**MOLDOVA STATE UNIVERSITY**

**DOCTORAL SCHOOL OF NATURAL SCIENCES**

With manuscript title

CZU: 004.42:004.6:581.1(043)

**ION GANEA**

# MODELS, TECHNIQUES AND PRODUCTS INTELLIGENT DATA ANALYSIS PROGRAM IN PLANT PHYSIOLOGY

## ABSTRACT

121.03. Computer programming

**Author:** Ion Ganea

**Scientific advisor:** Gheorghe Căpățână, PhD in technical science, professor.

**Guidance committee:**

Ana Bîrsan, PhD in biological science, associated professor.

Valeriu Ungureanu, PhD in physical-mathematical science, associated professor.

Maria Beldiga, PhD in computer science, associated professor.

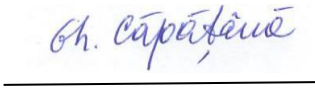
**CHIȘINĂU, 2024**

The thesis was developed in the Doctoral School of Natural Sciences, Moldova State University.

**Author**

**Ion GANEA**

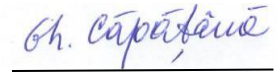
**Scientific advisor:**



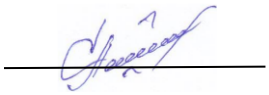
**Gheorghe CĂPĂȚÂNĂ**, PhD in technical sciences, professor

**Doctoral committee:**

**Vsevolod ARNAUT**, PhD in physical-mathematical sciences, associate professor, Moldova State University.



**Gheorghe CĂPĂȚÂNĂ**, PhD in technical science, professor, Moldova State University.



**Ana BÎRSAN**, PhD in biology, associate professor, Moldova State University.

**Emil GUȚULEAC**, PhD in technical sciences, professor, Technical University of Moldova.

**Igor COJOCARU**, PhD in computer sciences, Information Society Development Institute.

The defense will take place on 21.03.2024, 14.00, MSU, hall 222/4.

The abstract and doctoral thesis can be consulted at the Moldova State University Library and on the website of ANACEC https://cnaa.md.

The abstract was sent on 06.02.2024.

Author:  **Ion GANEA**

## CONTENT

## CONCEPTUAL MARKINGS OF THE RESEARCH

### Timeliness and importance of the theme

Currently, humanity is faced with complex global problems that threaten its development and existence. These problems require new concepts for sustainable economic development, assessing environmental impact. Climate change requires adaptive agriculture, supported by transdisciplinary research and smart technologies, so that nature protection coexists with development. Technologies represent essential methods for obtaining products and fulfilling human requirements [1, 2]. Intelligent data analysis in plant physiology is a growing research area that represents a combination of biology and computer science. This subject combines plant science research with modern approaches to data analysis and processing, using artificial intelligence (AI) and machine learning (ML) techniques and algorithms to extract relevant information about plant homeostasis (the physiological process by which organisms maintain a stable internal environment (temperature, pH and nutrient levels) despite external changes, their growth and development and their response to adverse factors.

**The purpose** of the PhD thesis is to develop and provide assistance to biological researchers to facilitate the processing of information and the identification of significant results obtained within complex biological research to optimize the process of discovering relevant knowledge and acquiring wisdom in biology and agriculture.

proposed research **objectives are the following:**

1. Development of a solid Knowledge Base in the field of „Integrated Data Modeling in Plant Physiology" (MIDDFP) for the development and integration of models and application software products.
2. Elaboration of complex Formal Language for a deeper understanding of plant physiological processes.
3. Developing Family-oriented Intelligent Data Analysis Models of problems in plant physiology for evaluation of the efficiency of the use of biologically active compounds on plants in an agricultural context.
4. Development of a Non-Relational Database (Neo4j) for efficient data storage in biological and agricultural research and the identification of essential information in the context of biological research to facilitate the process of selection and interpretation of relevant data.

5. Development of Intelligent Models for Prediction of Meteorological and Biological Phenomena for the predictability of sustainable and resilient agricultural development.
6. Integrating Products into a Problem Family-Oriented Data Repository for Data Modeling in Plant Physiology and to facilitate informed decision-making in the assessment of complex biological interactions.
7. Implementation of the product development paradigm on problem families, by exploring their potential solutions for expanding the results in various research fields, using Transfer Learning Techniques.

**Research hypothesis**

The integrated use of biostatistical analysis, graph databases and artificial intelligence algorithms in plant physiology investigations will lead to significant discoveries and the development of innovative tools for evaluating and interpreting complex data in the field under study. These tools will consist of advanced biostatistical analysis models and an efficient data management system, thus providing solid resources for further research and deeper understanding of plant physiological mechanisms.

**Research methodology**

Advanced methods and techniques were used in the thesis for research in the field of plant physiology and agriculture, by combining biostatistics and graph data repositories. This interdisciplinary approach has enabled valuable insight, information and relevant knowledge to be gained in the field.

Through the use of biostatistics, a series of rigorous statistical analyzes and tests were applied to investigate the relationships between various variables in the field of plant physiology. These methods helped to identify patterns, correlations and significant differences in the collected data, giving us a deep understanding of the studied phenomena.

A graph-based data warehouse was developed using the Neo4j DBMS. In the context of plant physiology research, this methodology has allowed for a clearer perspective on the mechanisms that direct plant growth and development, as well as on the interactions between plants and environmental factors in agriculture. By combining biostatistics and graph data repositories, we succeeded in highlighting the importance and potential of this holistic approach in developing knowledge in plant

physiology and agriculture, opening new research directions and applying our findings to improve agricultural practices and environmental sustainability.

**Scientific novelty and originality**: A formal language based on problem families, a knowledge base, biostatistical data analysis models, a data warehouse, predictive models related to drought and crop productivity have been developed for the MIDDFP research area. The data warehouse also enables the application of the intelligent data analysis model and presents the results through intuitive graphs that are easy to use and understand. Through an innovative solution based on a graph data warehouse, the performance of biological and agricultural systems is optimized through advanced technologies and biostatistical analyses. The aim is to find effective solutions to agricultural challenges such as climate change, soil degradation and sustainable production. Biostatistical analysis helps identify correlations between plant variables, facilitating the selection of resilient and environmentally adapted plants.

**Theoretical significance**: Advancing knowledge and understanding in the field of intelligent technologies applied in biology and agriculture. Research in the field of intelligent technologies applied in biology and agriculture consists in the discovery of new knowledge and a deeper understanding of the complex interactions between biological factors and the environment. Through research, the models and mechanisms of action of intelligent technologies are identified and validated, in the context of biology and agriculture.

**The applied value of research:** Development of concrete and innovative solutions for optimizing agricultural production, adapting to climate change and increasing the sustainability of agricultural systems. Predictive models are crucial in managing climate risks and adapting agriculture to climate change in agriculture and improving production planning and resource allocation. Advanced information and technological solutions increase productivity, reduce environmental impact and ensure long-term food security. Concrete and innovative solutions developed in research can be effectively implemented in agricultural practices, generating economic and social benefits for farmers and communities.

**Scientific results submitted for support:**

1. „Integrated modelling of data in the field of plant physiology" (abbreviated MIDDFP) was studied.
2. MIDDFP activity area, oriented to problem families, was developed. The MIDDFP formal language is of significant complexity. Following conventional programming technologies, the

language can be realized within a programming automation system developed by a collective of computer scientists.

