mic decât valoarea teoretică, atât la nivel de 0.05, cât și de 0.01. Astfel **se confirmă ipoteza nulă**.

Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative (tabelul 3.10). Diferențele mediilor $\bar{d}_{heteroauxina} = 0.167$, $\bar{d}_{reglalg} = 0.5$, $\bar{d}_{biovit} = 0.507$ sunt mai mici decât valorile lsd experimentale, prin urmare se confirmă rezultatul anterior. Astfel, se confirmă ipoteza nulă.

Tabelul 3.10. Rezultatele influenței compușilor biologic activi asupra volumului rădăcinii

Comparație	$\bar{d}$	s_d	t	$lsd_{0.05}$	$lsd_{0.01}$
<i>Martor vs Heteroauxina</i>	0.167	0.58	1.6	0.22	.039
<i>Martor vs Reglalg</i>	0.5	0.6	2.92*	0.36*	0.64
<i>Martor vs Biovit</i>	0.507	0.58	10.3*	0.1*	0.18**

Lungimea tulpinii

Se declară variabile pentru introducerea valorilor sub formă de liste:

$TulpM=\{33,27.5,26.5\}$, $TulpH=\{30,33,29\}$, $TulpR=\{25.4,29.3,24.4\}$,
 $TulpB=\{35,30,29.5\}$

Se creează diagrama care demonstrează media lungimii tulpinii cu eroarea standard în baza datelor obținute experimental (figura 3.7):

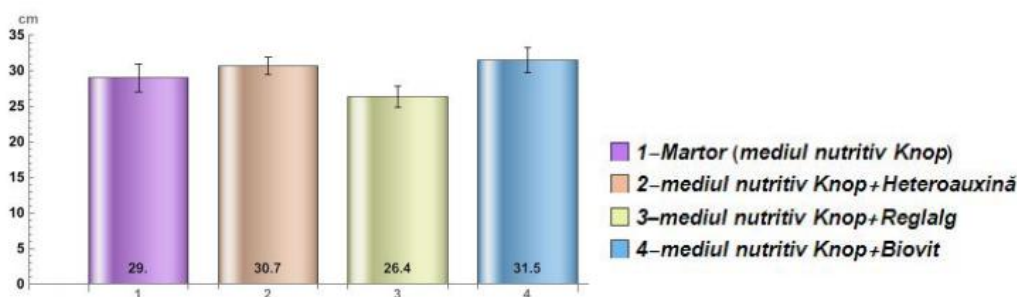


Figura 3.7. Lungimea tulpinii plantelor cultivate pe medii apoase suplinite cu substanțe biologice active

Rezultatele influenței tratamentului asupra lungimii tulpinii

Au fost elaborate funcțiile pentru analiza influenței tratamentului asupra lungimii tulpinii. Se determină influența biostimulatorilor asupra lungimii tulpinii. Din tabelul 3.11 rezultă că valoarea t_{exp} obținută în urma analizei datelor este **0.7** la utilizarea *Heteroauxinei*, **1.05** – la utilizarea *Reglalg*-ului și **0.9** la utilizarea *Biovit*-ului ceea ce este mai mic decât valoarea

teoretică, atât la nivel de 0.05, cât și de 0.01. Astfel, **se confirmă ipoteza nulă**. Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative (tabelul 3.11). Diferențele mediilor $\bar{d}_{heteroauxina} = 1.67$, $\bar{d}_{reglalg} = 2.63$, $\bar{d}_{biovit} = 2.5$ sunt mai mici decât valorile lsd obținute, prin urmare se confirmă rezultatul anterior. Astfel, **se confirmă ipoteza nulă**.

Tabelul 3.10. Rezultatele influenței compușilor biologic activi asupra lungimii tulpinii

Comparație	$\bar{d}$	$s_{\bar{d}}$	t_{exp}	$lsd_{0.05}$	$lsd_{0.01}$
Martor vs Heteroauxina	0.17	0.59	0.28	1.25	2.2
Martor vs Reglalg	0.5	0.6	0.83	1.28	2.25
Martor vs Biovit	0.57	0.58	0.88	1.23	2.2

Biomasa proaspătă a rădăcinii (g)

Se declară variabile pentru introducerea valorilor sub formă de liste:

$BmRadPrM = \{1.56, 1.18, 1.2\}$, $BmRadPrH = \{1.76, 2.2, 0.8\}$,
 $BmRadPrR = \{1.7, 1.82, 1.68\}$, $BmRadPrB = \{2.1, 2.2, 2.4\}$;

Se creează diagrama care demonstrează media biomasei proaspete a rădăcinii cu eroarea standard în baza datelor obținute experimental (figura 3.9):

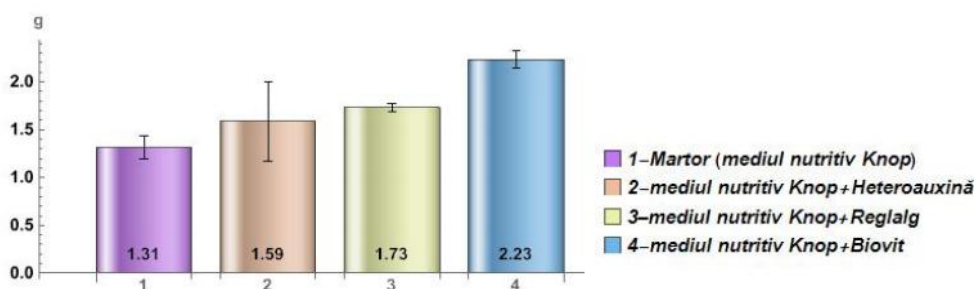


Figura 3.8. Biomasa proaspătă a rădăcinii plantelor cultivate pe medii apoase suplinite cu substanțe biologic active

Influența tratamentului asupra biomasei proaspete a rădăcinii

Au fost elaborate funcții pentru analiza influenței tratamentului asupra biomasei proaspete a rădăcinii. Se compară influența unui preparat cu eșantionul de control (*Martorul*) pentru a afla diferența mediilor eșantioanelor ($\bar{d}$), eroarea standard a diferențelor ($s_{\bar{d}}$), rezultatul testului t , limita celei mai mici diferențe semnificative (lsd).

Se determină influența biostimulatorilor asupra biomasei proaspete a rădăcinii. Din tabelul 2.3 rezultă că *Heteroauxina* nu influențează masa rădăcinii proaspete. valorile obținute sunt în limita intervalului de încredere. Rezultatul procesării datelor este redat în tabelul 3.12. Valoarea testului t_{exp} este **0.655**, fiind mai mică decât valoarea t_{teor} pentru 4 grade de libertate ($t_{0.5}$

= 2.13; $t_{0.01} = 3.75$). *Reglalg*-ul influențează semnificativ biomasei proapete a rădăcinii la nivel de certitudine de 95%. Valoarea testului t_{exp} este 3.32, fiind mai mare decât valoarea t_{teor} . *Biovit*-ul influențează semnificativ biomasei proapete a rădăcinii atât la nivel de certitudine de 95%, cât și 99%. Valoarea testului t_{exp} este 6.13, fiind mai mare decât valoarea t_{teor} . Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative. Diferențele mediilor la eșantionul la care s-a utilizat heteroauxina este $\bar{d}_{heteroauxina} = 0.273$, fiind mai mică decât valoarea *lsd* experimentală care este 0,89 la nivel de certitudine de 95%, cât și la nivel de 99%. valorile obținute sunt în limita intervalului de încredere.

Diferențele mediilor la eșantionul la care s-a utilizat *Reglalg*-ul este $\bar{d}_{reglalg} = 0.42$ care este mai mare decât valoarea *lsd*_{0.05} obținută. Diferențele mediilor la eșantionul la care s-a utilizat *Biovit*-ul este $\bar{d}_{biovit} = 0.92$ care este mai mare decât valorile *lsd* obținute la ambele nivele de certitudine. prin urmare se confirmă rezultatele testului t .

Tabelul 3.11. Rezultatele influenței compușilor biologic activi asupra biomasei proapete a rădăcinii

Comparație	$\bar{d}$	$s_{\bar{d}}$	t	<i>lsd</i> _{0.05}	<i>lsd</i> _{0.01}
Martor vs Heteroauxina	0.273	0.241	1.13	0.51	0.9
Martor vs Reglalg	0.42	0.073	5.75*	0.156*	0.274
Martor vs Biovit	0.92	0.08	10.6*	0.184*	0.325*

* Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

Biomasa uscată a rădăcinii (g)

Se declară variabile pentru introducerea valorilor sub formă de liste:

$BmRadUscM = \{0.04, 0.06, 0.04\}$, $BmRadUscH = \{0.1, 0.14, 0.1\}$,
 $BmRadUscR = \{0.04, 0.08, 0.1\}$, $BmRadUscB = \{0.076, 0.08, 0.085\}$

Se creează diagrama care demonstrează media biomasei proapete a rădăcinii cu eroarea standard în baza datelor obținute experimental (figura 3.9):

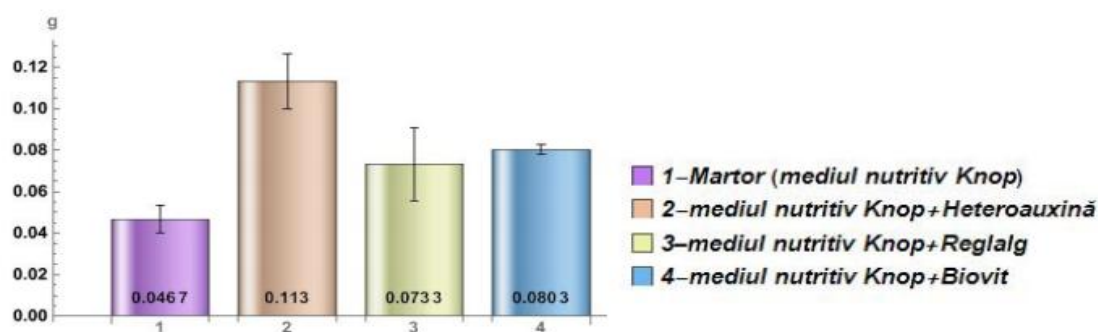


Figura 3.9. Biomasa uscată a rădăcinii plantelor cultivate pe medii apoase suplinate cu substanțe biologic active

Rezultatele influenței tratamentului asupra biomasei uscate a rădăcinii

Au fost elaborate funcții pentru analiza influenței tratamentului asupra biomasei uscate a rădăcinii. Se compară influența unui preparat cu eșantionul de control (Martorul) pentru a afla diferența mediilor eșantioanelor ($\bar{d}$), eroarea standard a diferențelor (s_d), rezultatul testului t , limita celei mai mici diferențe semnificative (lsd).

Se determină influența biostimulatorilor asupra biomasei uscate a rădăcinii. Din tabelul 3.13 rezultă că *Reglalg*-ul nu influențează masa rădăcinii proaspete. valorile obținute sunt în limita intervalului de încredere. Valoarea testului t_{exp} este **1.38**, fiind mai mică decât valoarea t_{teor} pentru 4 grade de libertate ($t_{0.5} = 2.13$; $t_{0.01} = 3.75$). *Heteroauxina* influențează semnificativ masa rădăcinii uscate la ambele nivele de certitudine. Valoarea testului t_{exp} este **4.52**, fiind mai mare decât valoarea t_{teor} . Biovit-ul influențează semnificativ biomasei proapete a rădăcinii atât la nivel de certitudine de 95%, cât și 99%. Valoarea testului t_{exp} este **4.51**, fiind mai mare decât valoarea t_{teor} . Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative. Diferențele mediilor la eșantionul la care s-a utilizat heteroauxina este $\bar{d}_{heteroauxina} = 0.0657$, fiind mai mare decât valoarea lsd experimentală care este **0.08** la nivel de certitudine de 95% și **0.154** la nivel de 99%. Diferențele mediilor la eșantionul la care s-a utilizat *Reglalg*-ul este $\bar{d}_{reglalg} = 0.0267$ care este mai mică decât valoarea $lsd_{0.05}$ obținută **0.08** la nivel de certitudine de 95% și **0.156** la nivel de 99%. Valorile obținute sunt în limita intervalului de încredere, ipoteza nulă se confirmă. Diferențele mediilor la eșantionul la care s-a utilizat *Biovit*-ul este $\bar{d}_{biovit} = 0.0337$ care este mai mare decât valorile lsd obținute la ambele nivele de certitudine (**0.08**, respectiv – **0.152**), prin urmare se confirmă rezultatele testului t .

Tabelul 3.12. Rezultatele influenței compuşilor biologic activi asupra biomasei uscate a rădăcinii

Comparație	$\bar{d}$	s_d	t_{exp}	$lsd_{0.05}$	$lsd_{0.01}$
<i>Martor vs Heteroauxina</i>	0.0667	0.04	1.62	0.08	0.154
<i>Martor vs Reglag</i>	0.0267	0.043	0.64	0.08	0.156
<i>Martor vs Biovit</i>	0.0337	0.04	0.83	0.08	0.152

Concluzie: În urma procesării biostatistice a datelor rezultă că preparatele *Reglalg* și *Biovit* nu influențează acumularea de biomasă uscată a rădăcinii, deși la acumularea de biomasă proaspătă se atestă o influență semnificativă la utilizarea *Biovitului*, atât la nivelul de veridicitate de 95%, cât și la nivelul de veridicitate de 99%. *Reglalgul* influențează acumularea de biomasă proaspătă doar la nivel de veridicitate de 95%. Fitohormonul *Heteroauxina* nu influențează nici biomasa proaspătă și nici uscată, valorile rezultate aflându-se în intervalul de încredere.

Biomasa proaspătă a părții aeriene(g)

Se declară variabile pentru introducerea valorilor sub formă de liste:

$BmTulpPrM=\{2.1, 2.62, 2.2\}$, $BmTulpPrH=\{3.46, 4.8, 1.7\}$,
 $BmTulpPrR=\{2.8, 2.94, 2.92\}$, $BmTulpPrB=\{3.1, 3.3, 3.5\}$;

Se creează diagrama care demonstrează media biomasei proaspete a tulpinii cu eroarea standard în baza datelor obținute experimental (figura 3.10):

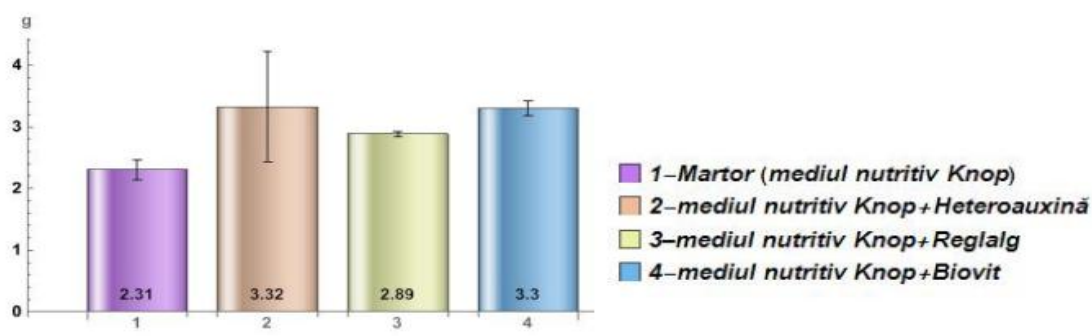


Figura 3.10. Biomasa proaspătă a tulpinii plantelor cultivate pe medii apoase suplinite cu substanțe biologice active

Analiza influenței tratamentului asupra biomasei proaspete a tulpinii

Au fost elaborate funcții pentru analiza influenței tratamentului asupra biomasei proaspete a tulpinii. Se compară influența unui preparat cu eșantionul de control (*Martorul*) pentru a afla diferența mediilor eșantioanelor a biomasei tulpinii proaspete ($\bar{d}$), eroarea standard a diferențelor (s_d), rezultatul testului t, limita celei mai mici diferențe semnificative (lsd).

Se determină influența biostimulatorilor asupra biomasei proaspete a tulpinii. Din tabelul 3.14 rezultă că *Heteroauxina* nu influențează masa tulpinii proaspete, valorile obținute sunt în limita intervalului de încredere. Valoarea testului t_{exp} este **1.92**, fiind mai mică decât valoarea t_{teor} pentru 4 grade de libertate ($t_{0.5} = 2.13$; $t_{0.01} = 3.75$). Deși diferența dintre mediile valorilor este 30.5 %, la eșantionul tratat cu *Heteroauxina* dispersia este mare (2.42 grame), (*Anexa 12*).

Tabelul 3.13. Influența tratamentului asupra biomasei proaspete a tulpinii

Comparație	$\bar{d}$	s_d	t	$lsd_{0.05}$	$lsd_{0.01}$
Martor vs Heteroauxina	0.3	0.52	1.92	1.12	1.98
Martor vs Reglalg	0.58	0.09	6.09*	0.2*	0.36*
Martor vs Biovit	0.993	0.11	8.6*	0.25*	0.43*

* Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

Concluzii: *Reglalg*-ul influențează semnificativ biomasei proaspete a tulpinii la nivel de certitudine de 95%, valoarea testului t_{exp} este **6.09**, fiind mai mare decât valoarea t_{teor} .

Biovit-ul influențează semnificativ biomasei proapete a tulpinii atât la nivel de certitudine de 95%, cât și 99%. Valoarea testului t_{exp} este **8.6**, fiind mai mare decât valoarea t_{teor} .

Rezultatul poate fi confirmat și prin utilizarea analizei celei mai mici diferențe semnificative (tabelul 3.14). Diferențele mediilor la eșantionul la care s-a utilizat heteroauxina este $\bar{d}_{heteroauxina} = 1.01$, fiind mai mică decât valoarea *lsd* experimentală care este 1.12 la nivel de certitudine de 95%, și 1.98 la nivel de 99%. Valorile obținute sunt în limita intervalului de încredere, **se confirmă ipoteza nulă**.

Diferențele mediilor la eșantionul la care s-a utilizat *Reglalg*-ul este $\bar{d}_{reglalg} = 0.58$ care este mai mare decât valoarea *lsd*_{0.05} obținută 0.351. Diferențele mediilor la eșantionul la care s-a utilizat *Biovit*-ul este $\bar{d}_{biovit} = 0.993$ care este mai mare decât valorile *lsd* obținute la ambele nivele de certitudine (0.203, respectiv 0.357). prin urmare se confirmă rezultatele testului *t*.

Biomasa uscată a părții aeriene (g)

Se declară variabile pentru introducerea valorilor sub formă de liste:

$BmTulpUscM = \{0.52, 0.34, 0.26\}$, $BmTulpUscH = \{0.58, 0.76, 0.22\}$,

$BmTulpUscR = \{0.32, 0.42, 0.38\}$, $BmTulpUscB = \{0.35, 0.4, 0.45\}$;

Se creează diagrama care demonstrează media biomasei uscate a părții aeriene cu eroarea standard în baza datelor obținute experimental (figura 3.11).

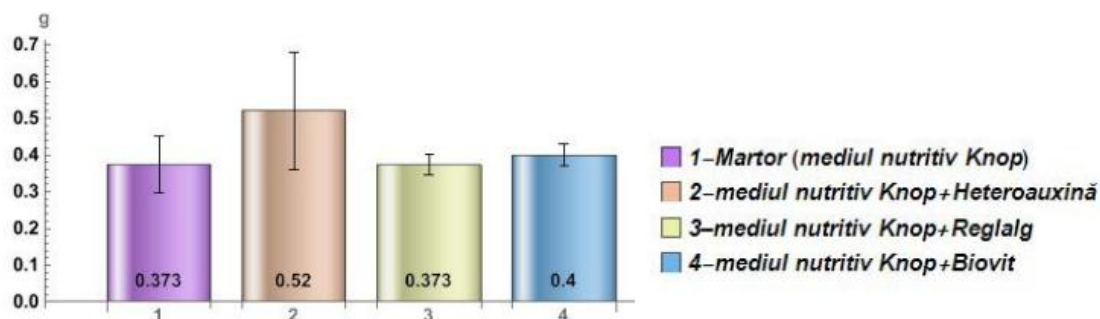


Figura 3.11. Biomasa uscată a tulpinii plantelor cultivate pe medii apoase suplinite cu substanțe biologice active

Analiza influenței tratamentului asupra biomasei uscate a tulpinii

Au fost elaborate funcțiile pentru analiza influenței tratamentului asupra biomasei uscate a tulpinii. Se compară influența unui preparat cu eșantionul de control (*Martorul*) pentru a afla diferența mediilor eșantioanelor ($\bar{d}$), eroarea standard a diferențelor ($s_{\bar{d}}$), rezultatul testului *t*, limita celei mai mici diferențe semnificative (*lsd*).

Se determină influența biostimulatorilor asupra biomasei uscate a tulpinii . Din tabelul 3.15 rezultă că niciun biostimulator nu influențează acest parametru. Astfel **se confirmă ipoteza nulă**, valorile se află în intervalul de încredere.

Tabelul 3.14. Influența tratamentului asupra biomasei uscate a tulpinii

<i>Comparație</i>	$\bar{d}$	s_d	t_{exp}	$lsd_{0.05}$	$lsd_{0.01}$
<i>Martor vs Heteroauxina</i>	0.147	0.1	1.42	0.22	0.39
<i>Martor vs Reglalg</i>	0	0.05	0	0.1	0.18
<i>Martor vs Biovit</i>	0.0267	0.05	0.54	0.1	0.18

Concluzii: Procesarea biostatistică a datelor a permis selectarea parametrilor morfologici asupra cărora a fost exercitate efecte semnificative.

Preparatele influențează substanțial și deferențiat parametrii morfologici ai deferitor genotipuri de soia. A fost influențată dezvoltarea sistemului radicular în ansamblu. Ambele preparate au avut influență semnificativă. Aceasta s-a reflectat în acumularea de biomasă proaspătă și uscată.

Reglalg-ul are o **influență negativă** asupra lungimii rădăcinii și a tulpinii, dar se atestă o **influență semnificativă pozitivă** asupra biomasei proaspete a rădăcinii și tulpinii, la ceilalți indicatori influența se află în marja intervalului de încredere.

Biovit-ul are o **influență semnificativă pozitivă** asupra lungimii și volumului rădăcinii, precum și a biomasei rădăcinii și tulpinii proaspete, la ceilalți indicatori influența se află în marja intervalului de încredere. Modelul este redat în anexa 12.

3.3. Model de analiză biostatistică a datelor referitor la scurgerea electroliților în plantele de soia

Electroliții reprezintă substanțe chimice care, prin dizolvare sau prin topire, se disociază, scindându-se în ioni, și care conduce curentul electric prin transportul acestor ioni [218]. Aceștia au rol în reglarea diferitelor procese care se desfășoară în plante. Electroliții dețin în componența lor ioni liberi mobili și se pot prezenta sub formă fluidă, lichidă sau solidă. În urma dizolvării în apă, are loc separarea electroliților în ioni pozitivi sau negativi. Din punct de vedere nutrițional, electroliții sunt reprezentați de mineralele esențiale: natriu (sodiu, Na), potasiu (K), clor (Cl), calciu (Ca), magneziu (Mg), fier seric (Fe), fosfat și bicarbonat [219]. Ca urmare a influenței unor factori, cum ar fi temperatura, are loc modificarea numărului de electroliți care conduc la producerea unor dezechilibre în plante.

S-au procesat datele obținute experimental referitor la scurgerea electroliților la temperatura de 25° C, 46° C, 48° C, 50° C, 52° C și 100° C. Mai jos sunt expuse graficele care reprezintă rezultatele referitor la influența tratamentului cu biostimulatorul *Humat* pentru toate

patru genotipuri: *Clavera*, *Colina*, *Horboveanca* și *Dorința* (figurile 3.12 și 3.13). De asemenea, sunt redată în tabelul 3.16 rezultatele procesării datelor pentru soiul *Clavera*. Întregul model este reprezentat în anexa 13.

