MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV

Qualification
scientific
work on manuscript
rights

HUANG MINGXIN

UDC 004.58+378

PHD THESIS

INFORMATION TECHNOLOGY FOR MONITORING CROP YIELDS USING GEOINFORMATION SYSTEMS

126 Information Systems and Technology

12 Information Technology

Applying for the Doctor of Philosophy degree

The PhD Thesis contains the results of own research. The use of ideas, results and texts of other authors are linked to the corresponding source

黄明鑫
Huang Mingxin

_________________ Mingxin Huang

Supervisor – Andrii Onyshchenko, Doctor of Economic Sciences, Professor

KYIV – 2024

SUMMARY

Huang Mingxin. Information technology for monitoring crop yields using geoinformation systems. – *Qualifying scientific work as a manuscript.*

Thesis for the Doctor of Philosophy Degree in Specialty 126 «Information Systems and Technology». – Taras Shevchenko National University of Kyiv, Kyiv, 2024.

Abstract Content. The dissertation is devoted to the development of models, methods, and tools for data processing aimed at monitoring crop yields and integrating with Geographic Information Systems (GIS).

As the global population grows, so does the need for food. Efficient agriculture can produce more food per unit of land, ensuring food security for an increasing number of people.

Yield monitoring in the context of project management in agriculture is crucial for enhancing its efficiency. This process allows farmers to analyze the impact of various agronomic factors, such as soil type, fertilizer use, and water availability, on crop yields, facilitating resource optimization and reducing environmental impact. Moreover, monitoring is a critical tool for developing effective strategies for managing agricultural projects.

Project management is becoming increasingly widespread in agriculture, as it promotes effective organization and management of agricultural projects, enhancing their productivity and profitability. In this context, yield monitoring serves as a fundamental tool for project management, providing valuable information for making informed decisions, planning, and controlling the execution of agricultural projects. Identifying the impact of external and internal factors on yield allows project managers to adapt strategies and optimize resources, ensuring the resilience and efficiency of agricultural initiatives. Thus, integrating project management in agriculture, supported by effective monitoring, opens new opportunities for enhancing productivity, adapting to climate change, fostering innovation, and achieving global food security.

Therefore, scientific research on yield monitoring not only contributes to improving the productivity and efficiency of agricultural practices but also plays a key role in ensuring food security, sustainable development, and economic well-being on a global scale.

This work addresses a critical task: the development of information technology that includes mathematical models, methods, and procedures for yield monitoring based on geoinformation data (scientific component), as well as the development of a yield monitoring information system that enables the automation of data collection, processing, and utilization of geoinformation data for yield forecasting (practical component).

The object of the study is yield monitoring.

The subject of the study is models, methods and information technology of yield monitoring based on interaction with Geographic Information Systems.

The study aim is to develop models, methods and data processing procedures necessary for yield monitoring.

Research Methods. The conducted research is based on methods of systems analysis, technical analysis, artificial intelligence, big data processing, and object-oriented programming.

Scientific novelty of the obtained results:

- *For the first time*, an integration model of artificial intelligence for yield monitoring has been developed, based on the combination of multispectral images and geoinformation data. This model integrates Convolutional Neural Networks and Recurrent Neural Networks to enhance the accuracy and sensitivity of monitoring.
- The mathematical model of the relationship between phenological indicators and crop yields *has been improved*. Unlike other models, this combined model includes an adaptive threshold method for determining the membership of crop pixels and identifying the trend and seasonal components of the phenological indicators' time series. This improvement enhances the accuracy of forecasting.

- The information technology for yield monitoring based on geoinformation data *has been improved*. The enhancement involves the use of a combined model of the relationship between phenological indicators and crop yields, as well as an integration model of artificial intelligence. Unlike other technologies, the developed technology takes into account a wider variety of data, which simplifies integration with Geographic Information Systems (GIS).
- The methods for representing and storing geoinformation data *have been further developed* in terms of correlating key properties of agricultural objects with aerospace images of the locality.
- Information technologies for project management have *further developed* in the aspects of monitoring and forecasting yields and integration with geographic information systems.