3. The use of artificial intelligence tools, non-relational databases and problem solving methodology oriented to problem families allowed the creation of the language created by the author within the deadline set in the Doctoral Research Plan.
4. *Generic (abstract) models and specific models oriented to problem families* were developed;
5. 9 biostatistical analysis models of the influence of some biologically active substances on the growth and development of soybean phenotypes (*Glycine max* (L.) Merrill.) were developed;
6. A graph data warehouse has been developed for storing and processing large amounts of data.
7. Two predictive models were developed regarding the probability of drought and the productivity of agricultural crops.

**Approval of thesis results:**

The results obtained from the research were presented and discussed at 11 national and international conferences:

1. European Exhibition of Creativity and Innovation „EuroInvent", 15th edition. 16 -18 May 2023, Iasi, Romania. **Diploma of excellence.**
2. International Exhibition „Inventcor" 2022, December 15-17, 2022, Deva, Romania. **Gold medal.**
3. International Exhibition „Inventcor" 2022, December 16-18, 2021, Deva, Romania. **Gold medal.**
4. International Exhibition of Inventions and Innovations „Traian Vuia" 2022, December 8-10, 2022, Timișoara, Romania. **Bronze medal.**
5. European Exhibition of Creativity and Innovation „EuroInvent", 14th edition. May 26-28, 2022, Iasi, Romania. **Silver medal.**
6. The national conference with international participation „Integration through Research and Innovation", 10 – 11.11.2022, Moldova State University, Chisinau, Republic of Moldova.
7. The International Student Innovation and Scientific Research Exhibition Cadet INOVA'22 Nicolae Bălcescu Academy of Land Forces, April 7 - 9, 2022, Sibiu, Romania.
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, State University of Moldova, Chisinau, Republic of Moldova.
10. The national scientific conference of doctoral students dedicated to the 75th anniversary of USM „Contemporary research and evaluation methodologies”, April 22-23, 2021, State University of Moldova, Chisinau, Republic of Moldova.
11. International Students' Conference StudMath-IT 2020 „Aurel Vlaicu” University of Arad, Faculty of Exact Science Arad, Romania, November 26-27, 2020.

**Implementation of scientific results:**

The developed models were awarded with:

- Two gold medals at the Deva International Exhibition of Innovation and Scientific Research in 2021 and 2022.
- Silver medal and Diploma of Excellence at the „EuroInvent” International Fair, Iași 2022, 2023.
- Bronze medal at the International Exhibition of Inventions and Innovations „Traian Vuia” 2022.
- At AGEPI, copyrights were registered on the works „Statistical analysis model of data on nitrogen and phosphorus content in soybeans” (Series 0, No. 7091 of 08.12.2021) (appendix 8) and „Prediction model of productivity of some agricultural crops” (Series 0Ș, No. 7552 of 21. 06. 2023).

## THESIS CONTENT

The thesis is composed of an introduction, four chapters, conclusions and final recommendations, a bibliography, 32 appendices. The content of the thesis includes 135 pages, 37 figures and 17 tables.

In **the Introduction**, the topicality and importance of the theme is presented. The purpose of the doctoral thesis, research objectives, research methodologies, scientific results obtained and their applied value were identified.

**Chapter I „Intelligent data processing in biology and agriculture"** reflects the current state of holistic knowledge and research in the field of implementing smart technologies in biology and agriculture. The fundamental concepts in the field of intelligent technologies were analyzed such as the notions of data, information, knowledge, knowledge, insight, wisdom. The concepts of Artificial Intelligence and its components *Machine Learning* (ML), *Deep Learning* (DL), *Transfer Learning* (TL). The relevant studies and experiments that were carried out to explore the potential of these technologies in improving the performance and sustainability of biological and agricultural systems were analyzed. Also, the advantages and benefits brought by smart technologies in different aspects of biology and agriculture were determined, such as optimization of production processes, efficient monitoring and management of resources, as well as anticipation and prevention of specific problems.

Global factors such as climate change, demographic changes and food security issues are driving researchers and agricultural producers to seek new innovative approaches to developing agricultural crops and increasing their yield. The latest advances in agricultural technology and practices are described in the specialized literature by the terms Agriculture *4.0* (Precision Agriculture) and *Agriculture 5.0* (Regenerative *Agriculture* or *Intelligent Agriculture)*, this develops precision agriculture by introducing robotics and *Artificial Intelligence* (IA) [3]. The ultimate goal is to create sustainable and regenerative systems that benefit both farmers and the environment. This progressive approach involves the use of *Big Data*, *Artificial Intelligence* and other disciplines of information science with Intelligent Technologies, such as artificial intelligence, machine learning and autonomous robots, bring with them a number of advantages, including: *increased efficiency, data-driven decision-making, automation and autonomy* [4].

Recent studies examine the challenges and limitations associated with using *Machine Learning* (ML) and *Deep Learning* (DL) in plant disease identification, including issues related to data availability, image quality, and distinguishing between healthy and diseased plants [4]. They provide

valuable insights to researchers, practitioners and industry professionals on plant disease detection, presenting solutions to these challenges and limitations, and providing a comprehensive understanding of the current state of research in this field [5, 6]. It also highlights the benefits and limitations of ML and DL methods and proposes potential solutions to overcome the challenges of their implementation [7, 8].

**Intelligent Data Processing** (IDP) studies the complex interactions within biological systems, using a holistic approach to **predict** how these systems evolve and adapt to various economic [9, 10], social and environmental [11-12] conditions for development of sustainable solutions to global problems such as food, health and the environment [13-15]. In order to solve the problems in the research field and to understand things in depth, it is necessary to study the components of a biological system globally at the molecular and informational level [16, 17]. These disciplines bear the generic name of *omics*.

*Omics research* and *holistic research* complement each other, as omics provides large-scale data and information about biological components, while holistic research integrates these data to gain a more complete understanding of biological systems [18-19].

In order to develop predictive models or high-precision analyzes oriented to families of problems in the field of plant physiology based on the available data, *Transfer Learning techniques can also be applied* [20]. In this context, TL can help to adapt pre-learned models or knowledge in a specific field or plant species to improve the understanding or prediction of phenomena related to plant growth, development and condition [21, 22]: prediction of health *status of plants*, *crop yield prediction*, *optimizing the use of water and nutrient resources*, *analyzing transcriptomic and proteomic data*, *monitoring climate change and adapting to new environments or environmental variations* [23-25].

In recent years, a lot of research has been carried out in the field of intelligent data processing in the field of plant physiology. This research focuses on the use of machine learning and deep learning technologies in agriculture and plant phenotyping. Various aspects related to monitoring, diagnosis and analysis of plants and agricultural crops have been analyzed using data processing techniques and machine learning models [26, 27]. The papers cover various aspects of plant phenotyping, such as identification of various characteristic traits, image extraction and classification, use of multi-visual stereo vision to obtain three-dimensional plant representations, machine vision analysis of plant growth and development based on deep learning, and yield prediction agricultural crops [26-32].