Se declară variabile pentru introducerea valorilor pentru plantele din grupul *Martor* și cele tratate cu *Humat* sub formă de liste:

<i>Martor</i>	<i>Humat</i>
$M25=\{161,132,141, 120\};$	$H25=\{133,119, 141, 163\};$
$M46=\{195,142, 217, 137\};$	$H46=\{230,142,163,156\};$
$M48=\{212,174,260,190\};$	$H48=\{171,137, 204, 165\};$
$M50=\{246,177, 275, 172\};$	$H50=\{227,155, 220, 224\};$
$M52=\{266,226, 304,338,244\};$	$H52=\{254,171, 261, 283\};$

Se creează graficele care demonstrează media scurgerii electroliților la temperatura de 46° C – 52° C pentru toate genotipurile la grupul *Martor* și la grupul tratat cu *Humat*:

```
ListlinePlot[Transpose[{M46,M48,M50,M52}], Ticks→{{{1,"46°"}, {2,"48°"},
{3,"50°"},{4,"52°"}}, Automatic}, Mesh→All, MeshStyle→{PointSize[0.02], Red},
PlotStyle→Thickness[0.01], Plotlabel→HoldForm[Martor],
(*Plotlegends {"46°","48°","50°","52°"},*) AxesLabel→ {"Grade", "Valori"},
Plotlegends→ {"Clavera", "Colina", "Enigma", "Horboveanca", "Albisoara", "Dorinta"},
LabelStyle→Directive[Bold,10], AxesStyle→Directive[RGBColor[0.2,0.4,0.],
AbsoluteThickness[1.5]], AxesOrigin→{.8,100} //N
ListlinePlot[Transpose[{H46,H48,H50,H52}], Ticks→{{{1,"46°"}, {2,"48°"}, {3,"50°"},
{4,"52°"}}, Automatic}, Mesh→All, MeshStyle→{PointSize[0.02], Red},
PlotStyle→Thickness[0.01], Plotlabel → HoldForm[Humat],
(*Plotlegends{"46°","48°","50°","52°"},*) AxesLabel → {"Degrees","Values"},
Plotlegends→{"Clavera", "Colina", "Enigma", "Horboveanca", "Albisoara", "Dorinta"},
LabelStyle→Directive[Bold,10], AxesStyle→ Directive [RGBColor[0.2,0.4,0.],
AbsoluteThickness[1.5]], AxesOrigin→{.8,100} //N
```

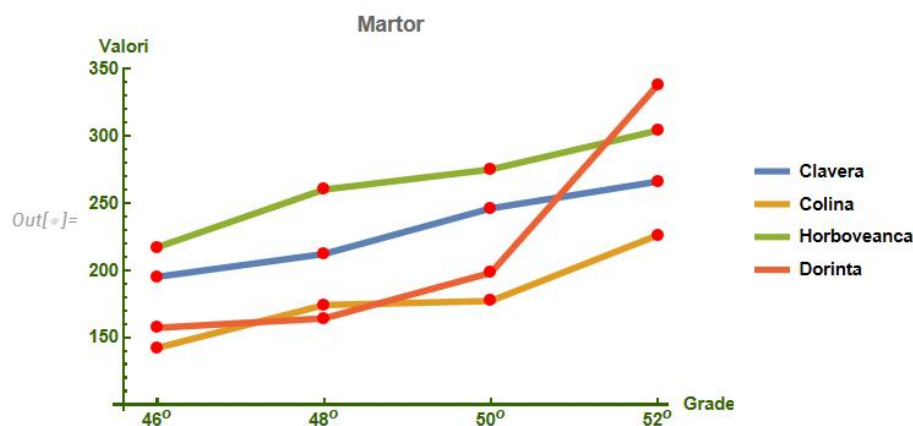


Figura 3.12. Scurgerea electroliților la grupul *Martor*

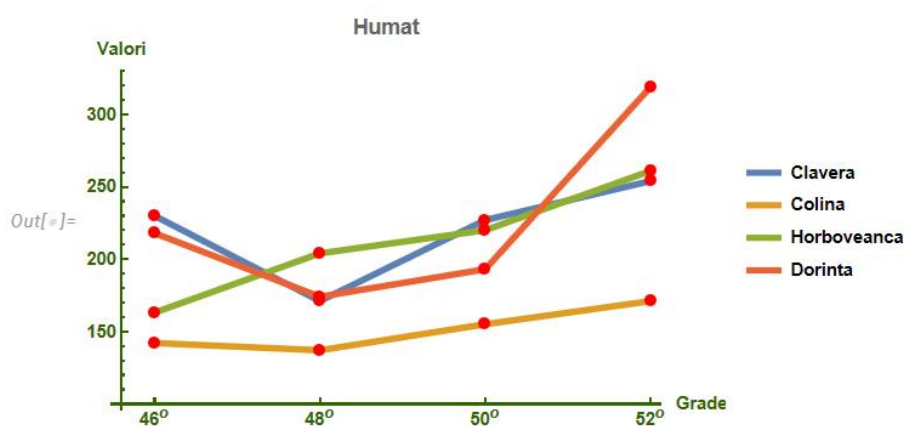


Figura 3.13. Scurgerea electroliților la grupul tratat cu *Humat*

Rezultatele influenței tratamentului asupra scurgerii electroliților

Au fost elaborate funcțiile pentru analiza influenței tratamentului cu *Humat* asupra scurgerii electroliților la soiul *Clavera*. Se compară influența unui preparat cu eșantionul de control (*Martorul*) pentru a afla diferența mediilor eșantioanelor ($\bar{d}$), eroarea standard a diferențelor (s_d^- rezultatul testului t , limita celei mai mici diferențe semnificative la pragul de semnificație de 0.05 și 0.01 (lsd). Rezultatele procesării datelor sunt redată în tabelul 3.16. O influență semnificativă este notată cu *:

```
Grid[{"Comparație", "d-", "S_d-", "t", "lsd0.05", "lsd0.01"},
{"Clavera 25°C Martor vs Humat ", meanClavera25MH=
Abs[Mean[Clavera25]-Mean[Clavera25]], erClavera25MH=sqrt((5^2 + 4^2)/n),
t=(meanClavera25MH/erClavera25MH), lsd0.05=(t5*erClavera25MH),
lsd0.01=(t1*erClavera25MH)}, {"Clavera 46oC Martor vs Humat ",
meanClavera46MH=Abs[Mean[Clavera46]-Mean[Clavera46]], erClavera46MH=
sqrt((6^2 + 1.5^2)/n), t=(meanClavera46MH/erClavera46MH),
lsd0.05=(t5*erClavera46MH), lsd0.01=(t1*erClavera46MH)}, {"Clavera
48oC Martor vs Humat", dif=100-(Mean[Clavera48]*100/
Mean[Clavera48H]), meanClavera48MH= Abs[Mean[Clavera48] -
```

```

Mean[Clavera48H]], erClavera48MH= $\sqrt{(5^2 + 11^2)/n}$ ),  $t=(\text{meanClavera48MH}/\text{erClavera48MH})^*$ ,  $\text{lsd0.05}=(t5*\text{erClavera48MH})^*$ ,  $\text{lsd0.01}=(t1*\text{erClavera48MH})\}$ , {"Clavera 50oC Martor vs Humat", dif=100-(Mean[Clavera50]*100/ Mean[Clavera50H]), meanClavera50MH= Abs[Mean[Clavera50]-Mean[Clavera50H]], erClavera50MH= $\sqrt{(2.3^2 + 6^2)/n}$ ),  $t=(\text{meanClavera50MH}/\text{erClavera50MH})$ ,  $\text{lsd0.05}=(t5*\text{erClavera50MH})$ ,  $\text{lsd0.01}=(t1*\text{erClavera50MH})\}$ , {"Clavera 52oC Martor vs Humat ", dif=100-(Mean[Clavera52]*100/ Mean[Clavera52H]), meanClavera52MH= Abs[Mean[Clavera52]-Mean[Clavera52H]], erClavera52MH= $\sqrt{(0.6^2 + 3^2)/n}$ ),  $t=(\text{meanClavera52MH}/\text{erClavera52MH})$ ,  $\text{lsd0.05}=(t5*\text{erClavera52MH})$ ,  $\text{lsd0.01}=(t1*\text{erClavera52MH})\}$ , {"Clavera 100oC Martor vs Humat ", dif= 100-(Mean[Clavera100]*100/Mean[Clavera100H]), meanClavera100MH= Abs[Mean[Clavera100]-Mean[Clavera100H]], erClavera100MH= $\sqrt{(5^2 + 11^2)/n}$ ),  $t=(\text{meanClavera100MH}/\text{erClavera100MH})$ ,  $\text{lsd0.05}=(t5*\text{erClavera100MH})$ ,  $\text{lsd0.01}=(t1*\text{erClavera100MH})\}$ , Frame→All, Spacings→{0.5,1.2}, Alignment→{Center,Center}, Background→{{White},{White}}, {None,{Green,Yellow}}}, Dividers→ {2→True,2→True}, ItemStyle→{Automatic, Directive [14,Bold]}, Frame→{{True},{True}},FrameStyle→ Directive[Thickness[2], Blue]]//N

```

Tabelul 3.15. Rezultatul scurgerii electroliților

Comparație	$d\bar{r}$	$s_{d\bar{r}}$	t_{exp}	$lsd_{0.05}$	$lsd_{0.01}$
Clavera 25° C Martor vs Humat	-27.7	3.7	7.48*	7.87*	13.9**
Clavera 46° C Martor vs Humat	-24.0	3.57	6.72*	7.61*	13.4**
Clavera 48° C Martor vs Humat	-17.7	6.98	2.5*	14.9*	26.2
Clavera 50° C Martor vs Humat	-19.7	3.71	5.3*	7.9*	13.9
Clavera 52° C Martor vs Humat	-12.3	1.77	6.98*	3.76*	6.62**

* Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

În urma procesării datelor s-au obținut următoarele rezultate:

- la 25° C valoarea t calculată este **7.48** care este mai mare decât valoarea t critică la 4 grade de libertate pentru nivelul de certitudine de 95% - **2.13** și 99% - **3.75**, prin urmare **se respinge ipoteza nulă** la ambele nivele de certitudine;
- la 46° C valoarea t calculată este **6.72** care este mai mare decât valoarea t critică, astfel **se respinge ipoteza nulă** la ambele nivele de certitudine;
- la 48° C valoarea t calculată este **2.5** care este mai mică decât valoarea t critică, doar la nivelul de certitudine de 95%;
- la 50° C valoarea t calculată este **5.3** care este mai mare decât valoarea t critică pentru nivelul de certitudine de 95% - **2.13**, dar mai mică la nivel de 99% - **3.75**, prin urmare **se respinge ipoteza nulă** doar la nivel de certitudine de 95%;
- la 52° C valoarea t calculată este **5.3** care este mai mare decât valoarea t critică, prin urmare **se respinge ipoteza nulă** la ambele nivele de certitudine.

- Limita semnificativă a diferențelor (*lsd*) la 25° C la 95% certitudine reprezintă **7.87** care este mai mica decât media diferențelor celor două eșantioane ($\bar{d} = 27.7$), la 99% nivel de certitudine este **13.9** care la fel este mai mică decât valoarea $\bar{d}$. Prin urmare, se confirmă rezultatul anterior.
- Limita semnificativă a diferențelor (*lsd*) la 46° C la 95% certitudine reprezintă **7.61** care este mai mica decât media diferențelor celor două eșantioane ($\bar{d} = 24$), la 99% nivel de certitudine este **13.4** care la fel este mai mică decât valoarea $\bar{d}$. Prin urmare, se confirmă rezultatul anterior.
- Limita semnificativă a diferențelor (*lsd*) la 48° C la 95% certitudine reprezintă **14.9** care este mai mare decât media diferențelor celor două eșantioane ($\bar{d} = 17.7$), la 99% nivel de certitudine este **26.2** care la fel este mai mare decât valoarea $\bar{d}$. Prin urmare, se confirmă ipoteza nulă conform rezultatului anterior.
- Limita semnificativă a diferențelor (*lsd*) la 50° C la 95% certitudine reprezintă **7.9** care este mai mica decât media diferențelor celor două eșantioane ($\bar{d} = 19.7$), la 99% nivel de certitudine este **13.9** care este mai mare decât valoarea $\bar{d}$. Prin urmare, se confirmă rezultatul anterior.
- Limita semnificativă a diferențelor (*lsd*) la 52° C la 95% certitudine reprezintă **3.76** care este mai mica decât media diferențelor celor două eșantioane ($\bar{d} = 12.3$), la 99% nivel de certitudine este **6.62** care la fel este mai mică decât valoarea $\bar{d}$. Prin urmare, se confirmă rezultatul anterior.

Notă: Valorile negative indică scurgerea a mai puțini electroliți ceea ce reprezintă un rezultat pozitiv.

Rezultatele obținute cu testul *t* și *lsd* se confirmă calculând valoarea *p*. Această valoare oferă semnificația statistică a rezultatelor testării unei ipoteze pentru testarea ipotezelor statistice pentru a confirma respingerea ipotezei nule.

Pentru a calcula valoarea *p* se utilizează funcția Wolfram Mathematica: *StudentTPValue[t, df]*, unde *t* reprezintă valoarea calculată a testului *t*, *df* reprezintă gradele de libertate.

```
StudentTPValue[7.48, 4] => OneSidedPValue -> 0.000854
StudentTPValue[6.72, 4] => OneSidedPValue -> 0.00128
StudentTPValue[2.53, 4] => OneSidedPValue -> 0.0323
StudentTPValue[5.3, 4] => OneSidedPValue -> 0.00304
StudentTPValue[6.98, 4] => OneSidedPValue -> 0.00111
```

Valorile *p* obținute confirmă respingerea ipotezei nule obținute din testele *t* și *lsd*.

Modelul de analiză biostatistică a datelor privind scurgerea electroliților la plantelor de soia este prezentat în anexa 13.

3.4. Model de analiză biostatistică a termotoleranței la plantele de soia

Răspunsul plantelor la stresul termic poate varia în funcție de mai mulți factori, inclusiv specia și soiul plantei, durata și intensitatea stresului termic, stadiul de dezvoltare al plantelor și capacitatea lor intrinsecă de a tolera temperaturi ridicate. Semințele colectate în câmp au fost germinate și tratate radiclele timp de 1 oră cu biostimulatorul *Reglalg* în doza de 300 μ l și apoi supuse stresului termic timp de 10 minute la temperaturi de 46° C. În baza datelor experimentale obținute a fost dezvoltat un model de analiză statistică pentru a aprecia influența biostimulatorului *Reglalg* asupra rezistenței plantelor la stresul termic (Anexa 14).

Dinamica dezvoltării sistemului radicular

Au fost realizate măsurări în trei repetiții. Lungimea inițială a raducolei pe repetiții a fost de 3, 2, 2 cm pentru a verifica exact influența preparatului asupra termotoleranței.

Lungimea rădăcinii principale

Se declară variabile pentru analiza variației lungimii rădăcinii principale pentru grupul de control netratat termic (*Martorul*), grupul de control tratat termic (*Martor Tratat*), grupul netratat termic, dar tratat cu *Reglalg*, grupul tratat termic și suplinit cu *Reglag*.

```
RadPrMartor = {14., 14.5, 14.}; RadPrMartorTrat = {2.5, 1.5, 1.5};  
RadPrRegNetr = {10., 12., 13.}; RadPrRegTrat={2.5,1.5,1.5};
```

Se creează diagrama care demonstrează media lungimii rădăcinii și eroarea standard (fig. 3.14):

```
LRad = {MeanAround[RadPrMartor], MeanAround[RadPrMartorTrat],  
MeanAround[RadPrRegNetr], MeanAround[RadPrRegTrat]}; BarChart[LRad,  
ChartElementFunction->"GlassRectangle", ChartStyle->"Pastel", ImageSize->  
450, ChartLabels {"1", "2", "3", "4"}, ChartLegends {"Martor", "Martor  
Tratat", "Reglag Netratat", "Reglag Tratat"}, PlotLabel->"Lungimea  
rădăcinii principale", LabelStyle->{14, Bold}, " " AxesLabel->{"", "cm"},  
AspectRatio->0.4, BarSpacing->0.3, LabelingFunction->Bottom]
```

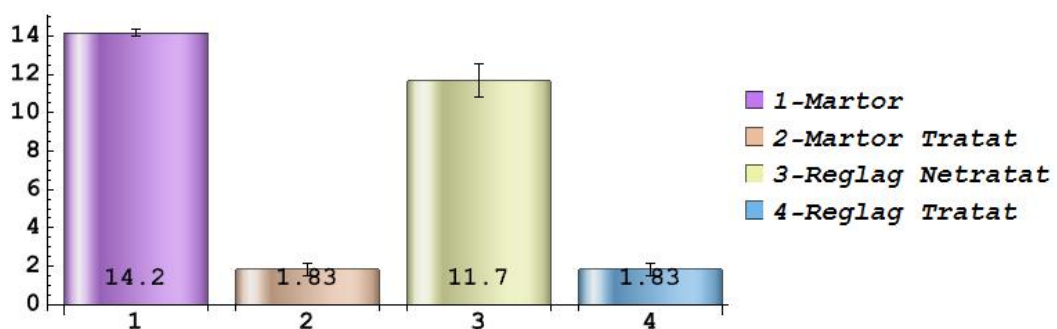


Figura 3.14. Lungimea rădăcinii

Au fost elaborate funcțiile pentru analiza variației lungimii rădăcinii principale. Rezultatul se materializează în tabelul 3.17:

```
Grid[{"Parametri", "Media", "Dispersia", "Abaterea Standard",  
"Coeficientul de variatie(%)", "Intervalul de încredere"}, {"Martor
```

```

Netratat", MeanAround[RadPrMartor], Variance[RadPrMartor],
StandardDeviation[RadPrMartor], StandardDeviation[RadPrMartor] /
Mean[RadPrMartor]*100, MeanCI[RadPrMartor]},
{"Martor Tratat", MeanAround[RadPrMartorTrat],
Variance[RadPrMartorTrat], StandardDeviation[RadPrMartorTrat],
StandardDeviation[RadPrMartorTrat] / Mean[RadPrMartorTrat]*100,
MeanCI[RadPrMartorTrat]},
{"Reglalg Netratat", MeanAround[RadPrRegNetr],
Variance[RadPrRegNetr],
StandardDeviation[RadPrRegNetr], StandardDeviation[RadPrRegNetr]/Mean
[RadPrRegNetr]*100, MeanCI[RadPrRegNetr]},
{"Reglalg Tratat", MeanAround[RadPrRegTrat], Variance[RadPrRegTrat],
StandardDeviation[RadPrRegTrat], StandardDeviation[RadPrRegTrat] /
Mean[RadPrRegTrat]*100, MeanCI[RadPrRegTrat]} //N, Frame->All,
Spacings-> {0.5,1.5}, Alignment -> {Center, Center}, ItemStyle ->
{Automatic, Directive [11,Bold]}, Background->{{White, {White}},
{None,{,}}, Dividers->{1->True, 1->True}, Frame->{{True},
{True}}, FrameStyle-> Directive[Thickness[1.`,`],]]

```

Tabelul 3.16. Variația lungimii rădăcinii

Parametri	$\bar{X}(\text{cm})$	σ^2	σ	CV*	$I\hat{I}^{**}$
Martor Netratat	14.17±0.17	0.08	0.29	2.04	{13.45,14.88}
Martor Tratat	1.83±0.33	0.33	0.58	31.49	{0.4,3.27}
Reglalg Netratat	11.7±0.9	2.33	1.53	13.09	{7.87,15.46}
Reglalg Tratat	1.83±0.33	0.33	0.58	31.49	{7.87,15.46}

* Coeficientul de Variație

** Intervalul de Încredere

Au fost elaborate funcții pentru determinarea influenței tratamentului cu compusul *Reglalg* asupra tratamentului termic al plantelor. Rezultatul este prezentat în tabelul 3.18:

```

Grid[{"Comparație", "d-", "Sd-", "t", "lsd0.05", "lsd0.01"},
{"Martor Netratat vs Martor Tratat", meanMnMt = Abs[Mean[RadPrMartor] -
Mean[RadPrMartorTrat]], erMnMt = Sqrt[0.17^2 + 0.33^2], t =
(meanMnMt/erMnMt), lsd05 = (t5*erMnMt),
lsd01 = (t1*erMnMt)}, {"Reglalg Netratat vs Reglalg Tratat",
meanRnRt = Abs[Mean[RadPrRegNetr] - Mean[RadPrRegTrat]], erRnRt =
Sqrt[0.9^2 + 0.33^2], t = (meanRnRt/erRnRt), lsd05 = (t5*erRnRt),
lsd01 = (t1*erRnRt)}, {"Martor Netratat vs Reglalg Netratat", dif =
100 - (Mean[RadPrMartor] * 100 / Mean[RadPrRegNetr]), meanMnRn =
Abs[Mean[RadPrMartor] - Mean[RadPrRegNetr]], erMnRn = Sqrt[0.17^2 +
0.9^2], t = (meanMnRn/erMnRn), lsd05 = (t5*erMnRn), lsd01 =
(t1*erMnRn)},
{"Martor Tratat vs Reglalg Tratat", meanMtRt =
Abs[Mean[RadPrMartorTrat] - Mean[RadPrRegTrat]],
erMtRt = Sqrt[0.33^2 + 0.33^2], t = (meanMtRt/erMtRt), lsd05 =
(t5*erMtRt), lsd01 = (t1*erMtRt)}, {"Martor Netratat vs Reglalg
Tratat", dif = 100 - (Mean[RadPrMartor]*100/ Mean[RadPrRegTrat]),
meanMnRt = Abs[Mean[RadPrMartor] - Mean[RadPrRegTrat]],
erMnRt = Sqrt[0.17^2 + 0.33^2], t = (meanMnRt/erMnRt), lsd05 =
(t5*erMnRt),
lsd01 = (t1*erMnRt)}, {"Martor Tratat vs Reglalg Netratat",
meanMtRn = Abs[Mean[RadPrMartorTrat] - Mean[RadPrRegNetr]], erMtRn =

```

$\text{Sqrt}[0.33^2 + 0.9^2]$, $t = (\text{meanMtRn}/\text{erMtRn})$, $\text{lsd05} = (t5*\text{erMtRn})$, $\text{lsd01} = (t1*\text{erMtRn})\}}\}$, $\text{Spacings} \rightarrow \{0.5, 1.2\}$, $\text{Alignment} \rightarrow \{\text{Center}, \text{Center}\}$, $\text{Background} \rightarrow \{\{\text{White}, \{\text{White}\}\}, \{\text{None}, \{\text{Green}, \text{Yellow}\}\}\}$, $\text{Frame} \rightarrow \text{All}$, $\text{FrameStyle} \rightarrow \text{Directive}[\text{Thickness}[1], \text{Blue}]]$;

Tabelul 3.17. Influența tratării termice asupra rădăcinii

Comparație	$\bar{d}$	$S_{\bar{d}}$	t_{exp}	$\text{lsd}_{0.05}$	$\text{lsd}_{0.01}$
Martor Netratat vs Martor Tratat	-12.33	0.21	57.5*	0.45*	0.8
Reglalg Netratat vs Reglalg Tratat	-9.83	0.55	17.8*	1.18*	2.08**
Martor Netratat vs Reglalg Netratat	-2.5	0.52	4.73*	1.13*	1.98**
Martor Tratat vs Reglalg Tratat	0	0.27	0	0.57	1.01
Martor Netratat vs Reglalg Tratat	-12.33	0.21	57.5*	0.45*	0.8*
Martor Tratat vs Reglalg Netratat	2.5	0.55	4.52*	1.18*	2.07**

*Influență semnificativă la nivel de 95%

** Influență semnificativă la nivel de 99%

În mod similar au fost procesate datele pentru alți parametri și alte soiuri. Modelul complet este prezentat în anexa 14.

În concluzie, Utilizarea compusului Reglalg a influențat pozitiv biomasa proaspătă și uscată a rădăcinii, conținutul de apă în rădăcină, lingimea tulpinii și conținutul de apă în tulpină. Asupra lungimii rădăcinii, a hipocotilului și a epicotilului a avut o influență negativă. La ceilalți indicatori rezultatele sunt în intervalul de încredere. Utilizarea preparatului Reglalg are influență pozitivă la plantele netratate termic în comparație cu martorul netratat termic, precum și cu martorul tratat termic.

În cadrul cercetărilor au mai fost elaborate și alte modele. Aceste modele sunt descrise în anexele 15 – 17:

- Productivitatea culturii de soia în baza datelor experimentale obținute în urma experimentelor desfășurate în anul 2014 (anexa 15);
- Cantitatea de apă și ulei în boabele de soia în baza datelor experimentale obținute în urma experimentelor desfășurate în anul 2021 (anexa 16);
- Influența unor substanțe biologice active asupra creșterii plantelor de soia în baza datelor experimentale obținute în urma experimentelor desfășurate în anii 2020-2021 (anexele 17-18);
- Determinarea polipeptidelor în rădăcinile și frunzele de soia (anexa 19).

3.5. Concluzii la capitolul 3

A fost realizată o submulțime a limbajului domeniului de cercetare „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor”.

Au fost dezvoltate modele generice orientate pe familii de probleme. În baza acestora au fost construite modele specifice pentru rezolvarea problemelor privind determinarea influenței unor substanțe biologice active asupra creșterii și dezvoltării unor fenotipuri de soia, fenologia acestor fenotipuri (termotoleranța și scurgerea electroliților).

Modelele generice și cele specifice folosesc oportunitățile inteligente ale sistemelor instrumentale *Wolfram Mathematica*, și sistemelor de gestiune a bazelor de date nerelaționale (de tip graf) *Neo4j*.

Au fost efectuate calcule pe datele experimentale obținute în Laboratorul „Securitate biologică” din cadrul Departamentului „Biologie și Ecologie”, facultatea de Biologie și Geștiințe.

Rezultatele obținute în urma calculelor au demonstrat corectitudinea modelelor elaborate. Modelele elaborate au fost folosite în activitățile de cercetare a Laboratorului „Securitate biologică”. Modelele elaborate pot fi utilizate și în alte laboratoare de cercetare. De asemenea, rezultatele obținute pot fi implementate de către agricultori.

A fost elaborat un Drept de Autor „Model de analiză statistică a datelor privind conținutul de azot și fosfor în plante”. Seria O Nr. 7091 din 08.12.2021.