The first chapter, an analysis of the scientific literature was conducted, which established that the use of digital images of geographical areas and the development of Geographic Information Systems (GIS) are becoming key in modern agriculture. This facilitates effective management of cultivated areas, analysis, and prediction of yields, especially in the context of the increasing demand for food against the backdrop of a growing global population. The advancement of technologies provides new opportunities for intensification and optimization in the agricultural sector. It was also established that neural networks are effective tools for yield prediction. They can model complex nonlinear dependencies in agronomic data and process satellite imagery and remote sensing data.

It was discovered that no existing service or software combines all the necessary capabilities for crop area management: anomaly detection, identification of phenological changes, and yield estimation. It is shown that a crucial task is to create specialized software that would allow uploading and working with large archives of images and would have built-in methods for intelligent data processing and pattern recognition.

The second chapter describes the use of aerospace imagery and time series analysis of images to determine phenological indicators and other important growth and health indicators of plants. The importance of using the NDVI index as a key indicator for assessing plant cover and yield is emphasized.

An approach to data processing and analysis within the context of Geographic Information Systems is considered. The vast volume of geographical data requires reduction for computational processing. The chapter also covers the structure and organization of geodata in vector and raster formats, revealing their unique capabilities and limitations for representing and analyzing geographic information models.

A conceptual model of the GIS for agricultural monitoring was developed. The system development is divided into four stages: defining objectives, describing functionality, implementation, and diagnostics. Each stage includes steps that facilitate the creation of an effective system for monitoring and managing agricultural crop yields. The importance of a systematic approach to the creation and use of Geographic Information Systems in agriculture is established.

The third chapter describes a mathematical model of the relationship between phenological indicators and the yields of agricultural crops and biomonitoring, which considers multispectral field images for dynamic yield forecasting. The decomposition of phenological indicators into trend, seasonal, and random components is aimed at effective yield monitoring. The model includes image binarization steps to define crop areas using a threshold function and the Otsu method for selecting the optimal threshold value.

The process of creating and training a hybrid neural network, integrating image data and soil information for yield prediction, is described. The network architecture includes convolutional neural networks (CNN) for image processing and fully connected layers for soil data analysis. This integration allows the network to consider diverse information, enhancing its ability to accurately predict yields. In

the second phase, the network uses recurrent neural networks to analyze data sequences, adding the ability to account for temporal dependencies and context.

The fourth chapter describes the development of a GIS-based yield monitoring information system to enhance the efficiency of the agricultural sector. The system's modular structure includes modules for data collection, storage, processing, visualization, and analysis, including the use of machine learning for forecasting and process optimization.

An algorithm for implementing an artificial intelligence integration model for yield monitoring based on the combination of multispectral images and geoinformation data is described, which includes seven stages: data collection and preparation, neural network development, testing and validation, optimization, implementation in agricultural systems, and further analysis of results.

Practical significance of the obtained results. The main scientific provisions of the dissertation have been elevated to the level of methodological generalizations and applied tools, enabling yield monitoring.

The agricultural crop yield monitoring information system was validated through comparative analysis methods. Comparing the yield predictions for winter wheat, corn, and barley in the Chernihiv region for 2019 with the predictions using the WOFOST simulation model and data from the State Statistics Service of Ukraine for 2019 shows that the monitoring model can provide sufficiently accurate yield forecasts. It was established that yield is significantly determined by plant development in the first three months after emergence, highlighting the importance of monitoring during this period. The obtained practical results emphasize the potential and limitations of using yield monitoring information technology with Geographic Information Systems.

The main provisions and results of the research have been implemented and applied in the activities of Yancheng Polytechnic College.