The review of existing literature demonstrates that intelligent data processing and data warehouses are important technologies in the field of plant physiology. They can be successfully used to collect, analyze and interpret data in real-time, which can help improve crop yields, reduce costs and environmental impact, and optimize resource use.

**Chapter II „Integration of biostatistics methods and data warehouses in plant physiology data processing"** contains the description of the research methods and resources used, as well as the methodological approaches applied in the research. This chapter describes the formal language of the generic and specific models created for data processing. Knowledge base development is an essential step to obtain relevant and consistent information for biostatistical analyzes of plant physiology data and graph data repositories.

In order to achieve the set objectives and make the necessary decisions, an intelligent information processing system based on *problem families* (FP) was developed [34]. Problem family modelling and programming is a new paradigm for developing software products. This paradigm has the advantage of standardizing design architecture and technology, operating conditions, and increasing the productivity of information systems development.

***Integrated Data Modeling of Plant Physiology (MIDDFP)*** research field, reflecting the complexity of biological processes in plant functioning. The language standardizes terms and concepts in modeling related data in the field of plant physiology, facilitating communication and collaboration between researchers. It also includes mathematical rules for representing processes, allowing simulations and prediction of plant behavior under various conditions.

Modelling plant physiology data using formal language has significant importance and relevance in understanding the complex processes that govern plant development, growth and responses to various stimuli. These aspects are essential for researchers, biologists, agronomists and others involved in the study of plants and the environment. Data modelling in this area has a number of advantages, including:

- *Accuracy in Understanding Physiological Processes*: Formal data models allow researchers to represent complex plant physiological processes in a logical and structured manner, providing opportunities for a deeper understanding of the interactions between different plant components and processes.
- *Predictions and simulations*: By modelling data, virtual simulations of plants and their growing environment can be created, which allows testing different hypotheses and scenarios

without conducting expensive experiments in reality and predicting the reaction of different genotypes to the action of different endogenous and exogenous factors, such as climate change or the use of different types of nutrients.

- *Data modelling* helps to understand how plants respond to different stresses, such as drought, disease or pest infestations. More effective strategies can be developed to combat these problems and protect crops.

The developed language can serve as the basis for software specialized in simulating plant physiology processes, helping to test hypotheses and predict plant responses in diverse scenarios, with implications in agriculture and related fields. As research continues, the formal language may evolve, reflecting new discoveries. Through interdisciplinary collaboration, this language can deepen understanding of plant-environment interactions.

The Problem Oriented Family Application (OFP) is *a generic* (*abstract*) application in relation to the applications $A_1$, $A_2$, ..., $A_n$, and the applications $A_1$, $A_2$, ..., $A_n$ represent ***n*** *specific applications* in relation to *the generic application A*.

In the case, when the set of problems $P_1$, $P_2$, ..., $P_n$ can be regarded as a family of problems, a single application $A$, which realizes the family of problems on the computer, can be developed.
The research area „*Integrated Data Modeling in Plant Physiology"* (abbreviated *MIDDFP*) integrates a range of entity types.

The formal *MIDDFP language*, which describes the solutions to some problems in these research fields, requested by scientific researchers specializing in plant biology and by farmers, is as follows:

<MIDDFP *language*> ::= <*studied entity*> | <*entity type relationship*> | <*define nodes and relationship between nodes*> | <*relationship between nodes*> | <MIDDFP*issue family*>

<*studied entity*> ::= <*properties studied objects*> | <*soil types*> | <*farm*> | <*geographic area*> | <*water resource*> | <*soil moisture*> | <*amount of precipitation*> | <*air temperature*> | <*hydrothermal coefficient*> | <*vitesis of the wind*> | <*weather forecast*> | <*manufacturer*> | <*pesticides and fertilizers*> | <*researchers and the research institution*> | <*agricultural equipment*> | <*sale and transaction*>

<*relationship between entity types*> ::= <*relationship between crops and farms*> | <*relationship between crops and soil types*> | <*relationship between farms and crops*> | <*relationship between farms and water resources*> | <*relationship between farms and geographical*

*areas> | <relationship between farms and water resources> | <relationship between geographical areas and water resources> | <relationship between geographical areas and amount of precipitation> | <relationship between geographic areas and air temperature> | <relationship between geographic areas and wind speed> | <relationship between crops and pesticides/fertilizers> | <relationship between varieties and crop quality> | <relationship between farms and producers of plant products> | <relationship between cultures and researchers/research institutions> | <relationship between farms and farm equipment> | <relationship between farms and sales/transactions> | <relationship between crops and weather forecasts>*

To achieve the proposed objectives, a knowledge infrastructure was developed to be used in a data warehouse. This process involves the integration of biostatistical methods into a coherent theoretical and practical framework, tailored specifically to the needs and characteristics of the field of plant physiology. The data warehouse will be used to store and share experimental information in the field of plant physiology.

**Graph data stores:**

- *Graphing data*: Using graphs to graph data can make it easier to quickly understand patterns and trends. Examples include histograms, scatter plots, bar charts, etc.
- *Networks and interaction graphs*: These can be used to represent the interactions between different elements of the plant system, such as pollination or competition relationships between species.
- The use of graph data repositories in the field of plant physiology is a relatively new research direction, but one that is already showing promising results.

**Combining biostatistical methods with graph data repositories** in intelligent data processing in plant physiology brings numerous benefits in understanding and optimizing plant growth. This integrated approach can be used to analyze data obtained from different experiments and plant observations, facilitating the identification of relevant patterns and trends. These two components can be integrated by:

- *Descriptive data analysis*: This involves calculating measures of central tendency, standard deviation, median, etc., to gain an initial understanding of the distribution of the data.
- *Statistical hypothesis testing*: Statistical tests can be applied to determine if there are significant differences between groups of plants or variables studied.

- *Regression and correlation analysis*: These methods can help identify relationships between different variables, such as the influence of environmental factors on plant performance.
- *Analysis of Variance (ANOVA)*: Can be used to assess differences between groups of plants within an experiment.

**Integrating biostatistics with graph data repositories:**

- Visualization of statistics results: The results of statistical analyses can be represented visually by means of graphs. For example, boxplots can be used to highlight differences between groups.
- Spatial data analysis: Using biostatistical methods to analyze spatial data on plant distribution and then representing it through a graph data warehouse can reveal geospatial patterns and interactions between plants.

**Machine Learning and Artificial Intelligence:** Integrating biostatistics with machine learning algorithms helps identify complex patterns and non-linear relationships in plant physiology data.

**Optimization of plant growth processes**: Using this set of methods can facilitate informed decision-making for improving plant growth conditions, such as optimizing the amount of water, nutrient levels, light, etc. By combining biostatistical methods with graph data repositories, a deeper and more detailed understanding of plant physiology and the factors influencing their development and performance can be obtained. This leads to improved agricultural practices and the development of more effective crop management strategies.

An important aspect addressed in the thesis refers to the modelling of experimental data, explaining how to collect, process and interpret the data in order to extract significant information with reference to the variation of some morphophysiological and biochemical parameters of plants, but also about the unfolding of some physiological processes in plant systems, in correlation with the living environment. This process is fundamental for a deeper understanding of physiological phenomena and for substantiating decisions in agriculture. Also, the thesis focuses on the use of graph data warehouses in agriculture, highlighting the general aspects of data warehouse development and its role in the management of data specific to the agricultural field.