Rezultatele au fost prezentate la Expoziții Internaționale de Invenții și Inovații:

1. International Exhibition InventCor, 2021 și 2022. Medalie de aur (anexele 3, 6).
2. Salonul Internațional al Inovării și Cercetării Științifice Studentești ”Cadet INOVA’22”, Academia Forțelor Terestre „Nicolae Bălcescu”.
3. Salonul European de Creativitate și Inovație „Euroinvent”, Iași, 2022 - Medalie de argint și 2023 - Diplomă de Excelență (anexele 4, 7) .
4. Salonul Internațional de Invenții, Inovații „Traian Vuia”. Timișoara, 08-10 octombrie 2022 / Medalie de bronz (anexa 5).
11. De asemenea, rezultatele obținute în cadrul cercetărilor au fost validate prin raportate la unele Conferințe:
 1. Conferința studentască „StudMath-IT” University "Aurel Vlaicu" din Arad, ediția13, 2020 și ediția 14, 2021 (anexele 1, 2);
 2. Conferința științifică națională a doctoranzilor „Metodologii contemporane de cercetare și evaluare” 22-23 aprilie 2021, Chișinău, 2022,
 3. Conferința Internațională „Mathematics & IT: Research and education (MITRE – 2021)”, Chișinău, 2021.

4. Conferința științifică națională cu participare internațională „Integrare prin Cercetare și Inovare”, 10.11.2022.

Rezultatele cercetărilor au fost înalt apreciate la Saloanele Internaționale de Creativitate și Inventică cu două medalii de aur, una de argint, una de bronz și o Diplomă de Excelență.

În concluzie, analiza biostatistică a datelor denotă faptul, că preparatele experimentate influențează în mod diferit. Fiecare parametru studiat: calitatea boabelor, proprietățile fiziologice și genetice ale plantei (lungimea rădăcinii, a tulpinii, biomasa rădăcinii și tulpinii proaspete și uscate, conținutul de apă și ulei, scurgerea electroliților, productivitatea) ș.a. De asemenea, a fost analizat comportamentul genotipurilor cultivate în diferite medii, cum ar fi câmpul, mediile apoase *Knop*, germinarea semințelor în *cutii Petri*.

Biostimulatorii utilizați influențează substanțial compoziția chimică a boabelor de soia. Acestea influențează diferențiat conținutul de azot, proteină, fosfor și pentaoxid de fosfor care crește la genotipul *Clavera* și *Dorința*, rezultatele demonstrând proprietățile fiziologice ale semnificativă. La genotipul *Horboveanca* rezultă că conținutul de fosfor (P_2) și pentaoxid de fosfor (P_2O_5) este influențat semnificativ de către ambele preparate în timp ce asupra conținutul de azot (N_2) și proteină influențează semnificativ doar preparatul *Biovit*.

Utilizarea acestor biostimulatori la cultivarea plantelor în medii apoase *Knop* are o influență diferită asupra parametrilor analizați. Astfel în rezultatul procesării datelor s-a constatat că: ***Heteroauxina*** are o **influență pozitivă** asupra lungimii rădăcinii și a biomasei uscate a rădăcinii, la ceilalți indicatori influența se află în marja intervalului de încredere.

Tratarea plantelor de soia cu preparatul *Reglalg*-ul are o influență semnificativă asupra biomasei părții aeriene proaspete și uscate. S-a atestat o influență negativă asupra lungimii rădăcinii și a tulpinii și la procesarea termică. La ceilalți indicatori influența se află în marja intervalului de încredere.

Tratarea plantelor de soia cu preparatul Biovit-ul are o influență semnificativă asupra lungimii și volumului rădăcinii, precum și a biomasei rădăcinii și tulpinii proaspete, la ceilalți indicatori influența se află în marja intervalului de încredere.

4. DEPOZITUL DE DATE GRAF: EVIDENȚIEREA CONEXIUNILOR PENTRU OPTIMIZAREA PROCESELOR AGRICOLE

4.1. Grafurile pentru procesarea datelor experimentale și a datelor meteorologice

Pentru procesarea datelor rezultate în urma experimentelor au fost realizate grafuri în SGBD Neo4j prin intermediul limbajului Cypher. Pentru simplificarea procesării datelor a fost elaborat câte un graf pentru fiecare experiment. Aceste grafuri sunt prezentate în anexele 20 - 27.

Un modul important pentru planificarea și gestionarea agriculturii îl reprezintă procesarea datelor meteorologice cu toate datele și informațiile relevante. Sunt incluse datele din perioada anilor 2002 – 2022 obținute de la Serviciul HidroMeteorologic de Stat și Biroul Național de Statistică al Republicii Moldova [210, 220].

Au fost definite diferite tipuri de noduri și relații între ele pentru a modela datele meteorologice (anexa 28). Graful pentru procesarea datelor meteorologice include date despre:

- Cantitatea lunară de precipitații (tipul de noduri *MYRC*), precum și cantitatea totală anuală de precipitații (tipul de nod *YRC*);
- Temperaturile medii lunare și temperaturile medii și mediane anuale (Tipurile de noduri *MYTC*, *YTC*);
- Coeficientul hidrotermic lunar și anual (Tipurile de noduri *MYHTCc*, *YHTCc*);
- Umiditatea solului la diferite nivele (Tipurile de noduri *MSMC*, *YMSC*);
- Viteza medie lunară a vântului (Tipurile de noduri *MYWSC*, *YWSC*);
- Cantitatea medie de roada la hectar (Tipul de noduri *Prod*) .

Definirea nodurilor și relațiilor în *Neo4j* pentru datele menționate

Cantitatea lunară de precipitații:

Fiecare lună este reprezentată de un nod de tip *MYRC*, cu proprietăți cum ar fi nume (lună, an) și cantitatea de precipitații. Se definesc variabile pentru cantitatea lunară de precipitații în diferite luni și ani. De exemplu:

Variabila „*jan2rC*” desemnează luna ianuarie 2002 cantitatea de precipitații în zona Centru. Este definit un nod de tipul *MYRC* - *Month Year Rainfall Center* - Precipitațiile lunare în zona Centru. Acest nod conține proprietatea „*name*” care definește numele nodului - '*Jan 2002 Rainfall Center*' și proprietatea „*rainfall*” care determină cantitatea de precipitații. În luna ianuarie 2002 au fost înregistrate 17 mm de precipitații, etc.

```
MERGE(jan2rC:MYRC{name:'Jan 2002 Rainfall Center', rainfall: 17});  
MERGE(mar6rC:MYRC{name:'Mar 2006 Rainfall Center', rainfall: 89});  
MERGE(may10rC:MYRC{name:'May 2010 Rainfall Center', rainfall: 69});
```

```
MERGE(jul23rC:MYRC{name:'Jul 2023 Rainfall Center'});
MERGE(aug24rC:MYRC{name:'Aug 2024 Rainfall Center'});
```

Similar cu nodurile de tip *MYRC*, a fost definit tipul de noduri *MYTC* (*Month, Year, Temperature, Center*) pentru temperaturile medii lunare.

Se definesc variabile pentru mediile lunare ale temperaturilor în diferite luni și ani de exemplu: jun12tC (temperatura aerului în luna iunie anul 2012 în zona Centru). Aceste noiduri au proprietatea *name* care determină numele nodului (de exemplu: "Jun 2012 Temp Center" și proprietatea *temp* care primește valoarea temperaturii medii din luna respectivă. Nodurile care reprezintă ultimii ani nu conțin proprietatea *temp* pentru testarea modelului.

Exemple: în iunie 2012 temperatura medie a fost de 23,3° C, în ianuarie 2014 temperatura medie a fost de -1,9 ° C, etc.

```
//Temperature MYTC
MERGE(jun12tC: MYTC{name:'Jun 2012 Temp Center', temp: 23.3});
MERGE(jan14tC: MYTC{name:'Jan 2014 Temp Center', temp: -1.9});
MERGE(aug18tC: MYTC{name:'Aug 2018 Temp Center', temp: 24.6});
MERGE(mar21tC: MYTC{name:'Mar 2021 Temp Center', temp: 3.8});
MERGE(may25tC: MYTC{name:'May 2025 Temp Center'});
```

În continuare se definesc nodurile cu datele referitor la coeficientul hidrotermic lunar și anual cu proprietățile acestora: Numele nodului (luna, anul, fenomenul și regiunea), valoarea coeficientului hidrotermic și tipul de secetă.

În instrucțiunile următoare sunt redată unele noduri. Valoarea coeficientului hidrotermic pe lună sunt definite prin variabila *htc*. Ultimele noduri nu conțin proprietatea tip secetă (*drought*) deoarece vor fi folosite pentru a testa și a verifica rezultatele modelului.

```
MERGE(apr2chtC:MYCHTC{name:'Apr 2002 HTC Center', htc:1.0,
drought:0}),
(jul6chtC:MYCHTC{name:'Jul 2006 HTC Center', htc:0.8, drought:1}),
(jun23chtC:MYCHTCp{name:'Jun 2023 HTC Center'});
```

Viteza medie lunară a vântului:

Se definesc variabile pentru viteza vântului lunar în diferite luni și ani.

Exemple: în martie 2005 viteza vântului a fost de 2,6 m/s, în februarie 2008 viteza vântului a fost de 2,0 m/s, etc.

```
MERGE(mar05wc: MWSC{name:'Mar 2005 Wind Center', wind: 2.6});
MERGE(feb08wc: MWSC{name:'Feb 2008 Wind Center', wind: 2.0});
MERGE(apr16wc: MWSC{name:'Apr 2016 Wind Center', wind: 2.9});
MERGE(oct18wc: MWSC{name:'Oct 2018 Wind Center', wind: 2.7});
MERGE(sep25wc: MWSC{name:'Sep 2025 Wind Center'});
```

Înregistrarea datelor colectate pe ani

În afară de datele meteorologice pe luni sunt definite și datele pe ani, inclusiv cantitatea anuală de precipitații, temperatura medie anuală, coeficientul hidrotermic și viteza medie a vântului. De asemenea calitatea aerului și productivitatea sunt determinate pe ani.

Sunt definite variabilele pentru noduri care semnifică fenomenul, anul, zona geografică: r2c, t10c, htc5c, etc.

Tipurile de noduri sunt definite astfel: Fenomen, An, Zonă geografică (YRC, YTC, etc.)

```
MERGE (r2c:YRC{name:'Rainfall 2002 Center', rainfall:604});  
MERGE (r10c:YRC{name:'Rainfall 2010 Center', rainfall:734});  
MERGE (r18c:YRC{name:'Rainfall 2018 Center', rainfall:609});  
MERGE (r20c:YRC{name:'Rainfall 2020 Center', rainfall:562});  
MERGE (r23c:YRC{name:'Rainfall 2021 Center'});
```

Media și mediana temperaturilor anuale

```
MERGE (t10c:YTC{name:'Temp 2010 Center', temp:10.6, tmedian:10.7});  
MERGE (t15c:YTC{name:'Temp 2015 Center', temp:12.0, tmedian:10.1});  
MERGE (t19c:YTC{name:'Temp 2019 Center', temp:12.2, tmedian:11.4});  
MERGE (t20c:YTC{name:'Temp 2020 Center', temp:12.7, tmedian:13.1});  
MERGE (t24c:YTC{name:'Temp 2024 Center'});
```

Coeficientul hidrotermic anual cu proprietățile „name”, „htc” și „drought”

```
MERGE (htc5c:YHTCc{name:'HTC 2005 Center', htc:1.6, drought:0});  
MERGE (htc11c:YHTCc{name:'HTC 2011 Center', htc:0.7, drought:1});  
MERGE (htc16c:YHTCc{name:'HTC 2016 Center', htc:0.7, drought:1});  
MERGE (htc20c:YHTCc{name:'HTC 2020 Center', htc:1.2});  
MERGE (htc25c:YHTCc{name:'HTC 2021 Center'});
```

//Wind Speed by years

```
MERGE (wind09c:WSC{name:'Wind 2009 Center', mean_wind:3.2, median_wind:3.1});  
MERGE (wind14c:WSC{name:'Wind 2014 Center', mean_wind:2.6, median_wind:2.6});  
MERGE (wind20c:WSC{name:'Wind 2020 Center', mean_wind:2.9, median_wind:2.8});  
MERGE (wind22c:WSC{name:'Wind 2022 Center', mean_wind:2.9, median_wind:2.8});  
MERGE (wind23c:WSC{name:'Wind 2023 Center'});
```

Calitatea aerului

Se definesc variabile pentru calitatea aerului în diferite luni și ani, inclusiv sursa de poluare: surse staționare (agenți economici) și transportul auto. cu proprietățile „name”, „statSources” (surse staționare ale agenților economici) și „trans” (transportul auto).

Exemple: în 2005 s-a înregistrat o calitate a aerului de 20,3 și o transmitere a datelor de 161,5, în 2014 s-a înregistrat o calitate a aerului de 15,0 și o transmitere a datelor de 179,0, etc.

```
//Air Quality
MERGE (aq05 :AQ{name: 'Air Quality Center', statSources: 20.3 ,
trans: 161.5});
MERGE (aq14 :AQ{name: 'Air Quality Center', statSources: 15.0 ,
trans: 179.0});
MERGE (aq18 :AQ{name: 'Air Quality Center', statSources: 15.2 ,
trans: 198.1});
MERGE (aq21 :AQ{name: 'Air Quality Center', statSources: 17.5 ,
trans: 172.1});
MERGE (aq22 :AQ{name: 'Air Quality Center'});
```

Productivitatea culturii de soia (q/ha)

Se definesc variabile (*soy09*) pentru determinarea productivității soiei și variabila *harvest* pentru înregistrarea cantității medii de soia exprimată în chintale (q) recoltată pe hectar în diferiți ani. În dependență de cantitatea de roadă anul se clasifică 3 clase: **1** – productivitate joasă, **2** – productivitate medie și **3** – productivitate înaltă. Exemple: în 2009 s-au recoltat 11 q de soia pe hectar, în 2015 s-au recoltat 7 tone de soia pe hectar, etc.

```
//Soybean harvest(q/ha);
MERGE (soy09:Prod{name: 'Soy Prod 2009 Center', harvest: 11, class:
1});
MERGE (soy15:Prod{name: 'Soy Prod 2015 Center', harvest: 7, class: 1});
MERGE (soy21:ProdP{name: 'Soy Prod 2021 Center', harvest: 23.7, class:
3});
MERGE (soy22:ProdP{name: 'Soy Prod 2022 Center', harvest: 13.9, class:
1});
MERGE (soy23:ProdP{name: 'Soy Prod 2023 Center'});
```

Interconectarea nodurilor

În continuare, se efectuează operații de conectare între noduri utilizând relații specifice. De exemplu, se realizează conexiuni între nodurile care reprezintă mediile temperaturilor lunare și cele care reprezintă mediile temperaturilor anuale. De asemenea, se stabilesc relații între nodurile care reprezintă cantitățile lunare de precipitații și cele care reprezintă producția de soia. Aceste operații de conectare facilitează interogările ulterioare ale bazei de date, permițând accesul la informații interconectate și realizarea de analize și calcule pe baza acestora. Este important de menționat că codul în sine nu este complet și există anumite instrucțiuni care au valori lipsă, cum ar fi numărul de identificare al nodului sau relația specificată. Codul prezentat este o secvență de instrucțiuni de creare a și de asociere a diferitelor noduri într-o bază de date (anexa 29). Un fragment din graful realizat este redat în figura 4.1. Au fost realizate relații între diferite tipuri de noduri. Unele instrucțiuni de realizare a relațiilor sunt redate mai jos:

- *Cantitatea lunară de precipitații → cantitatea anuală de precipitații;*
- *Temperatura medie lunară a aerului → temperaturi mediane anuale;*
- *Coeficientul hidrotermic lunar → coeficientul hidrotermic anual;*
- *Coeficientul hidrotermic anual → productivitatea;*
- *Viteza vântului → productivitatea;*
- *Calitatea aerului → productivitatea;*
- *Calitatea aerului → cantitatea de precipitații;*
- *Calitatea aerului → temperatura aerului, ș.a.*

Notă: Comentariile sunt redată în limbajul Cypher prin două bare oblice „//”.

// :YRC→ :YHTCc (Cantitatea anuală de precipitații → Coeficientul hidrotermic anual):

```
MATCH (r), (c) WHERE Id(r)= 652 and id(c)= 611 MERGE (r) -
[rel: DETERMINE_HTC]->(c);
// :YTC → :YCHTC (Mediana temperaturii anuale → Coeficientul
hidrotermic anual):
MATCH (t), (c) WHERE Id( t )= 672 and id(c)= 611 MERGE (t) -
[rel: DETERMINE_HTC]->(c);
// :YHTCc → :SoyProd (Coeficientul hidrotermic anual →
Productivitatea)
MATCH (c), (p) WHERE Id(c)= 611 and id(p)= 631 MERGE (c) -
[rel: DETERMINE_PROD] ->(p);
```

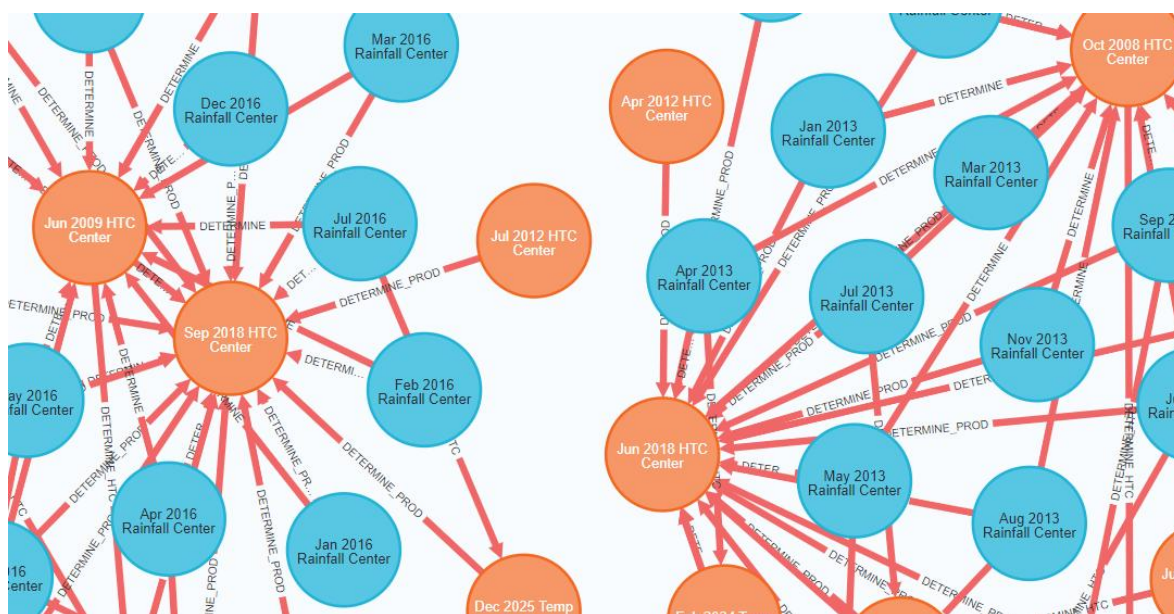


Figura 4.1. Graful rezultat (fragment)

Strucura depozitului de date graf este prezentată în anexa 18.

4.2. Predicția fenomenelor în baza datelor stocate în depozitul de date

Referitor la predicția fenomenelor, tehnologia progresează în acest domeniu, permițând utilizarea modelelor climatice și a datelor meteorologice pentru a prezice seceta și pentru a ajuta

la luarea de măsuri preventive. De exemplu, sistemele de alertă a secetei bazate pe date climatice pot ajuta la prevenirea pierderilor economice și la gestionarea mai eficientă a resurselor de apă. Elaborarea unui model de predicție a secetei implică multiple etape și pași, iar literatura de specialitate conține numeroase cercetări care descriu aceste aspecte și abordări pe această temă. Cercetările prezintă o analiză detaliată a modelelor și tehnicilor de predicție a secetei utilizând metode de învățare automată (*Machine Learning*), descriind etapele de selecție a datelor, de pre-procesare, de construire a modelelor și de evaluare a performanțelor acestora [36-38].

Pentru predicția secetei și a productivității în Republica Moldova este necesar un sistem de suport al deciziilor care să țină cont de specificul local și regional. Acest lucru poate ajuta fermierii, factorii de decizie și alte părți interesate să ia decizii în cunoștință de cauză cu privire la producția agricolă în condițiile instabilității climatice.

Pentru elaborarea acestui sistem au fost urmați anumiți pași:

1. *Definirea problemei*: determinarea scopului sistemului.
2. *Definirea factorilor specifici* care vor fi luați în considerare, cum ar fi temperatura, precipitațiile, umiditatea solului, tipurile de culturi, productivitatea, ș.a.
3. *Selectarea datelor*: selectarea datelor istorice despre modelele meteorologice, recoltele recoltelor și alți factori relevanți.
4. *Analiza datelor*: poate fi utilizată analiza statistică și algoritmi de învățare automată pentru a identifica modele în date. De exemplu, utilizarea analiza de regresie pentru a identifica relația dintre temperatură și randamentul culturilor. De asemenea, pot fi utilizați algoritmi de grupare pentru a grupa zone similare pe baza modelelor meteorologice și a tipurilor de culturi. Utilizarea funcțiilor *Cypher* și algoritmi din biblioteca GDS (*Graph Data Science*) pentru a analiza datele și a identifica tendințe, modele sau relații relevante.
5. *Dezvoltarea de modele*: utilizarea informațiilor obținute din analiza datelor pentru a dezvolta modele predictive. Aceste modele pot include indici de secetă, modele de creștere a culturilor și modele de predicție a randamentului. Acești algoritmi de învățare automată sunt folosiți pentru a antrena aceste modele și a le rafina în timp pe măsură ce depozitul se completează cu mai multe date.
6. *Evaluarea și perfecționarea modelelor*: sistemul este testat utilizând date istorice pentru a vedea cât de bine prezice seceta și recoltele. Pentru a îmbunătăți acuratețea predicțiilor modelele și algoritmi se rafinează sau se refactorizează după cum este necesar.
7. *Implementarea modelului predictiv*: Selectarea algoritmului potrivit pentru problema analizată și utilizarea lui pentru a crea modelul predictiv. Modelul realizat se încorporează în depozitul de date *Neo4j*.

8. Validarea modelului: implică testarea acestuia folosind date independente. Utilizarea datelor de testare pentru a evalua performanța modelului și pentru a determina dacă îndeplinește cerințele problemei.
9. *Utilizarea modelului*: După ce modelul a fost validat, acesta poate fi utilizat pentru a face predicții în timp real sau pentru a explora datele și relațiile din baza de date referitoare la zona sau regiunea țintă. Urmează instruirea utilizatorilor despre modul cum să folosească sistemul în mod eficient și să îl încorporeze în procesele lor de luare a deciziilor.

În cazul predicției secetei, un sistem graf poate fi folosit pentru a vizualiza istoricul pluviometric și a nivelurilor de apă din sol, astfel încât să se poată observa tendințe și modele de uscare a solului și de creștere a nivelului de secetă. Acest lucru poate ajuta la dezvoltarea unor modele de predicție mai precise și la identificarea zonelor care sunt cel mai susceptibile la secetă. *Scopul* elaborării unui model de predicție a secetei este de a contribui la dezvoltarea unui sistem avansat și eficient care să asiste decidentul cu informații și recomandări referitoare la diminuarea impactului secetei asupra comunităților umane, a agriculturii și a mediului înconjurător.

Modelul propus are următoarele obiective:

- a) *Identificarea zonelor vulnerabile la secetă*: Prin utilizarea unor algoritmi de analiză și modelare a datelor, modelul trebuie să fie capabil să identifice zonele geografice care sunt susceptibile la secetă. Parametrii precum nivelul de precipitații, temperatura și evapotranspirația pot fi utilizați pentru a evalua gradul de vulnerabilitate.
- b) *Avertizarea timpurie pentru agricultori și autorități*: Modelul de predicție trebuie să ofere informații timpurii și precise cu privire la probabilitatea apariției secetei într-o anumită regiune. Aceste informații pot fi utilizate de agricultori și autorități pentru a lua măsuri preventive și pentru a minimiza pierderile.
- c) *Suport în planificarea utilizării resurselor*: Modelul trebuie să permită o planificare optimă a utilizării resurselor de apă și sol în zonele predispuse la secetă. Acest lucru poate ajuta la reducerea impactului negativ al secetei prin gestionarea eficientă a acestor resurse.
- d) *Îmbunătățirea pregătirii și răspunsului la dezastre*: Modelul de predicție trebuie să faciliteze pregătirea și răspunsul rapid în situații de urgență, cum ar fi secetele severe. Informațiile furnizate de model pot ajuta la luarea deciziilor corecte și la implementarea unor măsuri adecvate pentru minimizarea impactului negativ asupra comunităților și mediului înconjurător.

Modelul de predicție a secetei beneficiază de tehnologii avansate, precum tehnologia graf, care poate contribui la stocarea și modelarea eficientă a datelor relevante într-o formă grafică. De

asemenea, analiza relațiilor complexe între diferiți factori și planificarea optimă a utilizării resurselor pot fi realizate prin intermediul algoritmilor de învățare automată și al tehnicilor de analiză a datelor.

Prin dezvoltarea unui model de predicție a secetei bazat pe aceste principii, se poate obține o înțelegere mai profundă a factorilor care influențează apariția și intensitatea secetei, precum și a măsurilor preventive și de adaptare care pot fi luate. Acesta poate contribui la o mai bună gestionare a resurselor și a riscurilor asociate cu seceta, în vederea asigurării unei dezvoltări durabile și a unei adaptări eficiente la schimbările climatice. Instrucțiunile acestui model sunt prezentate în anexa 20.