Keywords: yield monitoring, neural network, machine learning, GIS, project management, biomonitoring, information management, critical infrastructure**.**

LIST OF PUBLICATIONS OF THE APPLICANT BY PHD THESIS TOPIC

Articles in professional publications of Ukraine
(included in the list of the Ministry of Education and Science of Ukraine)

1. **Mingxin, Huang,** & Vatskel, V. (2019). Digital image analysis technologies for decision support systems in agricultural. Management of development of complex systems, 37, 164–167, https://doi.org/10.6084/m9.figshare.9783227 [Category B]
2. **Mingxin, Huang**. (2019). Review of monitoring and forecasting tools of the crop yield. Management of development of complex systems, 38, 161–167, https://doi.org/10.6084/m9.figshare.9788696 [Category B]
3. **Huang, M.,** & Shabala, Y. Y**.** (2019). Conceptual model of geographic information system for agriculture. Scientific Bulletin of Uzhhorod University. Series of Mathematics and Informatics, 2(35), 149–155. https://doi.org/10.24144/2616-7700.2019.2(35).149-155 [Category B]
4. **Mingxin, Huang.** (2024). Development of a yield monitoring model based on analysis of surveys and images of the field. Management of development of complex systems, 57, 67–71. https://doi.org/10.32347/2412-9933.2024.57.67-71 [Category B]

Articles in professional publications of Ukraine

(not included in the list of the Ministry of Education and Science of Ukraine)

1. **Mingxin, Huang.** (2020). Use of geoinformation systems for agricultural problems. *Science Journal Innovation Technologies Transfer*, 61–64. https://doi.org/10.36381/iamsti.4.2020.61-64

Approbation works

1. **Mingxin, Huang**. (2018). Use of geoinformation systems in agriculture. V International Scientific and Practical Conference "Information Technologies and Interactions", November 20–21, 2018, 63.

2. **Mingxin, Huang**. (2019). Information technology for efficient management of crop yields. XV International Scientific and Practical Conference "Project Management in the Development of Society", May 17-18, 2019, 40-41

3. **Mingxin, Huang**. (2019). Use of digital images if geographical areas for thr purposes of agricultural. I International Scientific and Practical Conference IMTCK2019, 65–66/

4. **Mingxin, Huang**. (2019). Participatory sensing for monitoring and forecasting of environmental pollution. VI International Scientific and Practical Conference "Information Technologies and Interactions", December 20, 2019, 95–96.

5. **Mingxin, Huang.** (2020). Using Geoinformation Systems for Agriculture Tasks. *Seventh international scientific-practical conference «Management of the development of technologies» Topic: "Information technology development of educational content» Kyiv, 25 – 26 March 2020*, 127-128. [In Ukrainian]

6. **Huang, M.,** Biloshchytskyi, A., Andrashko, Y. & Omirbayev, S. (2021). A Conceptual Research Model of Development of the Geographic Information System for Agriculture, *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Nur-Sultan, Kazakhstan, 1-5, https://doi.org/10.1109/SIST50301.2021.9465973. [**Scopus, Web of Science**]

АНОТАЦІЯ

Хуан Мінсінь. Інформаційна технологія моніторингу врожайності сільськогосподарських культур з використанням геоінформаційних систем. – *Кваліфікаційна наукова праця на правах рукопису.*

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 126 «Інформаційні системи та технології». – Київський національний університет імені Тараса Шевченка, Київ, 2024.

Зміст анотації. Дисертація присвячена розробці моделей, методів та засобів обробки даних для моніторингу врожайності та інтеграції з геоінформаційними системами.

З ростом світового населення зростає і потреба в продовольстві. Ефективне сільське господарство може виробляти більше продуктів харчування на одиницю площі, забезпечуючи продовольчу безпеку для наростаючої кількості людей.