The thesis presents ways of analyzing data from the physiology of plants grown on different nutrient media and supplemented with preparations aimed at increasing the quality and productivity of soybean plants and adapting to various environmental conditions for each studied genotype, as well as analyzing the influence of biostimulators on the plants and the result obtained.

Combining biostatistical methods with the use of graph data repositories in plant physiology research brings multiple benefits for understanding complex physiological processes and optimizing plant performance and adaptation to variable environmental factors. The use of biostatistical methods offers the possibility of analyzing experimental and observational data to identify patterns and relationships between variables related to the field of plant physiology and related fields.

One important aspect addressed in the thesis relates to the modeling of experimental data, explaining how data is collected, processed, and interpreted to extract meaningful information regarding the variation of morphophysiological and biochemical parameters of plants. It also delves into the understanding of physiological processes in plant systems, correlated with the living environment. This process is fundamental for a deeper understanding of physiological phenomena and for substantiating decisions in agriculture. Additionally, the thesis focuses on the use of graphical data repositories in agriculture, highlighting the general aspects of repository development and its role in managing domain-specific agricultural data.

For the development of specific biostatistical models, it is essential to establish a basic structure represented by a generic model that provides coherence and consistency in handling biostatistical information. This need arises from the complexity and diversity of biological datasets.

Key elements of the model include:

***Variable Declaration***: The main element that provides a list of variables to be analyzed within the biostatistical analysis model. These variables include:

- **<*LRad*>:** Variable associated with the parameter root length.
- **<*LTulp*>:** Variable associated with the parameter stem length.
- **<*VolRad*>:** Variable associated with the parameter radial volume.
- **<*BmPRad*>:** Variable associated with the parameter fresh root biomass.
- **<*BmURad*>:** Variable associated with the parameter dry root biomass.
- **<*BmPTulp*>:** Variable associated with the parameter fresh stem biomass.
- **<*BmUTulp*>:** Variable associated with the parameter dry stem biomass. These variables can represent various biological measurements or characteristics of biostatistical data and provide the foundation for further analysis.

***Determining Data Variation per Sample***: The main element that defines the general context of information, indicating that the structure focuses on analyzing data variation in a specific sample. ***Sub-elements***:

- ***Sample Mean*:** Retains the average value of data from the sample.
- ***Standard Error*:** Provides information about the standard error associated with sample data.
- ***Dispersion*:** Reflects the measure of data dispersion in the sample.
- ***Standard Deviation*:** Indicates the standard deviation value of the data.
- ***Coefficient of Variation*:** Provides a measure of the relative variability of data compared to the mean.
- ***Confidence Interval***: Specifies the range within which the population's mean is estimated to lie with a certain level of confidence.
- ***Null Values***: In the presented model, values for each sub-element are left null or empty, suggesting that these are places where specific data can be populated within a specific model. These sub-elements provide essential information for understanding the distribution of data and their associated statistical characteristics.

***Determining the Influence of Applied Factors***: The main element highlights the evaluation of the influence of applied factors and includes the following sub-elements:

- ***<DifferenceOfMeans>*:** Represents the difference between the means of two samples.
- ***<DifferenceError>*:** Provides information about the error associated with differences between samples.
- ***<t-Test>*:** Indicates the result of the t-test (Student), which is used to assess the significance of differences between means.
- ***<LSD_0.05>*:** Represents the threshold value for significant differences at a 95% confidence level.
- ***<LSD_0.01>*:** Represents the threshold value for significant differences at a 99% confidence level. These sub-elements are useful in evaluating the impact of factors on the data and determining the significance of observed differences.

**Confirmation of Null Hypothesis Rejection**: The main element that provides information to confirm the rejection of the null hypothesis and includes the sub-element

***<P-Test>*** with the specified formula ***<Formula>StudentTPValue[t,df]</Formula>***.

The element can be used to calculate the P-value and assess the statistical significance of the results. This sub-element provides a framework for evaluating the validity of the results and the rejection or acceptance of the null hypothesis in the context of biostatistical analysis.

The XML model offers a flexible and easily extendable structure to adapt biostatistical information according to the specific needs of the analysis or dataset. It is a useful tool for organizing and exchanging statistical information in a standardized manner.

Graph data stores provide a framework for organizing and analyzing complex relationships between different entities and variables. By representing the data as a graph, the interconnections between the components become more visible. Through graph algorithms, the complex dependencies and interactions that influence performance are revealed plants. Through data analysis and artificial intelligence algorithms, predictive models can predict the response of plants to different stimuli and environments.

In the process of data analysis, we often face the problem of having a small data set that does not provide enough information to obtain an accurate result. Small data sets can have a significant impact on the accuracy of drought or plant productivity predictions. Since the amount of data is limited, it becomes crucial to treat this problem carefully and use appropriate techniques and models to obtain relevant results [22-25]. Despite all efforts, it is notable that, in the case of small data sets, the predictions may show a higher uncertainty. It is imperative to highlight this level of uncertainty in the process of interpreting the results obtained and to expand, as much as possible, the size of the data set in order to reduce this uncertainty.

Among the important aspects to consider is the use of *Transfer Learning* (TL) techniques. TL techniques can be a valuable approach when data are limited. Pre-trained models can be used and adjusted to fit new data, which can improve performance. In recent years, due to the development of deep learning models and the increasing availability of massive data sets, Transfer Learning (TL) has become more prominent. The TL approach involves the use of a set of techniques that involve using knowledge and experience gained in one domain to improve the performance of a model or algorithms in another domain or similar task [22-25]. This technique is based on the premise that information and features learned from one domain can be useful and applicable in another domain, especially when the data set in the target domain is limited or insufficient to train an effective model from scratch.

**Chapter III „The intelligent application of biostatistics methods in plant physiology research"** describes some developed models of biostatistical analysis of experimental data, obtained

in the Biological Security laboratory of the „Biology and Geosciences" faculty of the State University of Moldova. Morphological parameters are analyzed of plants: root length and volume, hypocotyl and epicotyl length, internode length, total plant height. Biometric parameters were correlated with quantitative indices of fresh and dry root and stem biomass.

The thesis presents the components of the decision support system for the study of the chemical composition of soybeans (*Glycine max* (L.) Merrill.) for the analysis of organic and inorganic compounds with a decisive role in ensuring the quality of soybeans. The objectives of the software components were research assistance to determine the degree of influence of *Reglalg* (compound of algal nature) and *Biovit* (compound of humic nature) on the quality of soybeans [35].

**Table 1. Influence of compounds on nitrogen content (Clavera variety, subjected to foliar treatment with Reglalg and Biovit), (%)**

| Comparison | $\bar{d}$ | $S^{\bar{d}}$ | $t_{exp}$ | $lsd_{0.05}$ | $lsd_{0.01}$ |
|---|---|---|---|---|---|
| *Martor vs Reglalg* | 4.1 | 0.07 | 62.3* | 0.14* | 0.24** |
| *Martor vs Biovit* | 5.23 | 0.01 | 49.3* | 0.22* | 0.4** |

* Significant influence at a confidence level of 95%
** Significant influence at a confidence level of 99%

As a result of the processing of the data obtained as a result of conducting experiments in the laboratory, through the use of biometric methods, the necessary knowledge to achieve the set objectives was obtained. In this context, IT components were developed within the support system for the study of plant growth and development. The objectives of the components of the support system were research assistance to determine the degree of influence of the phytohormone Heteroauxina (0.0001%), the preparations Reglalg (1:300) and Biovit (1:500) on some morphophysiological parameters correlated with plant growth and development. The statistical hypothesis regarding the difference in means between independent pairs of samples was determined. Several parameters are analyzed: root and stem length, fresh and dry root and stem biomass.