4.3. Dezvoltarea modelului de predicție

Predicția legăturilor este o sarcină în teoria grafurilor și învățarea automată în care scopul este de a prezice legăturile sau relațiile lipsă între nodurile dintr-un graf. În contextul predicției secetei, *Neo4j* poate fi folosit pentru a reprezenta relațiile dintre diverși factori care contribuie la secetă, cum ar fi temperatura, precipitațiile, umiditatea solului, coeficientul hidrotermic, etc.

Există mai mulți algoritmi de învățare automată care pot fi utilizați pentru a crea un model de predicție în *Neo4j*. Acești algoritmi pot fi implementați în *Neo4j* folosind bibliotecile *APOC* (*Awesome Procedures On Cypher*) și *GDS* (*Graph Data Science*), care oferă o varietate de algoritmi de învățare automată și grafuri. Utilizarea algoritmilor de analiză de date, cum ar fi algoritmi de regresie sau de învățare automată, pentru a analiza relațiile dintre date și a determina care dintre acestea sunt cele mai importante pentru prezicerea secetei. Pentru a realiza un model de predicție este important să se aleagă algoritmul potrivit și să se testeze diferite metode pentru a se vedea care este mai precisă și mai eficientă.

O funcționalitate cheie a *GDS* este capabilitatea sa de a lucra cu grafuri mari și complexe, permițând analiza și vizualizarea lor cu ușurință.

Modelarea unui graf de învățare automată de predicție se desfășoară în mai multe etape.

Analiza complexă a grafurilor prin metodele de *învățare automată* se poate realiza prin crearea de **conductele de procesare a grafurilor** (în engleză, *Graph Processing Pipelines*) care reprezintă un ansamblu de procese și tehnologii utilizate pentru a efectua operațiuni de analiză și procesare pe grafuri, adică pe modele de date care constau din noduri și relații interconectate.

În biblioteca *GDS* din SGBD *Neo4j*, conductele oferă un flux de lucru *end-to-end*, de la extragerea de caracteristici la instruire și aplicarea modelelor de învățare automată.

Aceste conducte sunt utilizate în mod obișnuit în domeniul analizei datelor și al învățării automate, pentru a efectua o serie de operații, cum ar fi:

- Extrage date de la noduri și relații
- Calculează metrici și proprietăți ale grafurilor, cum ar fi centralitatea grafurilor și gradul nodurilor
- Efectuează transformări ale grafurilor, cum ar fi agregarea, filtrarea și împărțirea grafurilor
- Realizează analize și modelare pe grafuri, cum ar fi clustering-ul și detectarea comunităților
- Instruiește și testează modele de învățare automată pentru a efectua clasificări și predicții pe grafuri.

Crearea unei conduite de predicție implică următorii pași:

1. *Selectarea datelor* de învățare și datele de testare necesare pentru modelul de predicție. Aceste date pot fi colectate dintr-o varietate de surse, cum ar fi bazele de date, fișiere CSV, API-uri, sau alte surse.
2. *Preprocesarea datelor* implică pregătirea și prelucrarea datelor brute pentru a le face mai ușor de analizat și de utilizat în algoritmul de învățare automată. În majoritatea scenariilor de învățare automată, sunt aplicați mai mulți pași de preprocesare pentru a produce date care sunt adaptabile algoritmilor de învățare automată. Preprocesarea grafului se referă la procesul de pregătire și prelucrarea datelor brute pentru a le face mai ușor de analizat și de utilizat în algoritmul de învățare automată cum ar fi curățarea și transformarea unui set de date graf înainte de a fi utilizare. Scopul preprocesării este de a oferi caracteristici bune pentru algoritmul de învățare. Acesta este un pas crucial în dezvoltarea unui model de predicție precis și robust. Preprocesarea grafurilor poate include diverse etape și pași, în funcție de scopul și natura datelor și a analizei:
 - a) *Selectarea nodurilor relevante*: eliminarea nodurilor și a relațiilor inutile. În anumite cazuri, graful poate conține noduri sau relații care nu sunt relevante pentru analiză și care pot afecta performanța modelului. Eliminarea acestora poate reduce dimensiunea grafurilor și poate îmbunătăți performanța modelului.
 - b) *Curățarea datelor*: Datele trebuie curățate pentru a elimina valorile lipsă, valorile nevalide sau eronate, duplicatelor și zgomotul din date. De asemenea, datele pot fi normalizate pentru a fi puse într-un format coerent și comparabil.
 - c) *Normalizarea și scalarea datelor*: În cazul datelor care au valori diferite în mod semnificativ, cum ar fi datele economice sau de mediu, este necesară normalizarea sau scalarea datelor pentru a se asigura că acestea sunt comparabile. Acest lucru poate ajuta la evitarea influenței asupra analizei a valorilor extreme sau a datelor greșite.
 - d) *Împărțirea grafurilor în componente conectate*: În cazul unui graf neconectat, acesta poate fi împărțit în componente conectate, pentru a se asigura că analiza se

concentrează pe zonele relevante și că se evită erorile de analiză cauzate de zonele izolate.

e) *Eliminarea nodurilor și relațiilor redundante*: În cazul unui graf mare și complex, nodurile sau relațiile care sunt redundante sau care nu au nicio influență semnificativă asupra analizei pot fi eliminate sau reținute, pentru a simplifica grafurile și a crește performanța modelului.

Pașii specifici de preprocesare vor depinde de tipul de graf, de cerințele specifice și de datele disponibile.

3. *Selectarea caracteristicilor relevante*: Uneori, datele pot conține caracteristici irelevante sau redundante care pot afecta performanța modelului de predicție. În această etapă, se selectează doar caracteristicile relevante pentru problema de predicție, pentru a reduce complexitatea modelului și pentru a îmbunătăți performanța acestuia.
4. *Transformarea datelor*: Datele pot fi transformate pentru a le face mai ușor de analizat. Aceasta poate include transformarea datelor categorice în date numerice, scalarea datelor, sau convertirea datelor într-un format standardizat, cum ar fi formatul CSV.
5. *Splitarea datelor*: Datele pot fi împărțite într-un set de date de Instruire și un set de date de testare. Setul de date de Instruire este utilizat pentru a antrena modelul, în timp ce setul de date de testare este utilizat pentru a evalua performanța modelului.
6. *Balansarea datelor*: În cazul în care datele sunt nebalansate, adică unele clase sunt sub-reprezentate în date, atunci este important să se balanseze datele. Acest lucru poate fi realizat prin utilizarea tehnicilor de *oversampling* sau *undersampling*.
7. *Reducerea dimensiunii datelor*: Datele pot fi comprimate pentru a reduce dimensiunea setului de date. Aceasta poate include tehnici de extragere a caracteristicilor, cum ar fi analiza componentelor principale, sau eliminarea caracteristicilor redundante prin selecția caracteristicilor.
8. *Instruirea modelului*: Utilizați datele de învățare pentru a antrena modelul de predicție. Acest lucru poate fi realizat prin utilizarea diferitelor tehnici de învățare automată, cum ar fi regresia liniară, rețele neuronale sau algoritmi de clasificare.
9. *Validarea modelului*: După ce modelul a fost instruit, acesta trebuie validat pentru a asigura că funcționează corect. Acest lucru se face prin utilizarea setului de date de testare pentru a evalua performanța modelului. De exemplu, puteți utiliza metrici precum acuratețea, sensibilitatea și specificitatea pentru a evalua performanța modelului.
10. *Utilizarea modelului*: După ce modelul a fost validat, acesta poate fi utilizat pentru a face predicții pentru datele noi sau pentru a explora datele din baza de date. Acest lucru poate fi

realizat prin utilizarea interfețelor de programare a aplicațiilor (API-uri) sau prin integrarea modelului în alte aplicații.

Proiectarea unui graf în memorie

Proiecția unui graf specifică nodurile din baza de date care trebuie proiectate în graful GDS din memorie. Proiecția se bazează pe etichetele nodurilor care pot fi utilizate în funcție de cât de detaliată trebuie să fie proiectarea. Următoarea instrucțiune va proiecta un graf folosind o proiecție nativă și o va stoca în catalogul de grafuri sub numele „*predGraph*”, utilizând etichetele (tipurile de noduri) :YHTCc (pentru instruire) și :YHTCcp (pentru predicție) cu proprietățile *htc* și *drought*. Instrucțiunea și rezultatul rulării este redat în figura 4.2.

<pre> 1 //Proiecția grafului 2 CALL gds.graph.project('dtPred',{ 3 YHTCc:{properties:['htc','drought']}, 4 YHTCcp:{properties:'htc'} 5 }, 6 '*') 7); </pre>					
nodeProjection	relationshipProject	graphName	nodeCount	relationshipCount	projectMillis
<pre> {YHTCcp: {label: "YHTCcp", properties: {htc: {defaultVa HTCc: {label: "HTCc", properties: {htc: {defaultValu exInverse: false, a e: null, property: gregation: "DEFAUL "htc"}}}, YHTCc: {l T", type: "*", prop abel: "YHTCc", prop erties: {}}} erties: {drought: { defaultValue: null, property: "drought "}, htc: {defaultVa lue: null, property : "htc"}}}} </pre>					
		dtPred	24	21	3678

Figura 4.2. Proiecția grafului

Funcția ***gds.graph.project*** returnează noul graf proiectat, care poate fi utilizat ulterior în algoritmi și interogări de graf. Această funcție primește următoarele argumente:

- '***dtPred***': numele grafului sursă de pe care se va face proiecția;
- '***YHTCc***': numele nodului utilizat în proiecția grafului;
- '***{ properties: 'htc' }***': specifică faptul că proiecția trebuie să includă proprietatea '***htc***' a fiecărui nod;
- '*******': specifică faptul că graful proiectat va conține toate relațiile din graful sursă.

Un pas important al preprocesării grafului este necesară determinarea nodurilor similare. În acest sens se utilizează algoritmul *kNN* (*k-Nearest Neighbors*) care este o tehnică comună, puternică pentru identificarea modelelor și relațiilor complexe în seturi de date și este utilizată pe

scară largă în învățarea automată, analiza datelor și cercetarea științifică, oferind soluții eficiente pentru diverse aplicații. Algoritmul *kNN* identifică cei mai apropiați *k* vecini ai unui nod dat, pe baza unei metrici de similaritate, cum ar fi distanța sau caracteristicile numerice ale nodurilor. Acești vecini apropiați pot fi apoi utilizați pentru a grupa nodurile similare, clasificarea sau gruparea datelor sau detectarea anomaliilor sau valorilor aberante într-un set de date. Prin identificarea nodurilor similare în setul de date, algoritmul *kNN* ajută la reducerea dimensionalității și facilitează analiza datelor în sarcinile ulterioare. Scopul acestei etape este de a pregăti datele pentru analize ulterioare utilizând alți algoritmi de învățare automată sau metode statistice. Rezultatul rulării acestui algoritm este redat în figura 4.3.

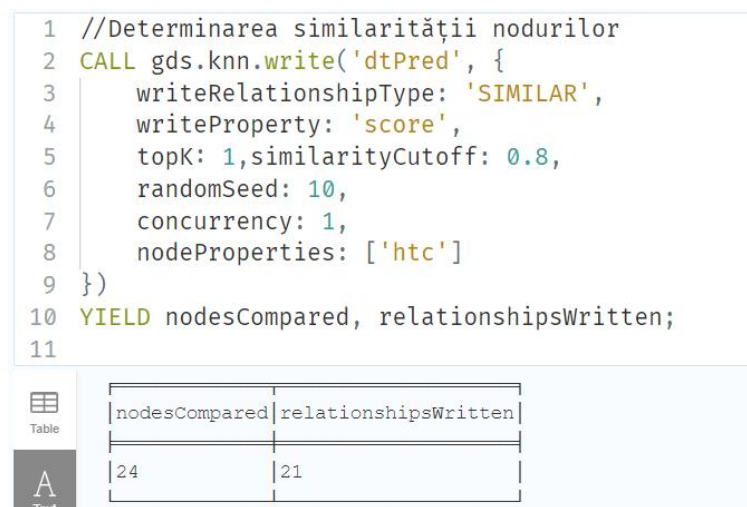


Figura 4.3. Determinarea nodurilor similare

Funcția *gds.knn.write* primește următoarele argumente:

- *'dtPred'*: numele grafului în care se dorește crearea relațiilor de similaritate;
- *writeRelationshipType*: tipul de relație care va fi creat între nodurile similare (în acest caz, *'SIMILAR'*);
- *writeProperty*: proprietatea care va fi adăugată relației (în acest caz, *'score'*);
- *topK*: numărul maxim de noduri similare care vor fi conectate prin relația *'SIMILAR'* pentru fiecare nod (în acest caz, 1);
- *similarityCutoff*: pragul de similaritate minim care trebuie atins pentru ca două noduri să fie considerate similare și, prin urmare, să fie conectate prin relația *'SIMILAR'* (în acest caz, 0.5);
- *randomSeed*: valoarea folosită pentru a inițializa generatorul de numere aleatorii, pentru a asigura reproducibilitatea experimentului;
- *concurrency*: numărul maxim de fire de execuție care pot fi utilizate simultan pentru a procesa datele;

- *nodeProperties*: lista de proprietăți ale nodurilor care trebuie luate în considerare în calculul similarității;
- Funcția *gds.knn.write* returnează numărul de noduri comparate și numărul de relații create în graf, care sunt salvate în variabilele *nodesCompared* și *relationshipsWritten*, respectiv.

Scorul de similaritate reprezintă o valoare numerică ce măsoară gradul de similitudine între două noduri. În primul rând, scorul de similaritate este folosit pentru a compara sau grupa anii în funcție de cantitatea de precipitații, coeficientul hidrotermic și productivitatea culturilor. De asemenea, este utilizat pentru gruparea culturilor în funcție de productivitate și a zonelor geografice în funcție de speciile și soiurile cultivate. Valoarea similarității se poate baza pe orice număr de factori, în funcție de natura datelor analizate. Scorul rezultat poate fi folosit pentru a identifica noduri care au o legătură semantică sau pentru a găsi date relevante într-o interogare de căutare. În contextul învățării automate, scorul de similaritate poate fi utilizat pentru a compara similitudinea dintre două puncte de date sau vectori. De exemplu, acesta poate fi utilizat pentru a compara asemănarea a două imagini ale unui soi de plante, cu scopul de a determina starea de sănătate a plantei respective.

În ceea ce privește regăsirea informațiilor, scorul de similaritate este utilizat pentru a clasifica rezultatele căutării în funcție de relevanța lor față de o anumită interogare. Aceasta se realizează prin compararea interogării cu fiecare nod dintr-un graf și calcularea unui scor de similaritate pentru fiecare nod. Ulterior, nodurile cu cele mai mari scoruri de similaritate sunt returnate ca rezultate de căutare de top. Găsind anii cu același *cht* putem determina legități de repetare a anilor secetoși, probabilitatea determinării cantității de recoltă pentru anumite specii sau soiuri de culturi. În figura 4.4 este redată similaritatea anilor după coeficientul hidrotermic (*cht*).

```

1 MATCH (n:YHTCc)-[r:SIMILAR]→(m:YHTCc)
2 RETURN n.name as year1,n.htc, m.name as year2,m.htc, r.score as similarity
3 ORDER BY similarity DESCENDING, year1, year2

```

year1	n.htc	year2	m.htc	similarity
"HTC 2006 Center"	1.1	"HTC 2017 Center"	1.1	1.0
"HTC 2007 Center"	0.5	"HTC 2015 Center"	0.5	1.0
"HTC 2009 Center"	0.6	"HTC 2019 Center"	0.6	1.0
"HTC 2010 Center"	1.2	"HTC 2004 Center"	1.2	1.0
"HTC 2011 Center"	0.7	"HTC 2018 Center"	0.7	1.0

Figura 4.4. Anii cu *htc* similar

În algoritmiile de învățare automată, *semințele* (sau *seed*-urile) sunt valorile inițiale folosite pentru a inițializa generatorii pseudo-aliatori. Acești generatori produc o secvență de numere care par să fie aliatoare, dar sunt generate în mod determinist. Utilizarea aceleiași semințe și aceluiași

algoritm va produce întotdeauna aceeași secvență de numere. În algoritmii de învățare automată, aceștia sunt folosiți pentru a introduce aleatorizare în procesul de antrenare. De exemplu, pot fi utilizați pentru a amesteca datele de antrenare, a selecta sub-seturi aleatorii de date sau a inițializa valorile ponderilor în rețelele neuronale. Utilizarea aceleiași semințe în algoritmi de învățare automată asigură reproductibilitatea rezultatelor. Procesele care necesită generarea de numere aliate, cum ar fi amestecarea datelor de antrenare sau inițializarea valorilor ponderilor, vor produce întotdeauna aceleași rezultate cu aceleași semințe, ceea ce facilitează comparația și reproducerea rezultatelor. Pentru a verifica nodurile similare, se utilizează o instrucțiune *Cypher* care returnează anii, coeficientul *cht* și similaritatea lor cu anii corespunzători. Aceste rezultate pot fi vizualizate sub forma graf (fig. 4.5).

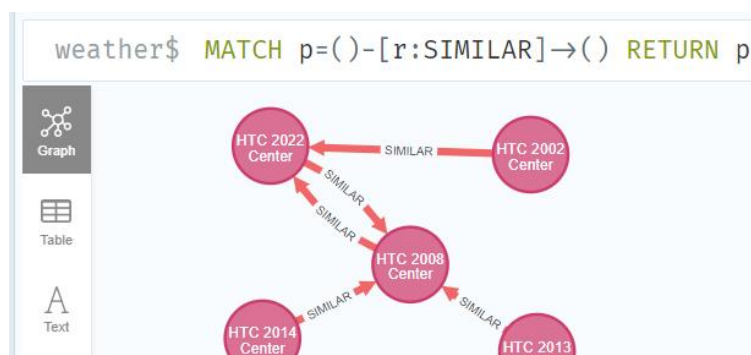


Figura 4.5. Nodurile similare

Pentru a putea analiza interacțiunile complexe dintre diferitele elemente și variabile care pot afecta apariția și evoluția fenomenelor studiate se utilizează ca pas de preprocesare **încorporarea nodurilor** (*embedding*) într-un model (figura 4.6). Încorporarea nodurilor reprezintă un proces de preprocesare în care obiectele sau entitățile interconectate într-un context sunt reprezentate printr-un set de reprezentări vectoriale cu dimensiuni reduse.

```

1 //Incorporarea nodurilor
2 CALL gds.fastRP.stream('dtPred',{
3   embeddingDimension: 20,
4   randomSeed: 10
5 })
6 )
7 YIELD nodeId, embedding;
8

```

nodeId	embedding
1067	[0.0, -0.3779645264148712, 0.0, -0.3779645264148712, 0.3779644668102264, 44, 0.0, -5.960464477539063e-8, 0.0, 0.0, 0.0, 0.3779645264148712, 0.3, 7796446681022644, -0.3779645264148712, -0.37796446681022644, 0.0, 0.0, 0.3779645264148712, -0.37796446681022644, -5.960464477539063e-8, -0.3, 7796446681022644]

Figura 4.6. Incorporarea nodurilor

Proprietățile de incorporare

Procedurile `gds.fastRP.stream()` și `gds.fastRP.write()` din biblioteca *GDS* este utilizată pentru a calcula înglobările *FastRP* pentru nodurile dintr-un graf (figura 4.7). La apelul acestei funcții au fost furnizați următorii parametri:

- *htcPred* - este numele grafului proiectat pe spațiul de încorporare.
- *embeddingDimension* - specifică faptul că dimensionalitatea înglobărilor *FastRP* ar trebui să fie de 20.
- *writeProperty* - indică faptul că înglobările rezultate ar trebui scrise într-o proprietate numită „*fastrp-embedding*” pe fiecare nod.
- *randomSeed* - setează *semințele* (*seed*) pentru generatorul de numere aleatorii utilizat în algoritmul *FastRP*.
- *relationshipWeightProperty* - specifică proprietatea relațiilor '*score*' care conține ponderile utilizate pentru calcularea înglobărilor.

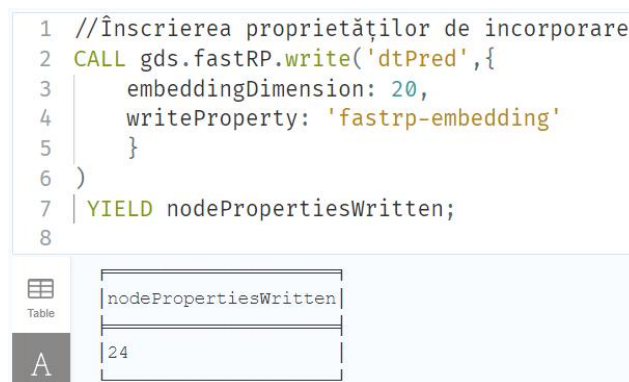


Figura 4.7. Proprietățile înscrise în noduri

Procedura va produce trei valori pentru fiecare nod:

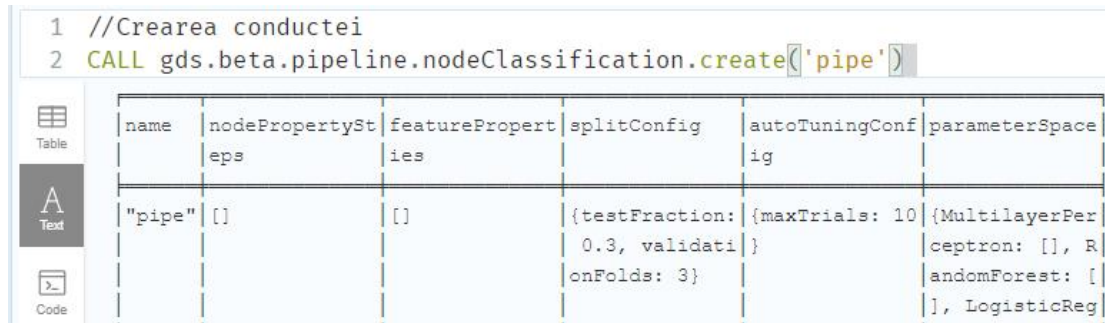
- ✓ *nodeId* - identificatorul nodului din grafic,
- ✓ *embedding* - încorporarea *FastRP* calculată pentru nod ,
- ✓ *nodePropertiesWritten* - indică numărul de proprietăți care au fost scrise în nod.

Aceste încorporări de noduri sunt utilizate ulterior în sarcini de învățare automată, cum ar fi clasificarea nodurilor, predicția legăturilor sau construirea grafului de similaritate *kNN*. În acest proces, se calculează reprezentări vectoriale pentru fiecare nod într-un graf, reflectând relațiile și interdependențele dintre acestea.

12. Crearea unei conduite pentru instruire.

13. Primul pas în construirea unei conduite se realizează folosind funcția "*gds.beta.pipeline.nodeClassification.create*" (fig. 4.8). Această funcție stochează un obiect de tip "*pipeline*" care poate fi ulterior antrenat și adăugat în catalogul de conduite de tipul

"Node classification training pipeline". Acest obiect reprezintă o conductă configurabilă, ce poate fi invocată în viitor pentru instruire și care, la rândul său, va crea un model de clasificare. Acest model de clasificare, la final, va fi, de asemenea, stocat în catalogul cu tipul "NodeClassification".



```

1 //Crearea conductei
2 CALL gds.beta.pipeline.nodeClassification.create('pipe')

```

name	nodePropertySt	featurePropert	splitConfig	autoTuningConf	parameterSpace
"pipe"	eps	ies	{testFraction: 0.3, validationFolds: 3}	{maxTrials: 10}	{MultilayerPerceptron: [], RandomForest: [], LogisticRegression: []}

Figura 4.8. Crearea conductei

Configurarea conductei

Într-o conductă de clasificare a nodurilor, pot fi executate unul sau mai mulți algoritmi din cadrul bibliotecii *GDS* în modul "mutate". Acești algoritmi pot crea proprietăți noi pentru nodurile grafului din memorie. Astfel de pași de creare a proprietăților pot fi conectați unul după altul, iar proprietățile rezultate pot fi ulterior utilizate ca caracteristici pentru procesul de clasificare. Un aspect important este că pașii de creare a proprietăților pentru noduri, care sunt adăugați în conducta de antrenament, vor fi executați atât în timpul antrenării modelului, cât și atunci când conducta de clasificare este aplicată pentru clasificare. Aceasta înseamnă că informațiile generate în etapele de creare a proprietăților vor fi utilizate pentru a învăța și a face predicții în cadrul procesului de clasificare. Aceste conduite de clasificare a nodurilor oferă un mod eficient de a crea și utiliza proprietăți ale nodurilor, care pot servi drept caracteristici în procesul de clasificare, permițând astfel obținerea de rezultate mai precise și relevante.