The main results of the influence of some biologically active substances on the growth and development of soybean plants (*Glycine max* (L.) Merrill.) were identified and described. Through rigorous biostatistical analysis, significant correlations were identified between the effects of some biologically active compounds and the physiological processes of plants, providing a better understanding of their impact in the context of the specific conditions in the Republic of Moldova. An important aspect addressed in this chapter is the highlighting of genotypes with useful potential and resilience to the climatic conditions of the Republic of Moldova, in the context of current climate

changes. These genotypes are of particular importance for the selection of soybean varieties that can survive and develop optimally under variable environmental conditions.

Also, the data obtained experimentally regarding the leakage of electrolytes at the temperatures of 25 ° C, 46 ° C, 48 ° C, 50 ° C, 52 ° C and 100 ° C were processed. Below are the graphs that represent the results regarding the influence of the humic substance treatment for all four genotypes: *Clavera*, *Colina*, *Horboveanca* and *Dorința* (figures 1 and 2).

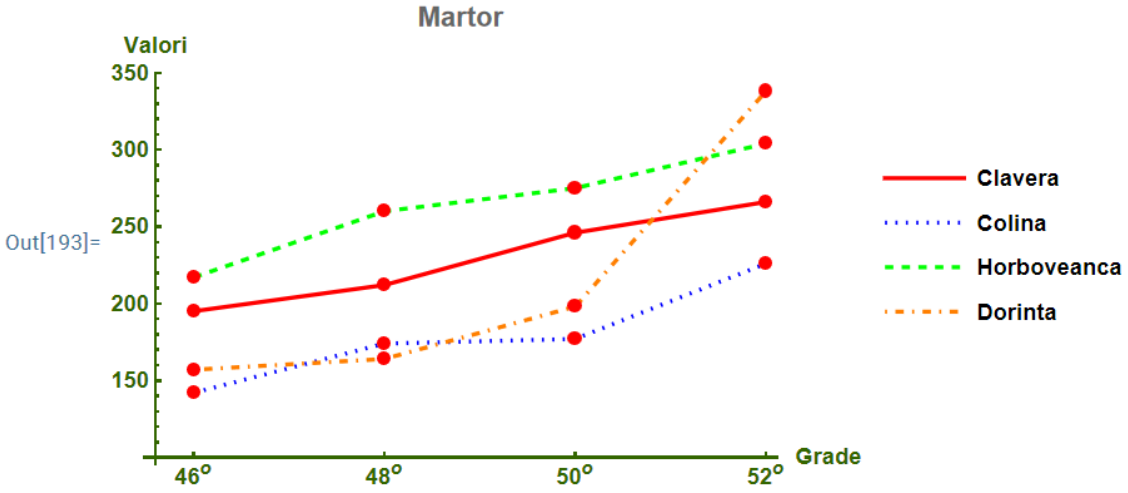



**Figure 1. Electrolyte leakage in the control group.**

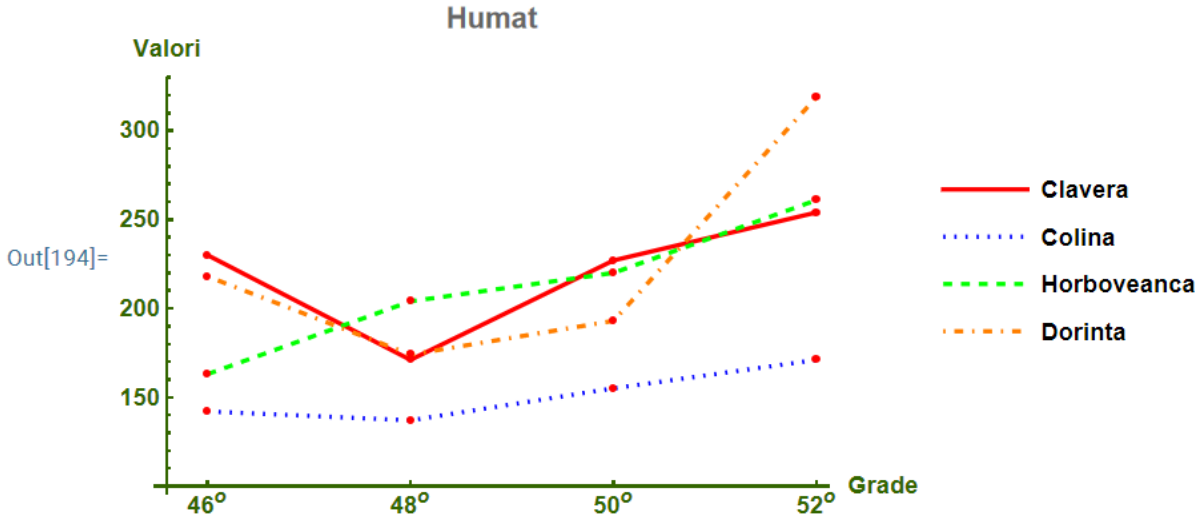



**Figure 2. Electrolyte leakage in the Humate-treated group.**

**Chapter IV** of the thesis presents in detail the graph data warehouse created to highlight the relevant connections in order to optimize agricultural processes. In this chapter, graphs are explored, used to process weather data and define nodes and relationships in Neo4j for said data.

These represent solutions for the achievement of research objectives in finding non-trivial interdependencies between genetic and physiological data according to various factors such as the influence of compounds with biological activity, nutrition, humidity, and temperature, as well as their impact on the agriculture of the Republic of Moldova.

*Neo4j* graph database management system was chosen to develop a data warehouse in agriculture. This system is a platform that focuses on graph management. Integrating *Neo4j* in this context can bring significant benefits in managing complex information and performing graph analysis and visualization. Graphs are useful for representing and visualizing complex relationships between different entities [36]. They allow a holistic approach to the management of the agricultural field and other fields related to it [37,38]. To process the data resulting from the experiments, graphs were made in the Neo4j DBMS by means of the Cypher language. To simplify data processing, a graph was developed for each experiment. These graphs are:

1. *Graph „Nitrogen, protein, phosphorus and pentaoxide content in soybeans";*
2. *Graph „Data processing related to the influence of some biologically active substances on plants grown in aqueous medium Knop";*
3. *Graph „Data Processing of Electrolyte Leakage in Soybean Plants";*
4. *Graph „Data processing related to thermotolerance in soybean plants";*
5. *Graph „Varieties and productivity";*
6. *Graph „Data processing related to the influence of biologically active substances on the water and oil content of soybeans";*
7. *Graph „Data processing regarding the influence of some biologically active substances on the growth of soybean plants";*
8. *The graph model „Data processing regarding the influence of biologically active substances on the growth of soybean plants".*

An important module for agricultural planning and management is weather data processing with all relevant data and information. Data from the period 2002 – 2022 obtained from the State Hydro-Meteorological Service and the National Bureau of Statistics of the Republic of Moldova are included [39, 40].

Node types and relationships between them were defined to model weather data. The weather data processing graph includes data on:

- *MYRC* node type, as well as the total annual precipitation amount (*YRC node type*);

- Average monthly temperatures and average and median annual temperatures (Types of nodes *MYTC*, *YTC*);
- The monthly and annual hydrothermal coefficient (Node types *MYHTCc*, *YHTCc*);
- Soil moisture at different levels (Node types *MSMC*, *YMSC*);
- Monthly mean wind speed (Node types *MYWSC*, *YWSC*);
- The average amount of fruit per hectare (Type of nodes *Prod*), *(Figure 3)*.

One of the important aspects addressed in this chapter is the prediction model of meteorological phenomena based on the data stored in the data warehouse. This model allows the anticipation of the evolution of meteorological conditions, being essential in the management and optimization of agricultural activities. Also, the result of the model for predicting the productivity of some agricultural products is shown, using the data available in the graph data warehouse. This model has the potential to support farmers in decision-making, optimizing crop yields.

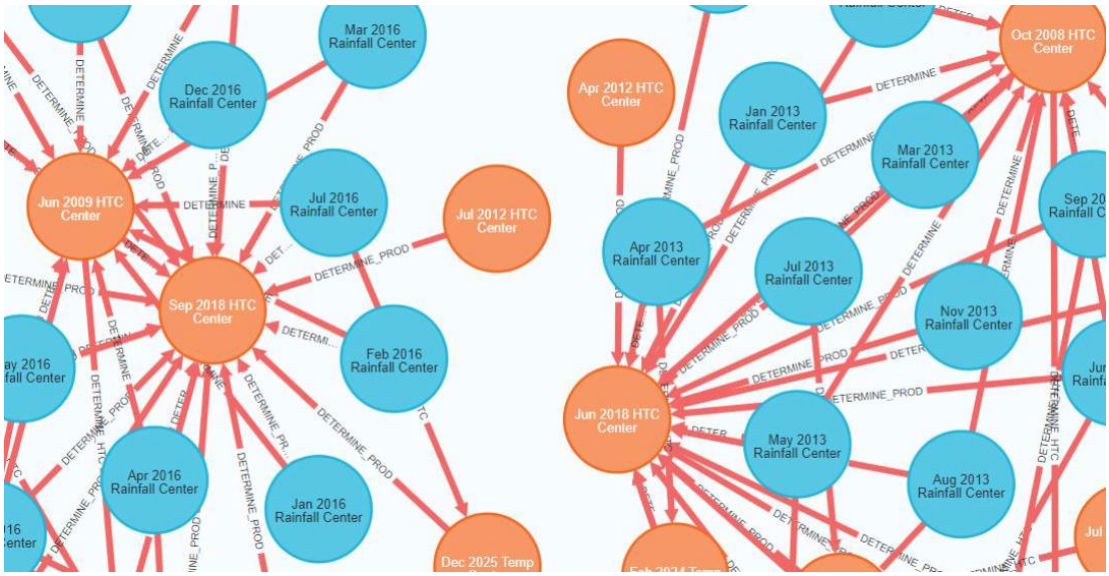



**Figure 3. A fragment of the „Meteorological data" graph.**

Once collected and processed, these data can be used to develop predictive models of plant physiological processes and to identify factors that affect these processes (Figure 4). This can help develop more efficient and sustainable agricultural technologies, as well as improve understanding of how plants work and interact with their environment.

```
1  //Proiecția grafului
2  CALL gds.graph.project('dtPred',{
3      YHTCc:{properties:['htc','drought']},
4      YHTCcp:{properties:'htc'}
5    },
6    '*'
7  );
```

Table | Text | Code

| nodeProjection | relationshipProjection | graphName | nodeCount | relationshipCount | projectMillis |
|---|---|---|---|---|---|
| {YHTCcp: {label: "YHTCcp", properties: {htc: {defaultValue: null, property: "htc"}}}, YHTCc: {label: "YHTCc", properties: {drought: {defaultValue: null, property: "drought"}, htc: {defaultValue: null, property: "htc"}}}} | {__ALL__: {orientation: "NATURAL", indexInverse: false, aggregation: "DEFAULT", type: "*", properties: {}}} | "dtPred" | 24 | 21 | 3678 |

**Figure 3. Creating the graph in memory for the prediction of some phenomena.**

## GENERAL CONCLUSIONS AND RECOMMENDATIONS

Intelligent data analysis in plant physiology is a current research topic due to the fact that it allows obtaining more detailed and diverse data on plant physiological processes, which allows the optimization of agricultural plant growth and production. The use of smart technologies can be successfully used to collect, analyze and interpret data in real-time, which can help improve crop yields, reduce costs and environmental impact, and optimize resource use. Following the research carried out, the following five general conclusions can be made:

1. **A knowledge base has been developed** in the research field „Integrated modelling of data in the field of plant physiology" (chapter „Integration of biostatistics methods and data warehouses in data processing from the field of plant physiology". The knowledge base was used to develop the models and application software products developed and was integrated into the developed software products.

2. The research field „Integrated data modelling in the field of plant physiology" was developed **(chapter** „Integration of biostatistical methods and data warehouses in data processing from the field of plant physiology"). The language is of considerable complexity. The application of conventional information technologies could require the involvement of an IT Company for its implementation on the computer. The original methodology for the development of software products oriented to problem families, used in the doctoral research allowed the realization of the formal language of the research field „Integrated Data Modeling in the Field of Plant Physiology" and contributes to advancing the deeper understanding of plant physiological processes.

3. **Were developed for the research area „Integrated data modelling in the field of plant physiology"** (chapter „Intelligent application of biostatistical methods in plant physiology research"). The processing of experimental data regarding the influence of biologically active compounds of natural origin highlighted the main biological variables significantly influenced in the administration of biopreparations, demonstrating high efficiency for indicators such as the content of nitrogen, proteins, phosphorus and phosphorus pentaoxide, as well as for the morphological parameters growth and development of plants on Knop aqueous nutrient medium. Positive influences on electrolyte leakage and plant tolerance to heat stress are also observed. However, it is noted that in some cases, the results were within the confidence interval or showed a negative influence. This aspect could be the subject of further investigations to better understand the mechanisms involved.

The developed models were awarded with: two Gold Medals, a Silver Medal, a Bronze Medal and a Diploma of Excellence at International Salons of Innovation and Scientific Research. Two copyrights were registered at AGEPI.

4. **Non-relational** (graph-type) *Neo4j* **database was developed, which**, unlike relational databases, demonstrates advantages in efficient data storage for biological and agricultural research, considerably expanding the nominal possibilities of graph databases in the application area of application software products that use this type of database (chapter „Graph Data Warehouse: Highlighting Connections for Optimizing Agricultural Processes").

5. Two intelligent models for predicting the probability of drought and the productivity of some agricultural crops have been developed and validated. These models provide essential tools for farmers, researchers and policy makers in their efforts to manage agricultural resources in a sustainable and resilient manner. By using advanced prediction technology, decision-making can be optimized, risks associated with meteorological and biological changes can be minimized, thus contributing to increased sustainability and efficiency in the agricultural sector.