Următoarea instrucțiune este o comandă de apel către o funcție din biblioteca *GDS*. Aceasta utilizează funcția *gds.beta.pipeline.nodeClassification.addNodeProperty* pentru a adăuga o proprietate la nodurile dintr-un graf (fig. 4.9). Se adaugă un pas de proprietate de nod la conductă. Aici graful de intrare conține o proprietate *htc*:

1	//Configurarea conductei. Se adaugă un pas de proprietate de nod htc la conductă:
2	CALL gds.beta.pipeline.nodeClassification.addNodeProperty(
3	'pipe', 'alpha.scaleProperties', {
4	nodeProperties: ['htc'], scaler: 'stdscore',
5	mutateProperty: 'scaledSizes'
6	})
7	YIELD name, nodePropertySteps;

Table	name	nodePropertySteps
Text	"pipe"	[{"name": "gds.alpha.scaleProperties.mutate", "config": {"scaler": "stdscore", "contextRelatio", "nshipTypes": [], "contextNodeLabels": [], "mutateProperty": "scaledSizes", "nodeProperties": [

Figura 4.9. Adăugarea unui nod de proprietate

Funcția „gds.beta.pipeline.nodeClassification.addNodeProperty” primește următorii parametri:

- *pipe*: Acesta este un argument obligatoriu și reprezintă numele conductei (pipeline) pe care dorim să o folosim pentru clasificarea nodurilor.
- *'alpha.scaleProperties'*: Acesta este un argument obligatoriu și specifică numele unei etape de scalare din cadrul conductei, care va fi folosită pentru a scala proprietățile nodurilor.
- {nodeProperties: 'htc', scaler: 'stdscore', mutateProperty: 'scaledSizes'}: Aceasta este o configurație suplimentară necesară pentru etapa de scalare. Această configurație are următoarele componente:
 - ✓ *nodeProperties*: Specifică numele proprietății nodurilor pe care dorim să o scalăm. În acest caz, proprietatea este denumită *'htc'*.
 - ✓ *scaler*: Specifică tipul de scalare pe care dorim să îl aplicăm proprietății. În acest caz, se utilizează scalarea standardizată, cunoscută și ca *Z-score*.
 - ✓ *mutateProperty*: Specifică numele proprietății rezultate după scalare. În acest caz, noua proprietate va fi denumită *'scaledSizes'*.

În final, instrucțiunea *YIELD* este utilizată pentru a obține rezultatele sau informațiile despre etapele procesului, cum ar fi numele conductei (*name*) și detaliile etapelor aplicate asupra proprietăților nodurilor (*nodePropertySteps*). Aceste rezultate pot fi utilizate ulterior în interogări sau pentru a analiza procesul de scalare al proprietăților nodurilor.

14. Selectarea caracteristicilor pentru conductă.

Conducta de clasificare a nodurilor oferă posibilitatea de a selecta un subset din proprietățile disponibile ale nodurilor, pe care să le utilizați ca caracteristici pentru modelul de învățare automată. Atunci când executați această conductă, proprietățile de nod selectate trebuie să fie deja prezente în Graful de intrare, fie create anterior prin intermediul unor pași de creare a proprietăților pentru noduri (figura 4.10). Această abordare permite o flexibilitate în alegerea

caracteristicilor relevante pentru modelul de clasificare și asigură că acestea sunt disponibile și pot fi folosite în procesul de învățare automată pentru a obține rezultate precise și coerente.

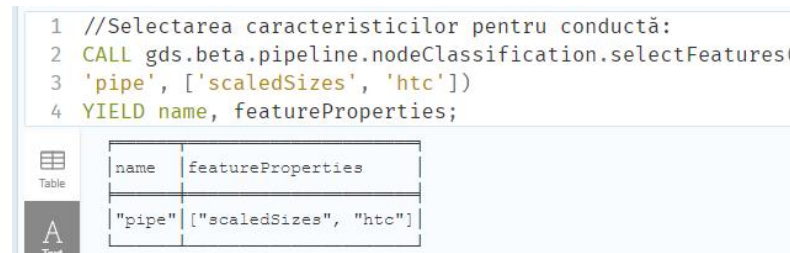


Figura 4.10. Instrucțiunea și rezultatul adăugării caracteristicilor

Această instrucțiune reprezintă un apel către funcția „*selectFeatures*” din biblioteca GDS: *gds.beta.pipeline.nodeClassification.selectFeatures*. Scopul acestei funcții este să selecteze anumite proprietăți ale nodurilor dintr-o conductă de clasificare și să le considere ca fiind caracteristici (features) pentru modelul de învățare automată. Explicarea parametrilor:

- *pipe*: Acesta este un argument obligatoriu și reprezintă numele conductei (pipeline) în cadrul căreia dorim să selectăm proprietățile nodurilor ca fiind caracteristici.
- *['scaledSizes', 'htc']*: Acesta este un argument obligatoriu și reprezintă o listă de nume ale proprietăților nodurilor pe care dorim să le considerăm drept caracteristici. În acest caz, proprietățile sunt "scaledSizes" și "htc".

În urma execuției instrucțiunii, se vor returna două rezultate:

- *name*: Aceasta este variabila în care va fi reținut numele conductei (pipeline) în care s-au selectat caracteristicile.
- *featureProperties*: Aceasta este variabila în care vor fi stocate proprietățile nodurilor selectate ca fiind caracteristici pentru modelul de învățare automată.

Aceste rezultate pot fi utilizate ulterior pentru a inspecta caracteristicile selectate și pentru a le folosi în continuare în procesul de clasificare a nodurilor.

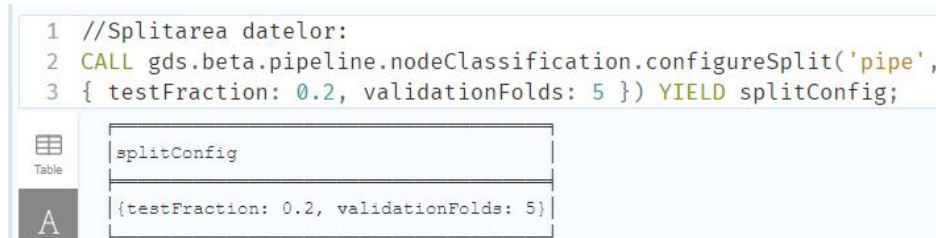
Divizarea nodurilor în graf.

Conductele de clasificare a nodurilor gestionează procesul de împărțire a nodurilor în mai multe seturi, utilizate pentru *instruire*, *testare* și *validare* a candidaților de modele definite în spațiul de parametri. Configurarea împărțirii este opțională, iar dacă nu este specificată, se vor folosi setările implicite. Pentru a inspecta configurația de divizare a unei conducte, se utilizează funcția *gds.beta.model.list* cu opțiunea *splitConfig*. Instrucțiunea și rezultatul splitării este reprezentat în figura 4.11.

Procesul de divizare a nodurilor este esențial în procesul de antrenament și urmează următorii pași:

- Graful de intrare este împărțit în două părți: *Graful de instruire* și *Graful de testare*. Astfel, se obține o vedere de ansamblu a acestui proces.
- Graful de instruire este ulterior împărțit într-un număr de pliuri de validare, fiecare constând dintr-o parte pentru antrenament și una pentru validare.
- Fiecare candidat model este instruit separat pe fiecare pli de instruire și evaluat pe pliul de validare corespunzător.
- Modelul cu cel mai mare scor mediu conform metricii primare va fi declarat câștigătorul procesului de antrenament.
- Modelul câștigător va fi apoi reantrenat pe întregul graf de instruire pentru a beneficia de toate datele disponibile.
- Modelul câștigător este evaluat atât pe *Graful de instruire*, cât și pe *Graful de test* pentru a verifica performanța sa.
- În cele din urmă, modelul câștigător este reantrenat pe întregul graf original, având la dispoziție toate datele pentru a fi pregătit pentru utilizare.

Această abordare de divizare a datelor și evaluare în etape multiple asigură o selecție optimă a modelului câștigător și un nivel ridicat de generalizare a acestuia pentru a fi folosit eficient în scenarii reale.



```

1 //Splitarea datelor:
2 CALL gds.beta.pipeline.nodeClassification.configureSplit('pipe',
3 { testFraction: 0.2, validationFolds: 5 }) YIELD splitConfig;

```

splitConfig
{testFraction: 0.2, validationFolds: 5}

Figura 4.11. Divizarea datelor

Adăugarea candidaților model.

O conductă conține o colecție de configurații pentru candidații de model, care este inițial goală. Această colecție se numește "spațiu de parametri". Spațiul de parametri al unei conducte poate fi inspectat folosind *gds.beta.model.list* și, opțional, producând numai *parameterSpace*. Fiecare configurație candidat de model conține fie valori fixe, fie intervale pentru parametrii de antrenament. Când este prezent un interval, valorile din interval sunt determinate automat de către un algoritm de reglare automată, denumit "Reglare automată". Candidații model reprezintă algoritmi de învățare automată utilizați în cadrul unei conducte de clasificare a nodurilor. Fiecare dintre aceștia reprezintă o configurație specifică pentru modelul de clasificare, iar alegerea unuia dintre aceștia depinde de particularitățile setului de date și de scopul problemei de clasificare.

Una sau mai multe configurații de model trebuie adăugate la spațiul de parametri al conduitei de antrenament, utilizând una dintre următoarele proceduri:

a) *gds.beta.pipeline.nodeClassification.addLogisticRegression*

Acest candidat model reprezintă o implementare a algoritmului de regresie logistică (fig. 4.12). Regresia logistică este un algoritm de învățare supervizată folosit pentru clasificarea binară (în care există doar două clase) sau *multiclass* (mai mult de două clase). Algoritmul încearcă să găsească o funcție logistică pentru a estima probabilitatea apartenenței unui exemplu la una dintre clasele de interes.

```

1 //Adăugarea unui model de regresie logistică cu un parametru de interval:
2 CALL gds.beta.pipeline.nodeClassification.addLogisticRegression(
3 'pipe', {maxEpochs: 500, penalty: {range: [1e-4, 1e2]}})
4 YIELD parameterSpace
5 RETURN parameterSpace.RandomForest AS randomForestSpace,
6 parameterSpace.LogisticRegression AS logisticRegressionSpace,
7 parameterSpace.MultilayerPerceptron AS MultilayerPerceptronSpace;

```

randomForestSpace	logisticRegressionSpace	MultilayerPerceptronSpace
{maxDepth: 2147483647, minLeafSize: 1, criterion: "GINI", minSplitSize: 2, numberOfDecisionTrees: 5, methodName: "RandomForest", numberOfSamplesRatio: 1.0}, {maxDepth: 2147483647, minLeafSize: 1, criterion: "GINI", minSplitSize: 2, numberOfDecisionTrees: 5, methodName: "RandomForest", numberOfSamplesRatio: 1.0}	{maxEpochs: 500, minEpochs: 1, classWeights: [], penalty: {range: [0.0001, 100.0]}, patience: 1, methodName: "LogisticRegression", focusWeight: 0.0, batchSize: 100, tolerance: 0.01, learningRate: 0.001}, {maxEpochs: 500, minEpochs: 1, classWeights: [], penalty: {range: [0.0001, 100.0]}, patience: 1, methodName: "LogisticRegression", focusWeight: 0.0, batchSize: 100, tolerance: 0.01, learningRate: 0.001}	{maxEpochs: 100, minEpochs: 1, classWeights: [0.4, 0.3, 0.3], penalty: 0.0, patience: 1, methodName: "MultilayerPerceptron", focusWeight: 0.5, hiddenLayerSizes: [100], batchSize: 100, tolerance: 0.001, learningRate: 0.001}, {maxEpochs: 100, minEpochs: 1, classWeights: [0.4, 0.3, 0.3], penalty: 0.0, patience: 1, methodName: "MultilayerPerceptron", focusWeight: 0.5, hiddenLayerSizes: [100], batchSize: 100, tolerance: 0.001, learningRate: 0.001}

Figura 4.12. Adăugarea unui model de regresie logistică

b) *gds.beta.pipeline.nodeClassification.addRandomForest*

Acest candidat model reprezintă o implementare a algoritmului de păduri aliatore (Random Forest). Pădurile aliatore sunt un ansamblu de arbori de decizie, unde fiecare arbore este antrenat pe o submulțime aliatore din date și apoi combină predicțiile individuale pentru a obține o predicție finală. *Random Forest* este eficient în gestionarea problemelor de clasificare și regresie și este robust la *overfitting*. Adăugarea unui astfel de model, precum și rezultatul acesteia, este descrisă în figura 4.13:


```

1 // Adăugarea modelelor candidat
2 //Adăugarea unui model de pădure aleatoriu:
3 CALL gds.alpha.pipeline.nodeClassification.addRandomForest(
4 'pipe', {numberOfDecisionTrees: 5}) YIELD parameterSpace;
5

```

parameterSpace
{MultilayerPerceptron: [{maxEpochs: 100, minEpochs: 1, classWeights: [0.4, 0.3, 0.3], penalty: 0.0, patience: 1, methodName: "MultilayerPerceptron", focusWeight: 0.5, hiddenLayerSizes: [100], batchSize: 100, tolerance: 0.001, learningRate: 0.001}], RandomForest: [{maxDepth: 2147483647, minLeafSize: 1, criterion: "GINI", minSplitSize: 2, numberOfDecisionTrees: 5, methodName: "RandomForest", numberOfSamplesRatio: 1.0}, {maxDepth: 2147483647, minLeafSize: 1, criterion: "GINI", minSplitSize: 2, numberOfDecisionTrees: 5, methodName: "RandomForest", numberOfSamplesRatio: 1.0}, {maxDepth: 2147483647, minLeafSize: 1, criterion: "GINI", minSplitSize: 2, numberOfDecisionTrees: 5, methodName: "RandomForest", numberOfSamplesRatio: 1.0}], LogisticRegression: [{maxEpochs: 500, minEpochs: 1, classWeights: [], penalty: {range: [0.0001, 100.0]},

Figura 4.13. Adăugarea unui model de pădure aleatoriu.

c) *gds.alpha.pipeline.nodeClassification.addMLP*

MLP (Multi-Layer Perceptrons) reprezintă o implementare a rețelelor neuronale cu perceptroni multipli. *MLP* este un tip de rețea neuronală artificială formată din mai multe straturi de neuroni, inclusiv un strat de intrare, cel puțin un strat ascuns și un strat de ieșire. *MLP* este capabil să rezolve probleme complexe de clasificare și regresie și poate captura relații non-liniare între caracteristici. Perceptronul este cea mai simplă rețea neuronală artificială formată dintr-un singur strat de neuroni artificiali [221]. Un perceptron are un singur strat de neuroni (sau noduri), care sunt conectați printr-o serie de ponderi la intrările rețelei. Fiecare intrare este asociată cu o anumită greutate (weight) care indică importanța relativă a acelei intrări pentru predicție.

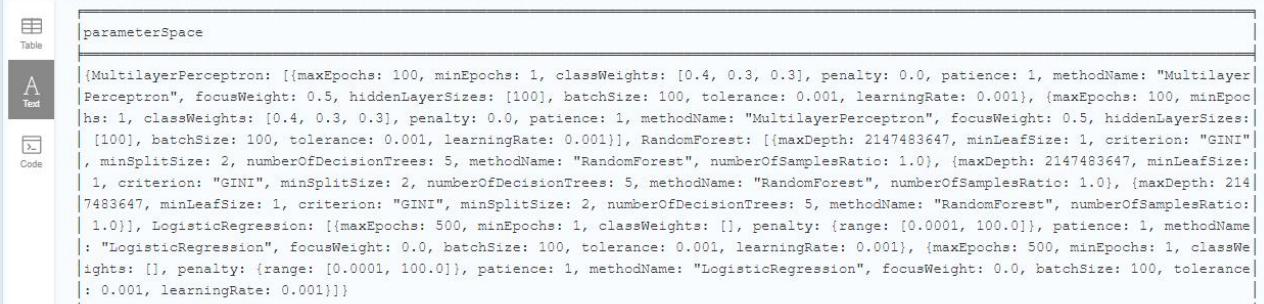
În figura 4.14 este reprezentată instrucțiunea pentru adăugarea unei configurații pentru un model de clasificare bazat pe rețele neuronale cu perceptroni multipli (MLP) în cadrul unei conduite. Această instrucțiune apelează funcția *gds.alpha.pipeline.nodeClassification.addMLP*. În urma acestei operații, se va obține spațiul de parametri al configurației modelului, care va fi stocat în variabila *parameterSpace*. Funcția dată primește următorii parametri:

- 'pipe': Acesta este un argument obligatoriu și reprezintă numele conduitei (pipeline) în cadrul căreia se va adăuga configurația modelului MLP.
- {classWeights: [0.4,0.3,0.3], focusWeight: 0.5}: Aceasta este o configurație suplimentară specificată ca un obiect JSON și are următoarele componente:

- ✓ *classWeights*: Acesta reprezintă ponderile claselor și este o listă de numere. În exemplul dat, avem trei clase și valoarea 0.4 este alocată primei clase, 0.3 celei de-a doua, și 0.3 celei de-a treia clase. Aceste ponderi pot fi utilizate pentru a trata neechitabilitatea claselor în procesul de învățare automată.
- ✓ *focusWeight*: Acesta reprezintă o valoare reală și indică ponderea focalizată asupra unei anumite clase. Această pondere adaugă o atenție mai mare asupra unei clase de interes în procesul de învățare.

În urma apelului funcției, rezultatul *parameterSpace* va conține configurația spațiului de parametri al modelului MLP, care poate fi ulterior utilizat în cadrul conduitei pentru instruirea și testarea modelului de clasificare. Configurarea modelului include atât parametrii prestabiliți, cât și cei personalizați în funcție de setul de date și cerințele problemei de clasificare. În figura 4.14 este dată instrucțiunea pentru adăugarea unui model de perceptron multistrat cu pierdere focală ponderată în clasă și rezultatul acesteia:

```
1 //Adăugarea unui model de perceptron multistrat cu pierdere focală ponderată în clasă:
2 CALL gds.alpha.pipeline.nodeClassification.addMLP('pipe',
3 {classWeights: [0.4,0.3,0.3], focusWeight: 0.5})
4 YIELD parameterSpace;
```



The screenshot shows a table titled 'parameterSpace' with a single row of configuration parameters. The parameters include settings for a MultilayerPerceptron (MLP) model, a Random Forest model, and a Logistic Regression model. The MLP model is configured with 100 epochs, a learning rate of 0.001, and a focus weight of 0.5. The Random Forest model is configured with a maximum depth of 2147483647 and a minimum leaf size of 1. The Logistic Regression model is configured with 500 epochs and a learning rate of 0.001. The class weights are set to [0.4, 0.3, 0.3] and the focus weight is 0.5.

Figura 4.14. Model de perceptron multistrat cu pierdere focală

Fiecare dintre candidați model descriși mai sus poate fi configurat și ajustat prin intermediul parametrilor pentru a obține performanțe optime în funcție de setul de date și de cerințele problemei de clasificare. Se recomandă utilizarea diferitor candidați model și utilizarea validării încrucișate pentru a determina care dintre aceștia se potrivește cel mai bine cu datele și cu obiectivul de clasificare. Aceste proceduri permit adăugarea de configurații de model pentru a explora diferite seturi de parametri și pentru a testa multiple combinații, astfel încât să se poată găsi cea mai potrivită configurație pentru problema de clasificare a nodurilor în discuție.

Înainte de instruirea modelului trebuie adăugat cel puțin un candidat model. Alegerea unui candidat model adecvat și ajustarea parametrilor potențează capacitatea modelului de a face predicții precise și relevante pe baza caracteristicilor din graf pentru a obține un model mai bun și mai eficient pentru rezolvarea problemei de clasificare a nodurilor.

Reglarea automată a parametrilor

Pentru a găsi cele mai bune modele, conductele permit utilizarea reglării automate a parametrilor algoritmului de instruire. În versiunea curentă a *GDS* (2.4.1), este posibilă doar configurarea numărului maxim de încercări de autoconfigurare. O instrucțiune și rezultatul acesteia este reprezentată în figura 4.14.

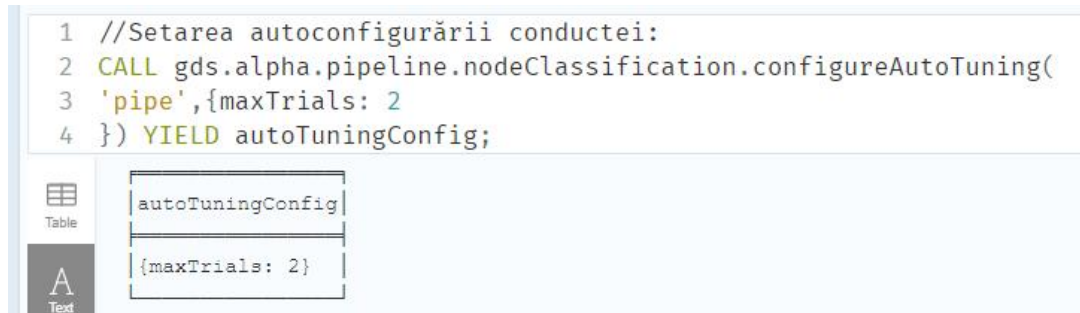


Figura 4.15. Reglarea automată a parametrilor

Instruirea conductei

Modulul de instruire, *gds.beta.pipeline.nodeClassification.train*, are responsabilitatea de a efectua următoarele operații: împărțirea datelor, extragerea caracteristicilor, selectarea modelului, antrenarea acestuia și stocarea lui pentru utilizări ulterioare. Execuția acestui modul, rezultă un model de clasificare de tip *NodeClassification*, care este apoi salvat în catalogul de modele. Acest model de clasificare poate fi aplicat pe diferite grafuri pentru a clasifica nodurile acestora.

```
//Instruirea modelului  
CALL gds.beta.pipeline.nodeClassification.train('dtPred', {  
  pipeline: 'pipe',  
  targetNodeLabels: ['YHTCc'],  
  modelName: 'nc-pipeline-model1',  
  targetProperty: 'drought',  
  randomSeed: 101,  
  metrics: ['ACCURACY', 'OUT_OF_BAG_ERROR']  
}) YIELD modelInfo, modelSelectionStats  
RETURN  
  modelInfo.bestParameters AS winningModel,  
  modelInfo.metrics.ACCURACY.train.avg AS avgTrainScore,  
  modelInfo.metrics.ACCURACY.outerTrain AS outerTrainScore,  
  modelInfo.metrics.ACCURACY.test AS testScore,  
  [cand IN modelSelectionStats.modelCandidates | cand.metrics.ACCURACY.  
  validation.avg] AS validationScores;
```



```
weather$ //Instruirea modelului CALL gds.beta.pipeline.nodeClassification.train('dt
```

winningModel	avgTrainScore	outerTrainScore	testScore	validationScores
{maxDepth: 2147483647, minSplitSize: 1, criterion: "GINI", minSplitSize: 2, numberOfDecisionTrees: 5, methodName: "RandomForest", numberOfSamplesRatio: 1.0}	0.554545458	0.64285715	1.0	[0.5666666699999999, 0.433333338, 0.500000004, 0.500000004]

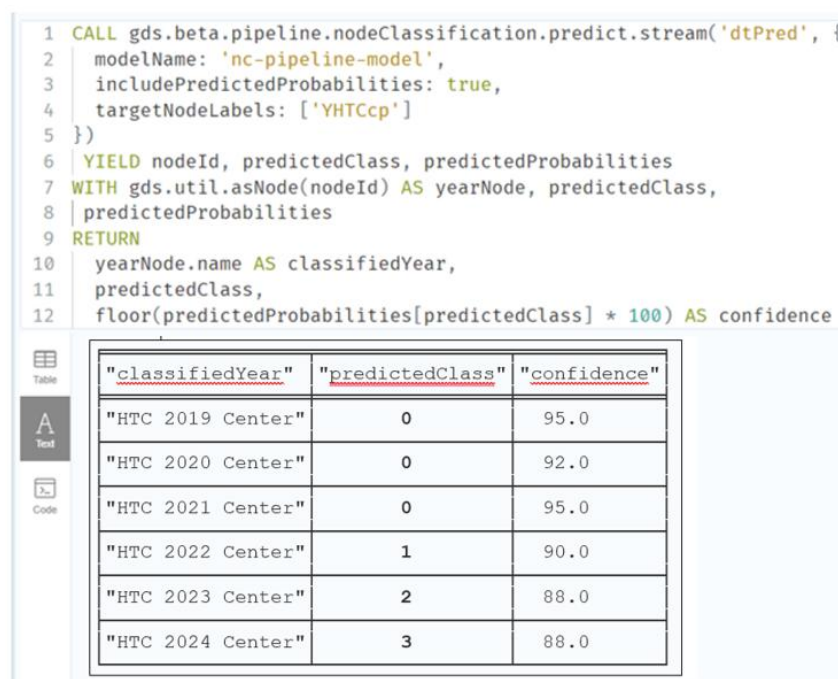
Figura 4.16. Rezultatul instruirii conductei

Codul de instruire a conductei folosește funcția *gds.beta.pipeline.nodeClassification.train*, apelată prin comanda *CALL*, pentru a antrena un model de clasificare a nodurilor.

- Parametrul *dtPred* specifică numele grafului în memorie;
- Parametrul *pipeline* specifică numele conductei „*pipe*” care va fi folosit pentru antrenare.
- Parametrul *targetNodeLabels* este o listă de etichete de noduri care vor fi luate în considerare ca etichete țintă pentru antrenare. În acest exemplu, se folosește doar o etichetă, 'YHTCc' care reprezintă tipul de noduri care determină valoarea coeficientului hidrotermic anual.
- Parametrul *modelName* specifică numele modelului antrenat rezultat. În acest caz, numele modelului este „*nc-pipeline-mod*”.
- Parametrul *targetProperty* specifică proprietatea „*drought*” a nodurilor care conține etichetele țintă.
- Parametrul *randomSeed* specifică o valoare pentru a asigura reproductibilitatea antrenării (capacitatea de a obține rezultate identice sau foarte similare în cadrul procesului de antrenare a unui model de *machine learning*, utilizând aceleași seturi de date și aceleași configurații). În acest exemplu, valoarea aleasă este 1337.
- Parametrul *metrics* este o listă de metrici care vor fi calculate pentru evaluarea modelului. În acest model, se calculează metricile „*ACCURACY*” (acuratețe) și „*OUT_OF_BAG_ERROR*” (eroare pe setul out-of-bag).
- Blocul *YIELD* este folosit pentru a specifica ce informații să fie returnate de funcția de antrenare care vor fi stocate în variabilele *modelInfo* și *modelSelectionStats*.
- În blocul *RETURN*, se extrag anumite informații din variabilele *modelInfo* și *modelSelectionStats*:
 - ✓ *modelInfo.bestParameters* reprezintă setul de parametri care a obținut cel mai bun rezultat în cadrul selecției de modele.

- ✓ `modelInfo.metrics.ACCURACY.train.avg` reprezintă scorul mediu de antrenare pentru metrica 'ACCURACY'.
- ✓ `modelInfo.metrics.ACCURACY.outerTrain` reprezintă scorul de antrenare extern pentru metrica 'ACCURACY'.
- ✓ `modelInfo.metrics.ACCURACY.test` reprezintă scorul de testare pentru metrica 'ACCURACY'.
- ✓ `[cand IN modelSelectionStats.modelCandidates | cand.metrics.ACCURACY.validation.avg]` reprezintă o listă a scorurilor de validare pentru metrica 'ACCURACY' ale tuturor candidaților de model din selecția de modele.

Testarea modelului de predicție folosind date istorice și ajustarea acestuia dacă este necesar pentru a-i îmbunătăți acuratețea. Rezultatul testului este specificat în figura 4.17:



```

1 CALL gds.beta.pipeline.nodeClassification.predict.stream('dtPred', {
2   modelName: 'nc-pipeline-model',
3   includePredictedProbabilities: true,
4   targetNodeLabels: ['YHTCcp']
5 })
6 YIELD nodeId, predictedClass, predictedProbabilities
7 WITH gds.util.asNode(nodeId) AS yearNode, predictedClass,
8   predictedProbabilities
9 RETURN
10 yearNode.name AS classifiedYear,
11 predictedClass,
12 floor(predictedProbabilities[predictedClass] * 100) AS confidence

```

"classifiedYear"	"predictedClass"	"confidence"
"HTC 2019 Center"	0	95.0
"HTC 2020 Center"	0	92.0
"HTC 2021 Center"	0	95.0
"HTC 2022 Center"	1	90.0
"HTC 2023 Center"	2	88.0
"HTC 2024 Center"	3	88.0

Figura 4.17. Rezultatul testării predicției

Modelul de predicție a secetei este prezentat în anexa 30.

4.4. Model de predicție a productivității culturii de soia

În mod similar a fost elaborat un model de predicție a productivității unor culturi agricole. S-a luat ca exemplu cultura de soia. **Scopul** principal al unui model de predicție a productivității culturilor agricole, în special a soiei, este evaluarea riscurilor și luarea deciziilor corecte. Modelul estimează producția viitoare a recoltei luând în considerare factori precum condițiile meteorologice, tipul de sol, input-urile agricole (cum ar fi fertilizanții, erbicidele) și alte variabile

importante, cum ar fi umiditatea, temperatura, luminozitatea și nivelul de infestare cu boli și dăunători. Scopurile modelului includ prognozarea producției viitoare, identificarea factorilor de influență, optimizarea practicilor agricole, îmbunătățirea rentabilității și reducerea riscului și creșterea rezilienței. Minimizarea impactului schimbărilor climatice asupra mediului și a comunităților umane, gestionarea resurselor naturale, precum și dezvoltarea economică și socială durabilă.

Obiectivele modelului de predicție a productivității includ:

- a) Prognoza producției viitoare a recoltei în funcție de condițiile meteorologice și alte variabile relevante. Acest lucru permite agricultorilor să obțină estimări asupra cât de mare va fi producția în viitor și să se pregătească în consecință.
- b) Identificarea factorilor de influență care afectează productivitatea culturii de soia. Modelul poate analiza diferiți factori și poate determina care dintre aceștia au cea mai mare influență asupra producției. Această informație poate fi folosită pentru a dezvolta strategii mai eficiente de cultivare a culturii de soia.
- c) Optimizarea practicilor agricole. Modelul poate identifica cele mai bune practici agricole și poate oferi recomandări pentru a optimiza utilizarea input-urilor agricole, cum ar fi irigarea și aplicarea de îngrășăminte. Aceasta ajută agricultorii să ia decizii mai bine informate în timp real, pe parcursul sezonului de cultură.
- d) Îmbunătățirea rentabilității. Prin furnizarea de estimări precise ale producției viitoare, modelul poate ajuta agricultorii să reducă costurile și să crească producția, ceea ce duce la o creștere a rentabilității activității agricole.
- e) Reducerea riscului și creșterea rezilienței culturilor agricole. Prin identificarea condițiilor care pot afecta productivitatea și furnizarea de informații în timp util, modelul permite agricultorilor să ia măsuri preventive pentru a minimiza impactul negativ al acestor condiții și să crească rezistența culturilor la factorii de stres.

În continuare sunt redată instruirea și testarea conduitei de predicției.

Instruirea conduitei. Datele de învățare sunt utilizate pentru a instrui modelul de predicție. Acest lucru poate fi realizat prin utilizarea diferitelor tehnici de învățare automată, cum ar fi algoritmi de clasificare (Clasificarea nodurilor). Rezultatul instruirii este reprezentat în figura 4.17.

```
CALL gds.beta.pipeline.nodeClassification.train('prodPred', {  
  pipeline: 'pipe', targetNodeLabels: ['Prod'],  
  modelName: 'nc-pipeline-model',  
  targetProperty: 'harvest',  
  randomSeed: 1337,  
  metrics: ['ACCURACY', 'OUT_OF_BAG_ERROR']
```

```

}) YIELD modelInfo, modelSelectionStats
RETURN
  modelInfo.bestParameters AS winningModel,
  modelInfo.metrics.ACCURACY.train.avg AS avgTrainScore,
  modelInfo.metrics.ACCURACY.outerTrain AS outerTrainScore,
  modelInfo.metrics.ACCURACY.test AS testScore,
  [cand IN modelSelectionStats.modelCandidates |
  cand.metrics.ACCURACY.validation.avg] AS validationScores

```

winningModel	avgTrainScore	outerTrainScore	testScore	validationScores
{maxEpochs: 100, minEpochs: 1, cla ssWeights: [], penalty: 0.0, patie nce: 1, methodName: "LogisticRegre ssion", focusWeight: 0.0, batchSiz e: 100, tolerance: 0.001, learning Rate: 0.001}	0.6927272739999999	0.15384616	0.0	[0.7000000002, 0.16666666800000002, 0.0, 0.7000000002, 0.7000000002]

Figura 4.18. Instruirea modelului.

Validarea modelului: După ce modelul a fost instruit, acesta este validat pentru a asigura că funcționează corect. Acest lucru se face prin utilizarea setului de date de testare pentru a evalua performanța modelului.

```

CALL gds.beta.pipeline.nodeClassification.predict.stream('prodPred',
{modelName: 'nc-pipeline-model',includePredictedProbabilities: true,
targetNodeLabels: ['Prodp']})
YIELD nodeId, predictedClass, predictedProbabilities
WITH gds.util.asNode(nodeId) AS harvestNode, predictedClass,
predictedProbabilities
RETURN
  harvestNode.name AS classifiedYear, predictedClass,
  floor(predictedProbabilities[predictedClass] * 100) AS confidence
ORDER BY classifiedYear

```

classifiedYear	<u>predictedClass</u>	confidence
"Soy Prod 2019 Center"	14	95
"Soy Prod 2020 Center"	14	96
"Soy Prod 2021 Center"	20	95

Figura 4.19. Rezultatul predicției productivității.

Testarea modelului de predicție folosind date istorice și ajustarea acestuia dacă este necesar pentru a-i îmbunătăți acuratețea. Pentru aceasta se pot utiliza unele metrice precum acuratețea, sensibilitatea și specificitatea pentru a evalua performanța modelului.

Modelul de predicție a productivității unor culturi agricole este prezentat în anexa 31.

4.5. Concluzii la capitolul 4.

A fost explorată utilizarea tehnologiei de depozit de date graf pe platforma *Neo4j*, pentru a gestiona și analiza datele meteorologice în contextul optimizării proceselor agricole prin

analiza datelor meteorologice și predicția fenomenelor relevante. A fost investigat modul în care datele meteorologice și datele fenomice pot fi reprezentate într-o structură de tip graf, unde fiecare nod reprezintă o măsurătoare specifică, cum ar fi temperatura, precipitațiile sau umiditatea solului, iar relațiile dintre noduri pot indica evoluțiile în timp, conexiunile sau corelații dintre diversele variabile meteorologice. Structurile de tip graf permit o înțelegere mai profundă a interacțiunilor complexe dintre diferitele aspecte meteorologice, ceea ce poate contribui la înțelegerea mai profundă a influenței acestora asupra proceselor agricole. conduce la identificarea de modele sau tendințe care altfel ar putea să nu fie observate.

Rezultatele acestor cercetări au fost prezentate la:

- Salonul Internațional de Inovații și „Inventcor” de la Deva în anii 2021 și 2022 unde au fost apreciate cu 2 medalii de aur;
- Salonul Internațional de Inovații și „Traian Vuia” de la Timișoara în anii 2021 și 2022 unde au fost apreciate cu o medalie de bronz;
- Salonul Internațional „EuroInvent”, Iași, 2022 și 2023, distinse cu o medalie de argint și Diplomă de mențiune.
- La AGEPI a fost înregistrat un drept de autor: „Model de predicție a productivității unor culturi agricole”, OȘ nr. 7552, 21.06.2023.

Integrarea SGBD *Neo4j* cu limbajul *Cypher* în procesul de analiză a datelor a demonstrat, că poate accelera procesul de identificare a corelațiilor și tendințelor de evoluare a relațiilor studiate. Pot fi prezise unele fenomene din domeniul cercetat pe baza datelor stocate în depozitul de date: Depozitul de date graf reprezintă o bază solidă pentru realizarea predicțiilor legate de fenomenele meteorologice și productivitatea unor culturi agricole. Utilizând informațiile depozitate în depozitul de date graf, pot fi dezvoltate modele de predicție pentru fenomene meteorologice sau referitoare la impactul acestora asupra proceselor agricole.

Modelele orientate pe familii de probleme elaborate pot furniza cercetătorilor date referitor la influența unor substanțe în creșterea și dezvoltare plantelor. Fiind orientate pe FP modelele pot fi folosite și pentru validarea altor experimente folosind tehnicile de învățare prin transfer. Agricultorilor le poate oferi informații cruciale despre momentul optim pentru plantare, recoltare, irigare și aplicarea îngrășămintelor, toate acestea contribuind la îmbunătățirea randamentului și eficienței agricole. Acest subiect se referă la procesul de creare și optimizare a modelelor de predicție pe baza datelor stocate în depozitul de date graf. Modelele elaborate au fost integrate într-un *Sistem Suport pentru Decizii* (SSD).

Utilizarea datelor stocate în depozitul de date graf pentru antrenarea modelului a adus avantaje semnificative. Modelul de predicție poate să învețe din conexiunile dintre diverse

variabile meteorologice și să furnizeze estimări viitoare mai exacte, contribuind la luarea deciziilor în procesele agricole.

În final, utilizarea unui depozit de date graf și a tehnicilor asociate acestuia în contextul optimizării proceselor agricole aduce avantaje semnificative prin posibilitatea de a evidenția conexiunile complexe dintre datele meteorologice și performanța culturilor agricole. Aceasta duce la o gestionare mai inteligentă a resurselor agricole și la luarea de decizii mai informate pentru sporirea rezilienței, a randamentului și sustenabilității în agricultură. Este important de menționat că aceste modele de predicție pot fi susceptibile la anumite erori și inexactități, iar unul dintre factorii principali poate fi volumul redus de date disponibile. Cu toate acestea, pe măsură ce se acumulează mai multe date în depozitul de date, există o perspectivă promițătoare pentru îmbunătățirea și rafinarea acestor modele. Prin extinderea bazei de date, vom avea o capacitate mai mare de a captura relațiile subiacente și de a obține predicții mai precise și robuste.

CONCLUZII GENERALE ȘI RECOMANDĂRI

Analiza inteligentă a datelor în fiziologia plantelor este un subiect de cercetare actual datorită actualității din motiv că permite obținerea de date mai detaliate și mai diverse cu privire la procesele fiziologice ale plantelor, care permit optimizarea creșterii și producției plantelor agricole. Utilizarea tehnologiilor inteligente pot fi utilizate cu succes pentru a colecta, analiza și interpreta datele în timp real, ceea ce poate ajuta la îmbunătățirea randamentului culturilor, reducerea costurilor și impactului asupra mediului și la optimizarea utilizării resurselor. În urma cercetărilor efectuate, se pot face următoarele cinci concluzii generale:

1. **A fost elaborat un limbaj formal** al domeniului de cercetare „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor” (capitolul „Integrarea metodelor biostatisticii și a depozitelor de date în procesarea datelor din domeniul fiziologiei plantelor”). Limbajul este de complexitate considerabilă. Aplicarea tehnologiilor informaționale convenționale ar putea solicita implicarea unei Companii IT pentru realizarea acestuia pe calculator. Metodologia originală de dezvoltare a produselor software orientate pe familii de probleme, folosită în cadrul cercetării doctorale a permis unui singur informatician (autorului prezentei teze de doctorat) să realizeze pe calculator limbajul formal al domeniului de cercetare „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor”.

2. **A fost dezvoltată o bază de cunoștințe** a domeniului de cercetare „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor” (capitolul „Integrarea metodelor biostatisticii și a depozitelor de date în procesarea datelor din domeniul fiziologiei plantelor”). Baza de cunoștințe a fost folosită la elaborarea modelelor și produselor program aplicative dezvoltate și a fost integrată în produsele program dezvoltate.

3. **Au fost dezvoltate** modele de analiză inteligentă a datelor orientate pe familii de probleme pentru domeniul de cercetare „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor” (capitolul „Aplicații inteligente ale metodelor biostatisticii în cercetări de fiziologie a plantelor”). Modelele elaborate au fost premiate cu: două Medalii de Aur, o Medalie de Argint, o Medalie de Bronz și o Diplomă de Excelență la Saloane Internaționale ale Inovării și Cercetării Științifice. La AGEPI au fost înregistrate două drepturi de autor.

4. **A fost dezvoltată o bază de date nerelațională** (de tip graf) *Neo4j* folosită în mediul *Wolfram Mathematica*. Spre deosebire de bazele de date relaționale bazele de date de tip graf demonstrează următoarele avantaje (capitolul „Depozitul de date graf: Evidențierea conexiunilor pentru optimizarea proceselor agricole”):

Posibilitățile nominalizate ale bazelor de date graf largesc considerabil aria de aplicare a produselor program aplicative care utilizează acest tip de baze de date.

5. **Produsele program dezvoltate** au fost integrate într-un SSD orientat pe familii de probleme pentru domeniul de cercetare „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor”.

Recomandări de cercetare viitoare

Cercetarea s-a axat pe utilizarea SSD orientat pe familii de probleme pentru domeniul de cercetare „Modelarea biostatistică a datelor în domeniul fiziologiei plantelor”.

S-a observat că această metodologie de cercetare poate fi extinsă și pentru alte culturi agricole. În baza acestor cercetări se recomandă:

- a) De a folosi în continuare modele, tehnicile și produsele program de analiză inteligentă a datelor în fiziologia plantelor pentru cercetările științifice la Facultatea Biologie și Geștiințe.
- b) De a extinde aria de aplicare a SSD pentru modelarea biostatistică a datelor în domeniul fiziologiei plantelor la alte Universități din Republica Moldova și peste hotare.
- c) De a extinde cercetările de modelare biostatistică a datelor în domeniul fiziologiei plantelor pentru alte culturi agricole în cooperare cu alte Instituții din Republica Moldova și peste hotare.
- d) Completarea bazei de date și a depozitului de date cu date și informații noi, precum și extinderea surselor și instrumentelor în vederea creării unui *lac de date (data lake)*.
- e) Dezvoltarea continuă a SSD și integrarea acestuia într-un sistem colaborativ.

BIBLIOGRAFIE

1. MASLOW, A. Motivație și personalitate, editura Trei, București, 2013.
2. Metode matematice în problematica dezvoltării. Coord. Solomon Marcus. Buc. 1982.
3. SCRITUBE. Nevoile umane. [citat: 01.12.2019] Disponibil: <https://www.scritub.com/sociologie/resurse-umane/Nevoile-umane> 195422133.php
4. DEXONLINE. Tehnologie. [citat: 16.06.2020] Disponibil: <https://dexonline.ro/definitie/tehnologie>
5. MARCU, Florin. Dicționar de neologisme. Știința, Ch. 2011.
6. WYMAN, O., De CLERCQ, Vats, M., BIEL, A. Agriculture 4.0 the future of farming technology, 2018 [citat: 12.10.2021]. Disponibil: <https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2021/apr/agriculture-4-0-the-future-of-farming-technology.pdf>
7. BASF. Totul despre agricultura de precizie: soluții tehnologice disponibile @2022. [citat: 12.06.2022] Disponibil: <https://www.agro.basf.ro/ro/stiri/basf-in-camp/agricultura-de-precizie-ce-este-cum-ajuta-fermierii.html>
8. AGFUNDERNEWS. Regenerative Agriculture 5.0: The Future of Farming?, @2020 [citat: 5.08.2020]. Disponibil: <https://agfundernews.com/regenerative-agriculture-5-0-the-future-of-farming.html>.
9. GOLAN, L. HELLER, A. How AI and ML can translate data into wisdom, 7 Dec 2020. [citat: 28.08.2022]. Disponibil: <https://www.nokia.com/blog/how-ai-and-ml-can-translate-data-into-wisdom/>
10. DEXONLINE. Dicționare ale limbii române. Date, @2022 [citat: 18.08.2022]. Disponibil: <https://dexonline.ro/definitie/date>
11. CATARON, A. Curs baze de date. [citat: 18.08.2022]. Disponibil: http://vega.unitbv.ro/~cataron/Courses/BD/BD_Cap_1.pdf
12. DEXONLINE. Dicționare ale limbii române. Date, @2022 [citat: 18.08.2022]. Disponibil: <https://dexonline.ro/definitie/informație>
13. SFETCU, N. Cunoaștere și Informații. Buc. MultiMedia Publishing, 2019.
14. ОСУГА, С. Обработка знаний. /Trad. din limba japoneza, M.: Мир, 1989.
15. GANEA, I. Elaboration of the Graph Type Model for the Analysis of Biological Data. In Extended Abstracts Volume of the International Student Conference „StudMath-IT”, „Aurel Vlaicu” University of Arad, November 18th-19th, 2021.
16. DATAROBOT. What are Data Insights? @2022 [citat: 01.09.2022]. Disponibil: <https://www.datarobot.com/wiki/insights/#:~:text=What%20are%20Data%20Insights%3F>

[&text=Data%20insights%20refer%20to%20the,technology%20to%20analyze%20a%20dataset.](#)

17. INSIGHT. What is Artificial Intelligence? @2023 [citat: 01.02.2023]. Disponibil: https://www.insight.com/en_US/content-and-resources/glossary/a/artificial-intelligence.html
18. PETCU M. Sisteme Expert. Capitolul 1. [citat: 21.08.2022]. Disponibil: https://www.academia.edu/7985477/Capitolul_1_SISTEME_EXPERT
19. DEXONLINE. Dicționare ale limbii române. Date, @2020 [citat: 14.08.2020]. Disponibil: <https://dexonline.ro/definitie/inteligenta>
20. ELEMENT of AI. How should we define AI? @2022 [citat: 09.05.2022]. Disponibil: <https://course.elementsofai.com/1/1>
21. TECHCHARGET. Enterprise AI. Artificial Intelligence (AI). @2022 [citat: 10.05.2022]. Disponibil: <https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence>
22. MCCARTHY, J. WHAT IS ARTIFICIAL INTELLIGENCE? Computer Science Department Stanford University, 2007 [citat: 19.05.2022]. Disponibil: <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>
23. ANDONIE, R. CATARON, A. Inteligența computațională. În „Cursul de Inteligența Artificială” - Universitatea Transilvania din Brașov, 2002.
24. SPRINGERLINK. Artificial wisdom: a philosophical framework. @2022 [citat: 11.05.2022]. Disponibil: <https://springer.com/article/10.1007/s00146-020-00949-5>
25. CĂRBUNE, V. ERHAN, C. Motor de inferență generic cu arhitectură reconfigurabilă pentru sisteme autonome de luare a deciziilor. În: Conferința Tehnico-Științifică a Colaboratorilor, Doctoranzilor și Studenților, Universitatea Tehnică a Moldovei, 15-17 noiembrie, 2012. Chișinău, 2013, vol. 1, pp. 145-148. ISBN 978-9975-45-249-6. ISBN 978-9975-45-250-2 (Vol.1). Disponibil: <http://repository.utm.md/handle/5014/1198?show=full>
26. JESTE, D., GRAHAM, S., NGUYEN, T., DEPP, C., LEE, E., & KIM, H. (2020). Beyond artificial intelligence: Exploring artificial wisdom. In: International Psychogeriatrics, 32(8), 993-1001. ISSN: 1041-6102 (Print), 1741-203X (Online) <https://doi.org/10.1017/S1041610220000927>. Disponibil: <https://escholarship.org/uc/item/6z338854>.
27. ACTUALITATE. PARLAMENTUL EUROPEAN. Ce este inteligența artificială și cum este utilizată? <https://www.europarl.europa.eu/news/ro/headlines/society/20200827STO85804/ce-este-inteligenta-artificiala-si-cum-este-utilizata>

28. PEDAMKAR, P. How Artificial Intelligence Works? Disponibil: <https://www.educba.com/how-artificial-intelligence-works/>
29. HAN, J., KAMBER M., PEI, J. Data mining : concepts and techniques – 3rd ed. p. cm. Elsevier, 2012. ISBN 978-0-12-381479-1.
30. SCHUH, G., REINHART, G., PROTE et all. Data Mining Definitions and Applications for the Management of Production Complexity, Procedia CIRP, Volume 81, 2019, Pages 874-879, ISSN 2212-8271, <https://doi.org/10.1016/j.procir.2019.03.217>. Disponibil: <https://www.sciencedirect.com/science/article/pii/S2212827119305220>
31. CIOS, K.J., PEDRYCZ, W., SWINIARSKI, R.W. (1998). Data Mining and Knowledge Discovery. In: Data Mining Methods for Knowledge Discovery. The Springer International Series in Engineering and Computer Science, vol 458. Springer, Boston, MA. Disponibil: https://doi.org/10.1007/978-1-4615-5589-6_1
32. FAYYAD, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. AI Magazine, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230>
33. KRZYSZTOF, J., CIOS, W., PEDRYCZ, R., SWINIARSKI, W. Data mining and knowledge discovery. Disponibil: https://www.doi:10.1007/978-1-4615-5589-6_1
34. KURGAN, Lukasz & MUSILEK, Petr. A survey of Knowledge Discovery and Data Mining process models. Knowledge Eng. Review. 21. 1-24. 10.1017/S0269888906000737, 2006.
35. JAVATPOINT. Difference between Artificial intelligence and Machine learning. @2021 [citat: 10.05.2022]. Disponibil: <https://www.javatpoint.com/difference-between-artificial-intelligence-and-machine-learning>
36. HURBANS, R. Grokking Artificial Intelligence Algorithms. pp. 227-232 [online]. Shelter Island: Manning Publication, 2020 [citat: 10.05.2022]. Disponibil: <https://books.google.md/books?id=hfflDwAAQBAJ&pg=PA355&dq=Artificial+Intelligence+for+Biology+and+Agriculture&hl=en&sa=X&ved=2ahUKEwjn7OmK1Pr2AhXylIsKHV63A4sQ6AF6BAgFEAI#v=onepage&q=Artificial%20Intelligence%20for%20Biology%20and%20Agriculture&f=false>
37. KOLABTREE BLOG. The Applications of Machine Learning in Biology. [citat: 11.05.2022]. Disponibil: <https://www.kolabtree.com/blog/applications-of-machine-learning-in-biology/>
38. XU, C., JACKSON, S.A. Machine learning and complex biological data. Genome Biol 20, 76 (2019). <https://doi.org/10.1186/s13059-019-1689-0>. [citat: 10.05.2022]. Disponibil: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1689-0>.
39. SHOAI, M., SHAH, B., EI-SAPPAGH, S., ALI, A., ULLAH, A., ALENEZI, F., GECHEV, T., HUSSAIN, T., ALI, F. An advanced deep learning models-based plant disease detection:

- A review of recent research. *Front. Plant Sci.* 14:1158933, 2023. doi: 10.3389/fpls.2023.1158933.
40. SIN YING GOY; YEN FOOK CHONG; THENG KAH KELVIN TEOH; CHEE CHIN LIM; VIKNESWARAN VIJEAN. Recognition of plant diseases by leaf image classification using deep learning approach. *AIP Conference Proceedings* 2562, 020001 (2023). <https://doi.org/10.1063/5.0112725>.
 41. BÎRSAN, A. SPRINCEAN, V. GANEA, I. Oportunități de utilizare în practică a tehnologiilor inteligente în evaluarea unor variabile biologice. În: *Studia Universitatis Moldaviae*, Nr. 1(2023), pp. 53-64. DOI: [https://doi.org/10.59295/sum1\(171\)2023_07](https://doi.org/10.59295/sum1(171)2023_07).
 42. WEBB, Sarah. Deep learning for biology. *Nature*. 554. 555-557. 2018. 10.1038/d41586-018-02174-z. [citat: 10.05.2022]. Disponibil: <https://www.nature.com/articles/d41586-018-02174-z>.
 43. MOHANTY, SP., HUGHES, DP., SALATHÉ, M. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* 7:1419.2016. doi: 10.3389/fpls.2016.01419. Disponibil: <https://www.frontiersin.org/articles/10.3389/fpls.2016.01419/full>.
 44. TAYLOR, Matthew & STONE, Peter. (2009). Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research*. 10. 1633-1685. 10.1145/1577069.1755839.
 45. DONAHUE, J., HOFFMAN, J., RODNER, Erik & Saenko, TREVOR, Kate & Darrell. Semi-Supervised Domain Adaptation with Instance Constraints. *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2013. 668-675. 10.1109/CVPR.2013.92.
 46. PAN, S. J., YANG, Q. A. Survey on Transfer Learning, in *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.
 47. RUDER, S., PETERS, M. E., SWAYAMDIPTA, S., WOLF, T. 2019. Transfer Learning in Natural Language Processing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, pages 15–18, Minneapolis, Minnesota. Association for Computational Linguistics. DOI: 10.18653/v1/N19-5004. <https://aclanthology.org/N19-5004>.
 48. MWITI, Derrick . Transfer Learning Guide: A Practical Tutorial With Examples for Images and Text in Keras. <https://neptune.ai/blog/transfer-learning-guide-examples-for-images-and-text-in-keras>.