6. **The developed program products** have been integrated into a Data Repository oriented to problem families for the research field „Integrated Data Modeling in the Field of Plant Physiology" that will influence informed decision-making by users and decision-makers in the assessment of complex interactions between plants, soil, climate.

7. The use of the paradigm of program product development on problem families and the use of Transfer Learning techniques represent potential solutions for the development of research in various research fields.

Overall, the results provide them with a coherent and informative picture of the results obtained in the research, as well as possible directions for the further development of the field.

**Recommendations for future research**

The research focused on the use of Problem Family Oriented Data Repository for the research area „Biostatistical Data Modeling in Plant Physiology".

This one research methodology can be extended to other agricultural crops. Based on these researches, it is recommended:

a) She continues to use the models, techniques and software products of intelligent data analysis in plant physiology for scientific research at the Faculty of Biology and Geosciences.
b) To expand the scope of application of the Data Warehouse for biostatistical data modelling in the field of plant physiology in other research in the Republic of Moldova and abroad.
c) To expand research on biostatistical data modelling in the field of plant physiology for other agricultural crops in cooperation with other Institutions from the Republic of Moldova and abroad.
d) Completing the database and data warehouse with new data and information, as well as expanding sources and tools to create a *data lake* (*data lake*).
e) Development continuous maintenance of the data warehouse and its integration into a collaborative system.

## BIBLIOGRAPHY


1. DEXONLINE. Technology. [cited: 16.06.2020] Available: https://dexonline.ro/definitie/tehnologie
2. MARCUS, Florin. Dictionary of neologisms. Chișinău: Știința, 2011.
3. AGUNDERNEWS. Regenerative Agriculture 5.0: The Future of Farming?, @2020 [cited: 2020-08-05]. Available: https://agfundernews.com/regenerative-agriculture-5-0-the-future-of-farming.html
4. WEI, W., CHENG, Y., ZHU, X., & LI, D. (2020). Deep Learning-Based Crop Recognition for Intelligent Agriculture. Sensors, 20, 1-19. https://doi/10.3390/s20133419
5. GOLAN, L. HELLER, A. How AI and ML can translate data into wisdom, 7 Dec 2020. [cited: 28.08.2022]. Available: https://www.nokia.com/blog/how-ai-and-ml-can-translate-data-into-wisdom/
6. SHOAIB, 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. https://doi/10.3389/fpls.2023.1158933
7. Webb, Sarah. Deep learning for biology. Nature. 554. 555-557. 2018. 10.1038/d41586-018-02174-z. [cited: 10.05.2022]. Available: https://www.nature.com/articles/d41586-018-02174-z
8. MOHANTY, SP., HUGHES, DP., SALATHÉ, M. Using Deep Learning for Image-Based Plant Disease Detection. Front. Plant Sci. 7:1419.2016. https://doi/10.3389/fpls.2016.01419
9. 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.
10. 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.
11. 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

12. PAN, SJ, YANG, QA Survey on Transfer Learning, in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010, https://doi/10.1109/TKDE.2009.191
13. RUDER, S., PETERS, ME, 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. https://doi/10.18653/v1/N19-5004
14. 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.
15. 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
16. TAVASSOLY I, GOLDFARB J, IYENGAR R. Systems biology primer: the basic methods and approaches. Essays Biochem. 2018 Oct 26;62(4):487-500. two: 10.1042/EBC20180003. PMID: 30287586. https://pubmed.ncbi.nlm.nih.gov/30287586/
17. ISB. What is Systems Biology. https://isbscience.org/about/what-is-systems-biology/
18. XU, C., JACKSON, SA Machine learning and complex biological data. Genome Biol 20, 76 (2019). https://doi.org/10.1186/s13059-019-1689-0
19. SCIENCE. [cited: 28.08.2022]. Available: https://www.scientia.ro/univers/76-sisteme-complexe/852-sistemele-complex-xe-scurta-introducere.html
20. DAI X and SHEN L (2022) Advances and Trends in Omics Technology Development. Front. Med. 9:911861. https://doi/10.3389/fmed.2022.911861
21. PAN, SJ, YANG, QA Survey on Transfer Learning, in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010, https://doi/10.1109/TKDE.2009.191
22. RUDER, S., PETERS, ME, 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. https://doi/10.18653/v1/N19-5004

23. 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.
24. 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
25. JIA, F., SHAO, L., LI, Y., WANG, Y., & WANG, H. Deep Learning for Crop Yield Prediction in Precision Agriculture: A Review. Sensors, 20, 1-21, 2020. https://doi/10.3390/s20154230.
26. PATEL, M., CHANDEL, T., BAJAJ, D., & KALIA, S. 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, 2021. https://doi/10.5815/ijisae.2021.03.06
27. ROY, A., SAHA, M., GHOSH, K., & GIRI, B. 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), 2022, pp. 102-107. https://doi/10.1109/iCoMET49435.2020.9133266
28. AHMED, KEA, HOSSAIN, MS, UDDIN, MS, & ALAMRI, A. An Efficient and Comprehensive IoT-Based System for Smart Agriculture. Sensors, 20, 1-16, 2020. https://doi/10.3390/s20174834
29. AMARA, K., BOUGUESSA, M., & MOUSSAOUI, S. A Survey on the Applications of Machine Learning in Agriculture: A Review. Journal of Information Science and Engineering, 36, 1041-1066, 2020. https://doi/10.6688/JISE.202007_36(4).0011
30. KUMAR, S., KAUR, H., & SINGH, R. 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), 2019. https://doi/10.1109/SPIN.2019.8711667

31. LEE, SH, LEE, JY, LEE, JH, KIM, MY, & LEE, DK. Agricultural IoT System for Real-Time Monitoring and Prediction of Environmental Parameters Affecting Crop Production. IEEE Access, 7, 113758-113767, 2019. https://doi/10.1109/ACCESS.2019.2934276
32. MAJUMDAR, J., NARASEEYAPPA, S. & ANKALAKI, S. Analysis of agricultural data using data mining techniques: application of big data. J Big Data 4, 20, 2017. https://doi.org/10.1186/s40537-017-0077-4
33. KAO, P.-H., BAIYA, S., LAI, Z.-Y., HUANG, C.-M., JHAN, L.-H., LIN, C.-J., LAI, Y.- S., & KAO, C.-F. 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, 2022. (https://doi/10.3389/fpls.2022.1019709
34. CAPĂȚÂNĂ, Gh. A Programming Paradigm Oriented to Families of Problems. In.: The Scientific Bulletin Addendum No. 4/2019, The Official Catalog 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://cadetinova.ro/documente/Supliment_Inova_19.pdf
35. McDONALD, JH. Handbook of Biological Statistics. University of Delaware: Sparky House Publishing, 2009. - 180 p.
36. ROBINSON, R; WEBBER J.; EIFREM, E. Graph Databases. O'Reilly, Sebastopol, California, USA 2015.
37. TOWARDS DATA SCIENCE. MORGANTE, V. What is a graph database? [cited: 22.09.2022]. Available: https://towardsdatascience.com/what-is-a-graph-database-249cd7fdf24d
38. GRAPH IN AGRO. Graphs Are Feeding the World. https://neo4j.com/blog/graphs-feeding-world/
39. MOLDOVA STATISTICAL DATA BANK. [cited: 28.06.2021]. Available: https://statbank.statistica.md/PxWeb/pxweb/ro/?rxid=2f5a0079-c3c1-461e-81a4-d854b7ba0fcd
40. STATE HYDROMETEROLOGICAL SERVICE. [cited: 12.09.2022]. Available: http://www.meteo.md/index.php/clima/cercetri-climatice/