49. ZHANG, J., Li, W., OGUNBONA, Ph., XU, D. Recent Advances in Transfer Learning for Cross-Dataset Visual Recognition: A Problem-Oriented Perspective. arXiv:1705.04396v3 [cs.CV] 20 May 2019. <https://arxiv.org/pdf/1705.04396.pdf>.
50. IANCOVIC, D. Comparing Machine Learning Models: Statistical vs. Practical Significance. [citat:10.05.2022]. Disponibil:<https://www.kdnuggets.com/2019/01/comparing-machine-learning-models-statistical-vs-practical-significance.html>.
51. KDnuggets. Comparing Machine Learning Models: Statistical vs. Practical Significance. Disponibil: <https://www.kdnuggets.com/2019/01/comparing-machine-learning-models-statistical-vs-practical-significance.html>
52. CRAVERO A, PARDO S, GALEAS P, LÓPEZ FENNER J, CANIUPÁN M. Data Type and Data Sources for Agricultural Big Data and Machine Learning. Sustainability. 2022; 14(23):16131. <https://doi.org/10.3390/su142316131>. <https://www.mdpi.com/2071-1050/14/23/16131>.
53. CĂPĂȚÂNĂ Gh. Tehnologii informaționale inteligente. // Fizica și tehnologiile moderne, vol. 1, no. 3, 2003, pp. 9-13. https://ibn.idsi.md/sites/default/files/imag_file/Tehnologii%20informationale%20inteligente.pdf.
54. KUMAR A, PATHAK RK, GUPTA SM, GAUR VS, PANDEY D. Systems Biology for Smart Crops and Agricultural Innovation: Filling the Gaps between Genotype and Phenotype for Complex Traits Linked with Robust Agricultural Productivity and Sustainability. OMICS. 2015 Oct;19(10):581-601. doi: 10.1089/omi.2015.0106. PMID: 26484978; PMCID: PMC4617413.
55. LI, Q., YAN, J. Sustainable agriculture in the era of omics: knowledge-driven crop breeding. Genome Biol 21, 154 (2020). <https://doi.org/10.1186/s13059-020-02073-5>
56. GEBREMARIAM, B., & DINSSA, M. F. (2020). Intelligent Data Analysis for Plant Physiology. Frontiers in Plant Science, 11, 1012. <https://doi.org/10.3389/fpls.2020.01012>.
57. DEVASENAPATHY, S., & KRISHNAN, S. (2018). Smart Agriculture Using IoT and Big Data Analytics. Journal of Big Data, 5(1), 1-23. <https://doi.org/10.1186/s40537-018-0120-x>.
58. FREITAS, P.L. Artificial intelligence in agriculture: A systematic review. Computers and Electronics in Agriculture, 153, 69-80, 2018. <https://doi.org/10.1016/j.compag.2018.08.017>.
59. SAVORY, A. "Holistic Management: A New Framework for Decision Making," Island Press, 1999.
60. MOISE D. E. Abordarea holistică. Scribd. 22.06.2017 [online]. Disponibil: <https://ro.scribd.com/document/351967625/Abordarea-Holistica#>

61. DEXONLINE. Holism. @2022. [citat: 09.03.2020] Disponibil: <https://dexonline.ro/definitie/holism>.
62. SMUTS, J. C. Holism and Evolution. Macmillan, 1926.
63. ENCICLOPEDIA TITANICA. Holistica. [citat: 06.12.2021] Disponibil: <https://ro.encyclopedia-titanica.com/signifi-cado-de-hol-stico>.
64. OTTO, T., BUBANDT, N. Experiments in Holism: Theory and Practice in Contemporary Anthropology. Blackwell, 2010.
65. ASTĂRĂSTOAE, V. Abordarea holistică în medicină. <https://sanatateabuzoiana.ro/prof-univ-dr-vasile-astarastoe-abordarea-holistica-in-medicina/>.
66. KITANO, H. Systems Biology: A Brief Overview, Science, vol. 295, no. 5560, pp. 1662-1664, Mar. 2002.
67. TYNER, J.J., MICHOD, R.E. Exploring the Complexity of Biological Systems, BioScience, vol. 53, no. 10, pp. 976-981, Oct. 2003.
68. JONES, Cristine. The Back Forty Down Under: Adapting Farming to Climate Variability. The Quivira Coalition Journal No. 35, February 2010 [citat: 11.08.2021]. Disponibil: <https://holisticmanagement.org/wp-content/uploads/2016/08/Jones-AdaptingFarming-Quivira-Feb10.pdf>
69. FOLKE, C. et al., "Ecological resilience, biodiversity, and scale," Annual Review of Ecology and Systematics, vol. 31, pp. 277-308, 2000.
70. TAVASSOLY I, GOLDFARB J, IYENGAR R. Systems biology primer: the basic methods and approaches. Essays Biochem. 2018 Oct 26;62(4):487-500. doi: 10.1042/EBC20180003. PMID: 30287586. <https://pubmed.ncbi.nlm.nih.gov/30287586/>.
71. NGUYEN, DANIEL D. KARAVANIC, KAREN L. Workflow Critical Path: A data-oriented critical path metric for Holistic HPC Workflows, BenchCouncil Transactions on Benchmarks, Standards and Evaluations, Volume 1, Issue 1, 2021, 100001, ISSN 2772-4859, <https://doi.org/10.1016/j.tbench.2021.100001>.
72. ISB. What is Systems Biology. <https://isbscience.org/about/what-is-systems-biology/>.
73. XU, C., JACKSON, S.A. Machine learning and complex biological data. Genome Biol 20, 76 (2019). <https://doi.org/10.1186/s13059-019-1689-0>.
74. ARGUESO, CT., ASSMANN, SM., BIRNBAUMKD et al. Directions for research and training in plant omics: Big Questions and Big Data. Plant Direct. 2019;3:1–16. <https://doi.org/10.1002/pld3.133>. <https://onlinelibrary.wiley.com/doi/full/10.1002/pld3.133>.
75. SCIENTIA. <https://www.scientia.ro/univers/76-sisteme-complexe/852-sistemele-complexe-scurta-introducere.html>.

76. SCIENTIA. Sisteme complexe. [citat: 20.12.2019] Disponibil: <https://www.scientia.ro/univers/76-sisteme-complexe/852-sistemele-omplexe-scurta-introducere.html>.
77. MARSH, Jacob I., et al. Crop breeding for a changing climate: Integrating phenomics and genomics with bioinformatics. *Theoretical and Applied Genetics* 134 (2021): 1677-1690.
78. DAI X and SHEN L (2022) Advances and Trends in Omics Technology Development. *Front. Med.* 9:911861. doi: 10.3389/fmed.2022.911861.
79. CRAVERO, A., PARDO, S., GALEAS, P., LÓPEZ FENNER, J., CANIUPÁN, M. Data Type and Data Sources for Agricultural Big Data and Machine Learning. *Sustainability* 2022, 14, 16131. <https://doi.org/10.3390/su142316131>.
80. BADGLEY, C. et al., "Organic Agriculture and the Global Food Supply," Cambridge University Press, 2007.
81. VAN EMON JM. The Omics Revolution in Agricultural Research. *J Agric Food Chem.* 2016 Jan 13;64(1):36-44. doi: 10.1021/acs.jafc.5b04515. Epub 2015 Nov 17. PMID: 26468989; PMCID: PMC4714296. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4714296/>.
82. MISRA, B. B., LANGEFELD, C., OLIVIER, M., & COX, L. A. (2019). Integrated omics: tools, advances and future approaches. *Journal of Molecular Endocrinology*, 62(1), R21-R45. Retrieved Sep 24, 2023, from <https://doi.org/10.1530/JME-18-0055>. <https://jme.bioscientifica.com/view/journals/jme/62/1/JME-18-0055.xml>.
83. MICHEEL CM, NASS SJ, OMENN GS, editors. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington (DC): National Academies Press (US); 2012 Mar 23. 2, *Omics-Based Clinical Discovery: Science, Technology, and Applications*. <https://www.ncbi.nlm.nih.gov/books/NBK202165/>.
84. SAVORY, A., BUTTERFIELD, J. The Holistic Management of Rangelands: A Model for Sustainable Livestock Production in Arid Zones, *Journal of Sustainable Agriculture*, vol. 4, no. 2-3, pp. 141-154, 1994.
85. KATSUYUKI Yugi, HIROYUKI Kubota, ATSUSHI Hatano, SHINYA Kuroda. *Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple 'Omic' Layers*, 2016. DOI:<https://doi.org/10.1016/j.tibtech.2015.12.013>. [https://www.cell.com/trends/bio-technology/fulltext/S0167-7799\(15\)00273-5?_returnURL=https%3A%2F%2FDisponibil-in-ghub.elsevier.com%2Fretrieve%2Fpii%2FS0167779915002735%3Fshowall%3Dtrue](https://www.cell.com/trends/bio-technology/fulltext/S0167-7799(15)00273-5?_returnURL=https%3A%2F%2FDisponibil-in-ghub.elsevier.com%2Fretrieve%2Fpii%2FS0167779915002735%3Fshowall%3Dtrue).
86. DALY, H., FARLEY, J. "Ecological Economics: Principles and Applications," Island Press, 2010.

87. SAITO, K. ARAI, S., KATO, H. A nutrigenomics database - Integrated repository for publications and associated microarray data in nutrigenomics research *British Journal of Nutrition*, 94 (2005), pp. 493-495.
88. KATO, Hisanori; TAKAHASHI, Shoko; SAITO, Kenji. Omics and Integrated Omics for the Promotion of Food and Nutrition Science. *Journal of Traditional and Complementary Medicine*, 2011, vol. 1, no. 1, p. 25-30. ISSN 2225-4110. [https://doi.org/10.1016/S2225-4110\(16\)30053-0](https://doi.org/10.1016/S2225-4110(16)30053-0).
89. BALKIR, Pinar, KEMAHLIOGLU, Kemal, YUCEL, Ufuk. Foodomics: A new approach in food quality and safety, *Trends in Food Science & Technology*, Volume 108, 2021, Pages 49-57, ISSN 0924-2244, <https://doi.org/10.1016/j.tifs.2020.11.028>.
90. MOCHIDA, Keiichi; SHINOZAKI, Kazuo. Advances in Omics and Bioinformatics Tools for System Analyses of Plant Functions. *Plant and Cell Physiology*, Volume 52, Issue 12, December 2011, Pages 2017–2038. DOI: 10.1093/pcp/pcp153.
91. ZHAO, Xiangyu; PAN, Shouhui; LIU, Zhongqiang; HAN, Yanyun; ZHANG, Qi; WANG, Kaiyi. Intelligent upgrading of plant breeding: Decision support tools in the golden seed breeding cloud platform. *Computers and Electronics in Agriculture*, 2022, vol. 194, 106672. ISSN 0168-1699. DOI: 10.1016/j.compag.2021.106672. [Online]. Available at: <https://www.sciencedirect.com/science/article/pii/S016816992100689X>.
92. BERSANELLI, Matteo et al. “Methods for the integration of multi-omics data: mathematical aspects.” *BMC bioinformatics* vol. 17 Suppl 2, Suppl 2 15. 20 Jan. 2016, doi:10.1186/s12859-015-0857-9.
93. TAIZ, L., ZEIGER, E. *Plant Physiology : International Edition*. Edition statement 5th Revised edition. Sunderland, United States 2010, 782 p.], 2014. *Plant physiology and development*. Sinauer Associates, Inc.
94. SADRAS, Victor, CALDERNI, D.F. *Crop Physiology: Applications for Genetic Improvement and Agronomy: Second Edition*. 1-551, 2014.
95. GHOSAL, A.; ZHENG, S.; CHAPMAN, N.; LEAFGREN, S. A.; JORDAN, S.; HIRNING, J. J. Transfer Learning for Plant Phenotyping. In: arXiv, 2018. Available at: <https://arxiv.org/abs/1711.05458>.
96. YU, K.; ZHANG, Y.; CHAKRABARTI, S.; WANG, X.; WANG, C. Transfer Learning for Improved Crop Mapping. In: *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018. Available at: <https://www.sciencedirect.com/science/article/pii/S0924271617302666>.
97. ZHU, X.; TUIA, S.; MOU, L.; XIA, L.; ZHANG, Y.; XU, F.; FRAUNDORFER, F. Transfer Learning for Hyperspectral Image Classification: A Convolutional Neural Network with

- Optimized Knowledge Accumulation. In: IEEE Transactions on Geoscience and Remote Sensing, 2017. Available at: <https://ieeexplore.ieee.org/document/7778781>.
98. BÎRSAN, Ana. Specii de plante alogene invazive și potențial invazive în rezervația cultural-naturală „Orheiul Vechi” Studia Universitatis Moldaviae, 2022, nr.6(156). Seria “științe reale și ale naturii”. ISSN 1814-3237 ISSN online 1857-498X p.64-71 64. CZU: 581.1:502.62(478). DOI: <https://doi.org/10.5281/zenodo.7442419>
 99. MOHANTY, S.; HUGHES, D. P.; SALATHÉ, M. Transfer Learning for Plant Disease Classification and Distribution Modeling in Precision Agriculture. In: arXiv, 2016. Available at: <https://arxiv.org/abs/1612.07333>.
 100. BHANDARI, N.; ZHANG, S.; KHATIWADA, B. Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. In: Remote Sensing, 2019. Available at: <https://www.mdpi.com/2072-4292/11/14/1667>.
 101. MOU, C.; ZHU, X.; XIA, R.; ZHANG, L. Transfer Learning in Remote Sensing Applications: A Survey of Recent Advances. In: ISPRS Journal of Photogrammetry and Remote Sensing, 2018. [citat: 01.06.2023] Disponibil: at: <https://www.sciencedirect.com/science/article/pii/S0924271617303027>.
 102. BADEA, A., NICULESCU, S., DANILA, D., BADEA, L. Hyperspectral imaging for plant phenotyping: current status and future perspectives" (Imagistica hiper-spectrală pentru fenotipizarea plantelor: starea actuală și perspective viitoare) - Journal of Agroalimentary Processes and Technologies, vol. 25, no. 2, 2019, pp. 260-266.
 103. COVINGTON, M. F., et al. "Intelligent data analysis for plant growth and development." IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 5, 2015, pp. 878-891.
 104. WANG, R., & CRAWFORD, N. M. (2018). Nitrate signaling: sensor molecules, signal transduction pathways, and regulatory mechanisms. Trends in plant science, 23(11), 927-939.
 105. BARGMANN, B. O., MUNNÉ-BOSCH, S. (2021). Heat stress in plants: an overview of stress signals, responses, and adaptation mechanisms. Journal of Plant Physiology, 257, 153352.
 106. ROSA, M., HILAL, M., GONZALEZ, J. A., PRADO, F. E. (2019). Low temperature response in salt-stressed quinoa (*Chenopodium quinoa* Willd.) seedlings. Plant physiology and biochemistry, 139, 169-178.

107. PURCAREA, C., PURCAREA, V., ROBU, T., MIHĂESCU, T. Modeling of photosynthesis using artificial neural networks (Modelarea fotosintezei utilizând rețele neuronale artificiale) - *Procedia Computer Science*, vol. 181, 2021, pp. 89-98.
108. SINGH, Gurjeet & Kalra, NEETU, Naresh & Yadav & Sharma, MANOJ, AShwani & Saini. SMART AGRICULTURE: A REVIEW. *Siberian Journal of Life Sciences and Agriculture*. 14. 423-454, 2022. 10.12731/2658-6649-2022-14-6-423-454.
109. LEE, D.I.; LEE, J.H.; JANG, S.H.; OH, S.J.; DOO, I.C. Crop Disease Diagnosis with Deep Learning-Based Image Captioning and Object Detection. *Appl. Sci.* 2023, 13, 3148. <https://doi.org/10.3390/app13053148>.
110. KUSWIDIYANTO, L.W.; NOH, H.-H.; HAN, X. Plant Disease Diagnosis Using Deep Learning Based on Aerial Hyperspectral Images: A Review. *Remote Sens.* 2022, 14, 6031. <https://doi.org/10.3390/rs14236031>.
111. CHANYOUNG, JU, JEONGEUN, KIM, JAEHWI, SEOL, HYOUNG IL SON, A review on multirobot systems in agriculture, *Computers and Electronics in Agriculture*, Volume 202, 2022, 107336, ISSN 0168-1699, [Online]. Available at: <https://www.science-direct.com/science/article/pii/S0168169922006457>.
112. BOTTA A, CAVALLONE P, BAGLIERI L, COLUCCI G, TAGLIAVINI L, QUAGLIA G. A Review of Robots, Perception, and Tasks in Precision Agriculture. *Applied Mechanics*. 2022; 3(3):830-854. <https://doi.org/10.3390/applmech3030049>.
113. LIU, X., et al. "Machine learning-based prediction of yield and plant traits in maize using UAV-based hyperspectral data." *Remote Sensing*, vol. 12, no. 18, 2020, pp. 1-22.
114. MEHMOOD, T., et al. "Machine learning for plant stress phenotyping: usage, challenges and future opportunities." *Journal of Experimental Botany*, vol. 72, no. 11, 2021, pp. 3843-3864.
115. SHI, Y., et al. "Deep learning for plant stress phenotyping: trends and future perspectives." *Frontiers in Plant Science*, vol. 10, 2019, pp. 1-11.
116. MOLDOVAN, I.T., BADEA, A., MĂRCULEȚ, C., ȘIMON, A., CÂMPEANU, C., SESTRAS, A.F. Monitoring of soil water content and plant water stress using unmanned aerial vehicle-based thermal and multispectral imagery" (Monitorizarea conținutului de apă din sol și a stresului hidric al plantelor utilizând imagini termice și multispectrale bazate pe vehicul aerian fără pilot) - *Revista de Chimie*, vol. 71, no. 8, 2020, pp. 364-372.
117. VASILESCU, M., GEORGESCU, M.I., GORGAN, D.L., SARARU, C., JURCUT, A.D., POPESCU, D. Intelligent data analysis of plant growth and development in response to

- abiotic stress factors (Analiza inteligentă a datelor pentru creșterea și dezvoltarea plantelor în răspuns la factori de stres abiotici). *Applied Sciences*, vol. 10, no. 3, 2020, pp. 1-16.
118. TUGRUL B, ELFATIMI E, ERYIGIT R. Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. *Agriculture*. 2022; 12(8):1192. <https://doi.org/10.3390/agriculture12081192>.
 119. LI, J., YANG, W., LIU, W., FANG, S., CHEN, J., & LIAO, H. (2020). Application of Machine Learning in Plant Phenotyping. *Plant Phenomics*, 2020, 1-9. doi: 10.34133/2020/8853547.
 120. SINGH, A., CHAKRABORTY, D., & KARMAKAR, G. (2021). Plant Phenotyping using Multi-View Stereo Vision. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-10. doi: 10.1109/TIM.2021.3101754.
 121. NGUYEN, T., HAN, L., KIM, Y., & KIM, Y. (2021). Deep Plant Phenomics: A Machine Vision Platform for Analyzing Plant Growth and Development. *Sensors*, 21, 1-18. doi: 10.3390/s21165815.
 122. YANG, W., LI, H., LI, S., & WANG, D. (2021). Plant Phenomics: Current Status and Future Directions. *Frontiers in Plant Science*, 12, 1-17. doi: 10.3389/fpls.2021.702632
 123. LI, Y., YAO, Y., GUO, Y., QIAN, B., YAN, J., & CAO, W. (2020). Deep-Learning-Based Prediction of Soil Nutrients in Precision Agriculture. *IEEE Transactions on Geoscience and Remote Sensing*, 58, 4767-4778. doi: 10.1109/TGRS.2020.2968718.
 124. JIA, F., SHAO, L., LI, Y., WANG, Y., & WANG, H. (2020). Deep Learning for Crop Yield Prediction in Precision Agriculture: A Review. *Sensors*, 20, 1-21. doi: 10.3390/s20154230.
 125. PATEL, M., CHANDEL, T., BAJAJ, D., & KALIA, S. (2021). Remote Sensing and Machine Learning Approaches for Monitoring Crop Growth and Yield in Precision Agriculture: A Review. *International Journal of Intelligent Systems and Applications in Engineering*, 9, 71-86. doi: 10.5815/ijisae.2021.03.06.
 126. GUO, Z., GUO, F., LU, S., GUO, X., & ZHANG, X. (2020). Plant Leaf Recognition Based on Deep Convolutional Neural Networks. *IEEE Access*, 8, 40608-40620. doi: 10.1109/ACCESS.2020.2979426.
 127. SHRESTHA, S., ARYAL, J., & RIJAL, K. (2020). Crop Disease Diagnosis Using a Novel Convolutional Neural Network. *IEEE Access*, 8, 159191-159205. doi: 10.1109/ACCESS.2020.3027511.
 128. XU, M., ZHANG, B., LI, L., & MA, Q. (2021). An Automated Approach to Detecting Banana Diseases Using Deep Convolutional Neural Networks. *Applied Sciences*, 11, 1-13. doi: 10.3390/app11020858.

129. WEI, W., CHENG, Y., ZHU, X., & LI, D. (2020). Deep Learning-Based Crop Recognition for Intelligent Agriculture. *Sensors*, 20, 1-19. doi: 10.3390/s20133419.
130. ROY, A., SAHA, M., GHOSH, K., & GIRI, B. (2020). Plant Image Analysis for Plant Diseases and Nutrient Deficiency Classification. In *Proceedings of the 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 102-107). doi: 10.1109/iCoMET49435.2020.9133266.
131. AHMED, K. E. A., HOSSAIN, M. S., UDDIN, M. S., & ALAMRI, A. (2020). An Efficient and Comprehensive IoT-Based System for Smart Agriculture. *Sensors*, 20, 1-16. doi: 10.3390/s20174834.
132. AKTER, S., ISLAM, M. Z., UDDIN, M. Z., & HOSSAIN, M. M. (2020). Automated Diagnosis of Tomato Leaf Diseases Using Machine Learning Techniques. In *Proceedings of the 2020 International Conference on Networking, Systems and Security (NSysS)* (pp. 1-6). doi: 10.1109/NSYSS50361.2020.9363356.
133. AMARA, K., BOUGUESSA, M., & MOUSSAOUI, S. (2020). A Survey on the Applications of Machine Learning in Agriculture: A Review. *Journal of Information Science and Engineering*, 36, 1041-1066. doi: 10.6688/JISE.202007_36(4).0011.
134. NAIK, V. R., GHATOL, A. A., & VYAS, O. P. (2020). Plant Disease Recognition Using Machine Learning Techniques: A Review. In *Proceedings of the 2020 International Conference on Computing, Communication, and Networking Technologies (ICCCNT)* (pp. 1-5). doi: 10.1109/ICCCNT49239.2020.
135. LOPEZ, L. J., RONCANCIO, C. A., & OSPINA, J. G. (2019). Intelligent Detection of Plant Pests and Diseases: A Review. In *Proceedings of the 2019 IEEE Colombian Conference on Communications and Computing (COLCOM)* (pp. 1-6). doi: 10.1109/COLCOM45640.2019.9016628.
136. KUMAR, S., KAUR, H., & SINGH, R. (2019). Machine Learning Techniques for the Prediction of Plant Diseases: A Review. In *Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 501-505). doi: 10.1109/SPIN.2019.8711667.
137. LEE, S. H., LEE, J. Y., LEE, J. H., KIM, M. Y., & LEE, D. K. (2019). Agricultural IoT System for Real-Time Monitoring and Prediction of Environmental Parameters Affecting Crop Production. *IEEE Access*, 7, 113758-113767. doi: 10.1109/ACCESS.2019.2934276.
138. KAMLESH GOLHANI, SIVA K. BALASUNDRAM, GANESAN VADAMALAI, BISWAJEET PRADHAN, A review of neural networks in plant disease detection using

- hyperspectral data, *Information Processing in Agriculture*, Volume 5, Issue 3, 2018, Pages 354-371, <https://doi.org/10.1016/j.inpa.2018.05.002>.
139. YANG M. Physiological Disorder Diagnosis of Plant Leaves Based on Full-Spectrum Hyperspectral Images with Convolutional Neural Network. *Horticulturae*. 2022; 8(9):854. <https://doi.org/10.3390/horticulturae8090854>.
 140. AKHTAR, AZEEM & SHABBIR, QAMAR. Assessment of Stress on Plants Through Neural Networks. 4. 120-128, 2022.
 141. CHEN,S., CHEN,Y., YANG, X., ZHANG, J. Agricultural yield prediction using machine learning: A review. *Computers and Electronics in Agriculture*, vol. 163, 2019, p. 104859. [citat: 06.06.2023] Disponibil: <https://www.sciencedirect.com/science/article/pii/S0168169918306782>.
 142. SMITH, J. D., & JOHNSON, A. B. Effects of drought stress on the physiology and growth of tomato plants. *Plant, Cell & Environment*, 2020, 43(3), 657-671. <https://doi.org/10.1111/pce.13874>.
 143. MAJUMDAR, J., NARASEEYAPPA, S. & ANKALAKI, S. Analysis of agriculture data using data mining techniques: application of big data. *J Big Data* 4, 20 (2017). <https://doi.org/10.1186/s40537-017-0077-4>.
<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0077-4>.
 144. KAO, P.-H., BAIYA, S., LAI, Z.-Y., HUANG, C.-M., JHAN, L.-H., LIN, C.-J., LAI, Y.-S., & KAO, C.-F. (2022). An advanced systems biology framework of feature engineering for cold tolerance genes discovery from integrated omics and non-omics data in soybean. *Frontiers in Plant Science*, 13. <https://www.frontiersin.org/articles/10.3389/fpls.2022.1019709>. (DOI:10.3389/fpls.2022.1019709) ISSN 1664-462X. ISSN=1664-462X. <https://www.frontiersin.org/articles/10.3389/fpls.2022.1019709/full>.
 145. THE PLANT LIST. [citat: 06.03.2023] Disponibil: <http://www.theplantlist.org/>.
 146. INVASIVE SPECIES SPECIALIST GROUP ISSG (2011). Global Invasive Species Database. Checklist dataset <https://doi.org/10.15468/aaobov> accessed via GBIF.org on 2022-05-18.
 147. PAGAD S, GENOVESI P, CARNEVALI L, SCALERA R, CLOUT M (2015) IUCN SSC Invasive Species Specialist Group: invasive alien species information management supporting practitioners, policy makers and decision takers. *Management of Biological Invasions*.
 148. CABI. Digital Library. Invasive Species Compendium. [citat: 09.05.2023] Disponibil: <http://www.cabi.org/isc/>

149. CABI. IBIS Island Biodiversity and Invasive Species. [citat: 18.03.2022]. Disponibil: <https://ckan.cabi.org/data/dataset/ibis-island-biodiversity-and-invasive-species>.
150. GRIIS. Global Register of Introduced and Invasive Species. [citat: 01.04.2022]. Disponibil: <https://griis.org/>.
151. ISGS. GLOBAL INVASIVE SPECIES DATABASE. [citat: 18.03.2022]. Disponibil: <http://www.iucngisd.org/gisd/>.
152. INTER-AMERICAN BIODIVERSITY INFORMATION NETWORK (IABIN) Data Integration & Analysis Gateway <https://consbio.org/products/projects/data-integration-and-analysis-gateway>.
153. OAS. Organization of American States [citat: 18.03.2022]. Disponibil: <https://www.oas.org/en/sedi/dsd/iabin/default.asp>.
154. EBI. Explore all our data resources and tools. [citat: 18.03.2022]. Disponibil: <https://www.ebi.ac.uk/services>.
155. UNIVERSITY of NEBRASKA–LINCOLN. Institute of Agriculture and Natural Resources. SOYSIM - SOYBEAN GROWTH SIMULATION MODEL, <https://soysim.unl.edu/>
156. SETIYONO, Tri & CASSMAN, Kenneth & SPECHT, James & DOBERMANN, A. & WEISS, A. & YANG, H. & CONLEY, Shawn & ROBINSON, A.P. & PEDERSEN, P. & DEBRUIN, Jason. (2010). Simulation of soybean growth and yield in near-optimal growth conditions. *Field Crops Research*. 119. 161-174. 10.1016/j.fcr.2010.07.007 RANT D, NELSON RT, CANNON SB, SHOEMAKER RC. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res*. 2010 Jan;38(Database issue):D843-6. doi: 10.1093/nar/gkp798. Epub 2009 Dec 14. PMID: 20008513; PMCID: PMC2808871.
157. SOYBASE. Baza de date genetică și genomică USDA-ARS [citat: 06.06.2023] Disponibil: <http://soybase.org>.
158. DESHMUKH, Rupesh et al. (2020). Soybean Transporter Database (SoyTD): A comprehensive database for identification and exploration of natural variants in soybean transporter genes. *Physiologia Plantarum*. 171. 10.1111/ppl.13287.
159. SOYKB. SOYBEAN KNOWLEDGE BASE. SoyDT, [citat: 18.03.2022]. Disponibil: <https://soykb.org/SoyDT>.
160. HA, J., JEON, H.H., WOO, D.U. et al. Soybean-VCF2Genomes: a database to identify the closest accession in soybean germplasm collection. *BMC Bioinformatics* 20, 384 (2019). <https://doi.org/10.1186/s12859-019-2859-5>.

161. VREJOIU, M. (2019). Rețele neuronale convoluționale, Big Data și Deep Learning în analiza automată de imagini. *Revista Română de Informatică și Automatică*. 29. 91-114. 10.33436/v29i1y201909.
162. ELEMENTS OF AI. Tehnicile rețelelor neuronale avansate. [citat: 19.03.2022]. Disponibil: <https://course.elementsofai.com/ro/5/3> 157
163. NGO, VUONG & LE-KHAC, NHIEN-AN & KECHADI, Tahar. Designing and Implementing Data Warehouse for Agricultural Big Data. (2019). https://www.researchgate.net/publication/333477491_Designing_and_Implementing_Data_Warehouse_for_Agricultural_Big_Data.
164. BOULENT J, FOUCHER S, THÉAU J AND ST-CHARLES P-L Convolutional Neural Networks for the Automatic Identification of Plant Diseases. *Front. Plant Sci.* 10:941. (2019) doi: 10.3389/fpls.2019.00941.
165. SLADOJEVIC S, ARSENOVIC M, ANDERLA A, CULIBRK D, STEFANOVIC D. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Comput Intell Neurosci.* 2016;2016:3289801. doi: 10.1155/2016/3289801. Epub 2016 Jun 22. PMID: 27418923; PMCID: PMC4934169.
166. PHENOSPEX. Smart Plant Analysis. https://phenospex.com/products/plant-phenotyping/planteye-f600-multispectral-3d-scanner-for-plants/?gad=1&gclid=Cj0KCQjwho-lBhC_ARIsAMpgMoekOqmHm1g-zNc9ETZ7YB3XX1AXirUzi3p2SGUSyiDZjbe03RrXnKAaAoOrEALw_wcB.
167. MIRANDE K. et all. A graph-based approach for simultaneous semantic and instance segmentation of plant 3D point clouds. *Frontiers in Plant Science*. Vol. 13. 2022. DOI=10.3389/fpls.2022.1012669. <https://www.frontiersin.org/articles/10.3389/fpls.2022.1012669>.
168. ZHANG X.-Meng, LIANG Li, LIU Lin TANG Ming-Jing. Graph Neural Networks and Their Current Applications in Bioinformatics. *Frontiers in Genetics*. Vol.12.2021. DOI=10.3389/fgene.2021.690049.
169. AMARA A, FRAINAY C, JOURDAN F, NAAKE T, NEUMANN S, NOVOA-DEL-TORO EM, SALEK RM, SALZER L, SCHARFENBERG S AND WITTING M (2022) Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation. *Front. Mol. Biosci.* 9:841373. doi: 10.3389/fmolb.2022.841373.
170. PITOLLAT, B., TANDO, N., ROUARD, M., DROC, G., VENKATESAN, Aravind et al.. AGROLD: A Knowledge Graph Database for plant functional genomics. *JOBIM 2021*, Jul 2021, Paris, France. 2018, ff10.1101/325423ff. fahal-03269813f.

171. OVENS K, EAMES BF AND MCQUILLAN I (2021) Comparative Analyses of Gene Co-expression Networks: Implementations and Applications in the Study of Evolution. *Front. Genet.* 12:695399. doi: 10.3389/fgene.2021.695399.
172. BHARAT, Mishra, KUMAR, Nilesh, M. MUKHTAR, Shahid. Network biology to uncover functional and structural properties of the plant immune system, *Current Opinion in Plant Biology*, Volume 62, 2021, 102057, ISSN 1369-5266, <https://doi.org/10.1016/j.pbi.2021.102057>. Disponiibl: <https://www.sciencedirect.com/science/article/pii/S1369526621000571>).
173. HE, Q., TANG, S., ZHI, H. et al. A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nat Genet* (2023). <https://doi.org/10.1038/s41588-023-01423-w>. <https://www.nature.com/articles/s41588-023-01423-w>.
174. NAIK, Hsiang Sing, LEE, Nigel. Graph Based Automated Analysis for Plant Root Phenotyping <https://dr.lib.iastate.edu/server/api/core/bitstreams/818a1b20-ef2f-4030-a338-42bc2a3582fa/content>.
175. ŞEN, Z. (2015). Chapter Seven - Drought Hazard Mitigation and Risk. In Z. Şen (Ed.), *Applied Drought Modeling, Prediction, and Mitigation* (pp. 393-459). Elsevier. ISBN 9780128021767. <https://doi.org/10.1016/B978-0-12-802176-7.00007-9>.
176. Drought analysis by the climate indexes with the yield of the main crop from the crişurilor plain, Romania. [citat: 20.04.2021], Disponibil: https://www.researchgate.net/publication/282362353_drought_analysis_by_the_climate_indexes_in_disponibil_with_the_yield_of_the_main_crop_from_the_crisurilor_plain_romania.
177. CHOI, Sun & Kim, YOUNG Jin & Briceno, MAVRIS, Simon & Dimitri. (2016). Prediction of weather-induced airline delays based on machine learning algorithms. 1-6. 10.1109/DASC.2016.7777956. Disponibil: https://www.researchgate.net/publication/311610165_Prediction_of_weather-induced_airline_delays_based_on_machine_learning_algorithms.
178. LINGUEE.[citat: 06.06.2023] Disponibil: <https://ro.linguee.com/rom%C3%A2n%C4%83englez%C4%83/traducere/reproductibilitate.html>.
179. ORACLE. [citat: 06.01.2023] Disponibil: <https://www.oracle.com/database/what-is-a-data-warehouse/>
180. ORACLE. [citat: 02.03.2023] Disponibil: <https://www.oracle.com/ro/cx/service/knowledge-management/what-is-know-ledge-management/>.

181. HASSAN, Md. M. Agricultural Statistics and Biostatistics in Agriculture –A Basic Introduction. In: Agrobusiness Education & Research International. Citat Disponibil: <https://agribusinessedu.com/agricultural-statistics-biostatistics-agriculture-basic-introduction/#:~:text=It%20decides%20crop%20quality%2C%20crop,yields%20and%20crop%20quality%20tests>.
182. MITRANA, G., MITRANA V. A formal language-based approach in biology. *Comp Funct Genomics*. 2004;5(1):91-4. doi: 10.1002/cfg.364. PMID: 18629037; PMCID: PMC2447320.
183. MARTIN-VIDE, C. MITRANA, V. PĂUN, G. Formal languages and applications. Springer, 2004.
184. CĂPĂȚÂNĂ Gh. A Programming Paradigm Oriented to Families of Problems. In.: The Scientific Bulletin Addendum No. 4/2019, The Official Catalogue of the „Cadet INOVA” Exhibition Research and Innovation in the Vision of Young Researchers The International Student Innovation and Scientific Research Exhibition - „Cadet INOVA’19”, „Nicolae Bălcescu” Land Forces Academy Sibiu, April 11-13, 2019, p. 72-83, ISSN 2501-3157, ISSN-L 2501-3157, http://cadetnova.ro/documente/Supliment_Inova_19.pdf
185. McDONALD, J.H. Handbook of Biological Statistics. University of Delaware: Sparky House Publishing, 2009. - 180 p.
186. ORGANIC AGRICULTURE INTRODUCTION. In: Agrobusiness Education & Research International. Citat Disponibil: <https://agribusinessedu.com/organic-agriculture-introduction/>.
187. BĂBEANU, N. ș.a. Biometrie. UASM, 2012, 180 p.
188. DODGE Y. The Concise Encyclopedia of Statistics. Springer: Berlin, Germany, 2008;
189. ДОСПЕХОВ Б. А. Методика полевого опыта (с основами статистической обработки результатов исследований). - 5-е изд.; Агропромиздат: Москва, Россия, 1985.
190. GATAULIN A., LICA D., ПОМОХАЦИ Cr. Biostatistică intuitivă. Ceres, 2002.
191. WOLFRAM MATHWORLD. Student's t-Distribution. <http://mathworld.wolfram.com/Studentst-Distribution.html>.
192. TATTREK. Student's t-Distribution. [citat: 06.12.2020] Disponibil: <https://stattrek.com/probability-distributions/t-distribution.aspx>.
193. KHAN ACADEMY. Sampling distribution of the sample mean. <https://www.khanacademy.org/math/ap-statistics/sampling-distribution-ap/sampling-distribution-mean/v/sampling-distribution-of-the-sample-mean>.

194. HAYES, A. Probability? How Do You Use It. In: Investopedia: <https://www.investopedia.com/terms/t/tdistribution.asp>.
195. STATISTICS HOW TO. T-Distribution / Student's T: Definition, Step by Step Articles, Video. [citat: 10.02.2020] Disponibil: <https://www.statisticshowto.com/probability-and-statistics/t-distribution/>.
196. SINGPURWALLA, D. A handbook of statistic: An overview of statistical methods, 1st edition; Bookboon:Copenhaga, Denmark, 2013.
197. BRIAN E. The Cambridge Dictionary of Statistics; Cambridge University Press: Cambridge, United Kingdom, 1998.
198. SAPKOTA, A. The Biology Notes.P Value- Definition, Formula, Table, Finding P-value, Significance.August 3, 2023 <https://thebiologynotes.com/p-value/>.
199. INVESTOPEDIA. P-Value: What It Is, How to Calculate It, and Why It Matters. [citat: 10.03.2020]. Disponibil: <https://www.investopedia.com/terms/p/p-value.asp>.
200. WOLFRAM. References. <https://reference.wolfram.com/language/guide/Statistics.html>.
201. MIAS G. Mathematica for Bioinformatics: A Wolfram Language Approach to Omics. Springer, 2018.
202. IGI GLOBAL. Database of Information Science & Technology. What is Graph Data Warehouse. <https://www.igi-global.com/dictionary/graph-database-management-systems/89787>.
203. INTELIA. Lu, K. Manage Data Warehouse Relationships With The Graph Database. <https://www.intelia.com.au/2022/11/30/manage-data-warehouse-relationships-with-the-graph-database/>.
204. ROBINSON, R; WEBBER J.; EIFREM, E. Graph Databases. O'Reilly, Sebastopol, California,USA 2015.
205. TOWARDS DATA SCIENCE. MORGANTE, V. What is a graph database? Disponibil: <https://towardsdatascience.com/what-is-a-graph-database-249cd7fdf24d>. (accesat 03. 12. 2022).
206. GRAF IN AGRO. Graphs Are Feeding the World. <https://neo4j.com/blog/graphs-feeding-world/>.
207. NEO4J GRAPH DATABASE. [citat: 08.04.2020]. Disponibil: <https://www.bbva.com/en/Neo4jwhatagraphdatabaseisandwhatitisusedfor/>.
208. GARTNER. Top 10 Data and Analytics Trends for 2021. [citat: 06.01.2022]. Disponibil: <https://www.gartner.com/smarterwithgartner/gartner-top-10-data-and-analytics-trends-for-2021>.
209. NEO4J. Graph Data Modeling. <https://neo4j.com/graphacademy/eduprogram/>.

210. BANCA DE DATE STATISTICE MOLDOVA. Disponiibl:
<https://statbank.statistica.md/PxWeb/pxweb/ro/?rxid=2f5a0079-c3c1-461e-81a4-d854b7ba0fcd>.
211. IONOVA E.V., LIKHOVIDOVA V.A., LOBUNSKAYA I.A. Drought and hydrothermal humidity factor as one of the criteria to estimate its intensity degree (literature review). Grain Economy of Russia. 2019;(6):18-22. <https://doi.org/10.31367/2079-8725-2019-66-6-18-22>. <https://www.zhros.online/jour/article/view/804>.
212. HANDBOOK OF DROUGHT INDICATORS AND INDICES. https://library.wmo.int/doc_num.php?explnum_id=3057.
213. SERVICIUL HIDEOMETEOROLOGIC DE STAT.
<http://www.meteo.md/index.php/clima/terminologie-si-unitati-de-masura/glosar-de-termini/c/> (accessed on 2023-03-24).
214. NEO4J. Disponibil: <https://neo4j.com/docs/graph-data-science/current/machine-learning/machine-learning/>.
215. MACHINE LEARNING. Evaluate Models Using Metrics. <https://developers.google.com/machine-learning/testing-debugging/metrics/metrics>.
216. GANEA, I. Model of statistical data analysis on nitrogen content in soybeans (Glicine max Merrill) in Clavera variety. Journal of Engineering Science 2023, 30 (1), 99. 165-177. [https://doi.org/10.52326/jes.utm.md.2023.30\(1\).14](https://doi.org/10.52326/jes.utm.md.2023.30(1).14).
217. HORTICULTORUL. Cultura hidroponică a legumelor. [citat: 11.07.2021]. Disponiibl:
<https://www.horticultorul.ro/legume/cultura-hidroponica-a-legumelor/>.
218. DEXONLINE. Electrolit. <https://dexonline.ro/definitie/electrolit>.
219. ROMEDIC. Purtan, T. Electroliții. <https://www.romedic.ro/electrolitii-0P36969>
220. SERVICIU HIDROMETEOROLOGIC DE STAT. [citat: 12.09.2022]. Disponibil:
<http://www.meteo.md/index.php/clima/cercetri-climatice/>.
221. SHARMA, S. What the Hell is Perceptron? The Fundamentals of Neural Networks. <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>

CURRICULUM VITAE

Informații personale

Ion Ganea



30-11-1974



0690 91 730



iganea9@gmail.com, ganea.ion@usm.md



Chișinău, Republica Moldova.

Experiența profesională

2014 - prezent	Centrul de Excelență în Economie și Finanțe. Profesor de Tehnologii Informaționale.
2011-2014	Colegiul de Construcții din Chișinău. Inginer-programator.
2002-2006	Gimnaziul Făgureni, Strășeni. Profesor de istorie
2002-2004	Școala Medie Căpriană, Strășeni. Profesor de limba franceză.
1997-1999	Liceul Mihail Sadoveanu, Chișinău. Profesor de limba franceză.

Educație și formare

2019 - 2023	Universitatea de Stat din Moldova Școala Doctorală de Matematică și Știința Informației Doctorand.
2016-2018	Universitatea de Stat din Moldova Masterat „Baze de date și cunoștințe”. Master în științe exacte.
2007-2009	Institutul de Formare Continuă. Tehnologii Informaționale. Informatician.
1996-2002	Universitatea de Stat din Moldova. Istorie. Licențiat în istorie.
Domenii de interes științific	Inteligența Artificială, Machine Learning, Data Mining, Text Mining, Baze de date graf, Fenomica Plantelor.

Competențe lingvistice

Limba maternă	• Română
Limbi străine	• Rusă B1 • Franceză A2

Publicații

1. Ganea, I. Influența unor substanțe biologice active asupra creșterii și dezvoltării plantelor de soia cultivate pe medii nutritive apoase KNOP. In: StudMath-IT 2020 Students Conference organized by the Faculty of Exact science within "Aurel Vlaicu" University of Arad, 13th edition, p. 12-13
2. Ganea, I. Elaborarea bazei de date de tip graf pentru analiza datelor biologice. Volumul lucrărilor: International Students' Conference "StudMath-IT 2021", "Aurel Vlaicu" University of Arad, Faculty of Exact Science Arad, Romania, November 18-19, 2021
3. Ganea, I. Sistem suport pentru studiul dezvoltării plantelor cultivate pe medii M1. Conferința științifică națională a doctoranzilor „Metodologii contemporane de cercetare și evaluare” 22-23 aprilie 2021. Chișinău, 2022, CEP USM. CZU:004.89:[57.087.1:631.53], pag. 103 - 107.
4. Ganea, I. Use of bio-statistics in data analysis of the biological active substances on plants. Mathematics & IT: Research and education (MITRE – 2021), International Conference. Abstracts. Chișinău, 2021. CEP USM, 2021. ISBN 978-9975-158-19-0. Pag. 104.
5. Ganea, I. Bîrsan, A. Biological data processing. International Exhibition InventCor - IIInd edition. Posters. ISBN 978-606-35-0467-9, pag. 44. Medalie de aur.
6. Ganea, I. Bîrsan, A. Căpățână, G. Using the neo4j pearson similarity algorithm for the analysis of protein content in soybean. The Official Catalogue of the “Cadet INOVA” Exhibition Research and Innovation in the Vision of Young Researchers. “Nicolae Bălcescu” Land Forces Academy of Sibiu. Scientific Bulletin Addendum No. 7/2022 April 7-9, 2022. Academia Forțelor Terestre „Nicolae Bălcescu” din Sibiu. ISSN 2501-3157. Pag. 120 – 131.
7. Ganea, I. Bîrsan, A. Use of neo4j algorithms in biological data processing. Proceedings of the 14th edition of Euroinvent European Exhibition of Creativity and Innovation, Iași, 2022. ISSN Print: 2601-4564, Online: 2601-4572. Pag. 167 – 168. Medalie de argint.
8. Ganea, I. Bîrsan, A. Use of neo4j link prediction algorithms in biological data processing. În „Salonul Internațional de Invenții, Inovații „Traian Vuia””. Timișoara, 08-10 octombrie 2022 / coord. : Romi Rădulescu. Timișoara, Editura Politehnica, 2022, p. 77, ISBN 978-606-35-0496-9. Medalie de bronz.
9. Ganea, I. Bîrsan, A. Use of neo4j algorithms in biological data processing. Proceedings of the 14th edition of Euroinvent European Exhibition of Creativity and Innovation, Iași, 2022. ISSN Print: 2601-4564, Online: 2601-4572. Pag. 167 – 168. Medalie de argint.
10. Ganea, I. Bîrsan, A. Use of embedding algorithms in biological data processing. Proceedings of the

15th edition of Euroinvent European Exhibition of Creativity and Innovation, Iași, 2023. ISSN Print: 2601-4564, Online: 2601-4572. Pag. 167 – 168. Diplomă de mențiune.

11. Ganea, I. Bîrsan, A. Căpățână, G. Model de analiză a scurgerii electroliților la plantele de soia. Prezentat la Conferința științifică națională cu participare internațională „Integrare prin Cercetare și Inovare”, 10.11.2022.
12. Bîrsan, A. Sprincean, V. Ganea, I. Oportunități de utilizare în practică a tehnologiilor inteligente în evaluarea unor variabile biologice. În: Studia Universitatis Moldaviae, Nr. 1(2023), pp. 53-64. DOI: [https://doi.org/10.59295/sum1\(171\)2023_07](https://doi.org/10.59295/sum1(171)2023_07).
13. Ganea, I. Model of statistical data analysis on nitrogen content in soybeans (Glicine max Merrill) in Clavera variety. Journal of Engineering Science 2023, 30 (1), 99. 165-177. [https://doi.org/10.52326/jes.utm.md.2023.30\(1\).14](https://doi.org/10.52326/jes.utm.md.2023.30(1).14).
14. Ganea, I., Căpățână, Gh. A predictive model for determination probabilities of drought (în tipar).
15. Ganea, I. Prediction Model for Soybean Productivity. Database Systems Journal (în tipar).

- | | |
|-------------------|---|
| Drepturi de autor | <ol style="list-style-type: none"> 1. Ganea, I. Bîrsan, A. Smântâna, L. Căpățână, G. Model de analiză statistică a datelor privind conținutul de azot și fosfor în plante. Drept de autor. Seria O Nr. 7091 din 08.12.2021. 2. Ganea, I. Model de predicție a productivității unor culturi agricole. Drept de autor. Seria OȘ Nr. 7552 din 21.06.2023 |
|-------------------|---|

Distincții	La Saloane de Invenții și Tehnologii din România, performanțele științifice au fost înalt evaluate cu 2 Medalii de Aur; 1 Medalie de Argint; 1 Medalie de Bronz; Diplomă de Excelență.
------------	--

Certificări	Certificare profesională <i>Neo4j</i> .
-------------	---