## LIST OF PUBLICATIONS

1. **GANEA, I**. Aplicarea metodelor text mining în căutarea avansată a documentelor de tip text. În: Analele științifice ale Universității de Stat din Moldova. Științe ale naturii și exacte. Chișinău: CEP, 2018, pp. 146-150. ISBN 978-9975-142-55-7.
2. **GANEA, I.** Model of statistical data analysis on nitrogen content in soybeans (Glicine max Merill) 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. (Category **B+**).
3. BÎRSAN, A. SPRINCEAN, V. **GANEA, I**. Opportunities for the practical use of intelligent technologies in the evaluation of some biological variables. In: Studia Universitatis Moldaviae, No. 1(2023), pp. 53-64. https://doi.org/10.59295/sum1(171)2023_07. (**Category B**).
4. **GANEA, I.** The influence of some biologically active substances on the growth and development of soybean plants grown on aqueous nutrient media 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.
5. **GANEA, I.** The development of the graph database for the analysis of biological data. Volume of papers: International Students' Conference „StudMath-IT 2021", „Aurel Vlaicu" University of Arad, Faculty of Exact Science Arad, Romania, November 18-19, 2021
6. **GANEA,** I. Support system for the study of the development of plants grown on M1 media. The national scientific conference of doctoral students „Contemporary research and evaluation methodologies" April 22-23, 2021. Chisinau, 2022, CEP USM. CZU:004.8: [57.087.1:631.53], pp. 103-107.
7. **GANEA,** I. Use of bio-statistics in data analysis of the biologically active substances on plants. Mathematics & IT: Research and education (MITRE – 2021), International Conference. Abstracts. Chisinau, 2021. CEP USM, 2021. ISBN 978-9975-158-19-0. p. 104.
8. **GANEA, I.** BÎRSAN, A. Biological data processing. International Exhibition InventCor. – II-nd edition. Posters. ISBN 978-606-35-0467-9, p. 44, **Gold Medal**.
9. **GANEA, I.** BÎRSAN, A. CĂPĂȚÂNĂ, G. Using the neo4j Pearson similarity algorithm for the analysis of protein content in soybean. The Official Catalog 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. „Nicolae Bălcescu" Land Forces Academy in Sibiu. ISSN 2501-3157, p. 120-131.

10. **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, Iasi, 2022. ISSN Print: 2601 -4564, Online: 2601-4572. Page 167 – 168, **Silver medal**.
11. **GANEA, I.** BÎRSAN, A. Use of neo4j link prediction algorithms in biological data processing. In „International Salon of Inventions, Innovations „Traian Vuia"". Timișoara, October 8-10, 2022 / coord.: Romi Rădulescu. Timisoara, Politehnica Publishing House, 2022, p. 77, ISBN 978-606-35-0496-9, **Bronze medal**.
12. **GANEA, I.** BÎRSAN, A. CĂPĂȚÂNĂ, G. Analysis model of electrolyte leakage in soybean plants. Presented at the national scientific conference with international participation „Integration through Research and Innovation", 10.11.2022, pp. 51-53.
13. BÎRSAN, Ana, FRUNZĂ, Maria, **GANEA, Ion**, DIACONU, Ecaterina. Invasive species of woody plants in the „Old Orheiul" protected area. In: *Natural sciences in the dialogue of generations*, September 14-15, 2023, Chisinau. Chisinau: CEP USM, 2023, p. 135. ISBN 978-9975-3430-9-1.
14. **GANEA, I.** BÎRSAN, A. SMANTĀNA, L. CĂPĂȚÂNĂ, Gh. Model for statistical analysis of data on nitrogen and phosphorus content in plants. Copyright. Series O No. 7091 of 08.12.2021
15. **GANEA, I.** Productivity prediction model of some agricultural crops. Copyright. Series OȘ No. 7552 of 21.06.2023
16. **GANEA**, **I**., CĂPĂȚÂNĂ, Gh. Intelligent Support System „Biostatistical Modeling of Data in the Field of Plant Physiology". In: „Decision Making and Decision Support in the Information and AI Era", (preprint).
17. **GANEA, I.** Prediction Model for Soybean Productivity. Database Systems Journal (preprint).

## ADNOTARE

**Ion GANEA „Modele, tehnici și produse program de analiză inteligentă a datelor în fiziologia plantelor”. Teza de doctor în informatică, Chișinău, 2024.**

**Structura tezei:** Teza este scrisă în limba română și constă din Introducere, patru capitole, concluzii și recomandări. Bibliografia cuprinde 220 de titluri. Lucrarea conține 135 de pagini de text de bază, 37 de figuri, 17 tabele și 32 de anexe. Rezultatele obținute au fost publicate în 17 lucrări științifice cu un volum de peste 5 coli de autor.

**Cuvinte cheie:** tehnologii inteligente, fiziologia plantelor, 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; Crearea Depozitului de date eficient pentru cercetările biologice și agricole.

**Noutatea și originalitatea științifică:** model generic în limbaj formal al domeniului „Modelarea integrată a datelor în domeniul fiziologiei plantelor” (MIDDFP), 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. 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, Chisinau, 2024.**

**Thesis Structure:** The thesis is written in Romanian and consists of an Introduction, four chapters, conclusions, and recommendations. The bibliography includes 220 titles. The work contains 135 pages of main text, 37 figures, 17 tables, and 32 appendices. The obtained results have been published in 17 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 research; Evaluating factors influencing agricultural efficiency, considering complex interactions.

**Scientific Novelty and Originality:** A generic model in the formal language of the field „Integrated Data Modeling in Plant Physiology" (MIDDFP), 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, 2024.**

**Структура диссертации:** Диссертация написана на румынском языке и состоит из Введения, четырех глав, заключения и рекомендаций. Библиография включает 220 названий. Работа содержит 135 страниц основного текста, 37 иллюстраций, 17 таблиц и 32 приложений. Полученные результаты были опубликованы в 17 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 для обнаружения знаний и результатов в области биологии на основе анализа большого объёма 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 „Интегрированное Моделирование Данных в Физиологии Растений” (MIDDFP), 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.

**ION GANEA**

# MODELS, TECHNIQUES, AND SOFTWARE PRODUCTS FOR INTELLIGENT DATA ANALYSIS IN PLANT PHYSIOLOGY

**121.03. COMPUTER PROGRAMMING**

**Abstract of the doctoral thesis in computer science**


Approved for printing: 29.03.2024 Paper format 60x84 1/16
Offset paper. Offset printing. Circulation 15 ex
Form sheets.: 1.8 Order no. ….

USM Editorial-Polygraphic Center
60, Al. Mateevici str., Chisinau, MD-2009
Email: cep1usm@mail.ru, usmcep@mail.ru