Моніторинг врожайності в контексті проектного управління сільським господарством має вирішальне значення для підвищення його ефективності. Цей процес дозволяє фермерам аналізувати вплив різноманітних агрономічних факторів, таких як тип ґрунту, використання добрив і наявність води, на урожайність, що сприяє оптимізації ресурсів і зменшенню екологічного впливу. Водночас, моніторинг є критичним інструментом для розробки ефективних стратегій управління сільськогосподарськими проектами.

Проектний менеджмент набуває все більшої поширеності в сільському господарстві, оскільки він сприяє ефективній організації та управлінню аграрними проектами, підвищуючи їх продуктивність і рентабельність. В цьому контексті, моніторинг врожайності виступає як фундаментальний інструмент проектного менеджменту, оскільки він надає цінну інформацію для прийняття обґрунтованих рішень, планування та контролю за виконанням аграрних проектів. Визначення впливу зовнішніх і внутрішніх факторів на

урожайність дозволяє менеджерам проектів адаптувати стратегії та оптимізувати ресурси, забезпечуючи стійкість і ефективність аграрних ініціатив. Таким чином, інтеграція проектного менеджменту в сільському господарстві, підкріплена ефективним моніторингом, відкриває нові можливості для підвищення продуктивності, адаптації до змін клімату, інноваційного розвитку та досягнення продовольчої безпеки на глобальному рівні.

Таким чином, наукові дослідження моніторингу врожайності не лише сприяють підвищенню продуктивності та ефективності сільськогосподарських практик, але й відіграють ключову роль у забезпеченні продовольчої безпеки, сталому розвитку та економічному благополуччі на глобальному рівні.

В даній роботі вирішується важливе завдання, а саме: розроблення інформаційної технології, яка включає математичні моделі, методи та процедури для моніторингу урожайності на основі геоінформаційних даних (наукова складова), а також розроблення інформаційної системи моніторингу урожайності, що дозволяє автоматизувати збір, обробку та використання геоінформаційних даних для прогнозування врожайності (практична складова).

Об’єктом дослідження є моніторинг врожайності.

Предметом дослідження є моделі, методи та інформаційні технології моніторингу врожайності на основі взаємодії з Геоінформаційними системами.

Метою дослідження є розробка моделей, методів і процедур обробки даних, необхідних для моніторингу врожайності.

Методи дослідження. Проведені дослідження базуються на методах системного аналізу, технічного аналізу, штучного інтелекту, обробки великих даних та об'єктно-орієнтованого програмування.

Наукова новизна отриманих результатів :

- *Вперше* розроблено інтеграційну модель штучного інтелекту для моніторингу врожайності на основі поєднання мультиспектральних зображень та геоінформаційних даних. Ця модель поєднує Згорткові нейронні мережі (CNN) та Рекурентні нейронні мережі (RNN) для покращення точності та чутливості моніторингу.

- *Удосконалено* математичну модель взаємозв'язку між фенологічними показниками та врожайністю сільськогосподарських культур, яка, на відміну від інших моделей є комбінованою та включає адаптивний пороговий метод визначення належності пікселів посіву та виділення трендової та сезонних складових часового ряду фенологічних показників. Це дає змогу підвищити точність прогнозування.

- *Удосконалено* інформаційну технологію моніторингу урожайності на основі геоінформаційних даних. Вдосконалення полягає у використанні комбінованої моделі взаємозв'язку між фенологічними показниками та врожайністю сільськогосподарських культур та інтеграційної модель штучного інтелекту. На відміну від інших, розроблена технологія враховує більшу різноманітних даних, що спрощує інтеграцію з геоінформаційними системами (GIS) та

- *Набули подальшого розвитку* методи представлення та зберігання геоінформаційних даних в частині співставлення ключових властивостей об'єктив господарювання із аерокосмічними знімками місцевості.

- *Набули подальшого розвитку* інформаційні технології проєктного управління в частині моніторингу та прогнозування врожайності та інтеграції з геоінформаційними системами.

В першому розділі проведено аналіз наукової літератури в результаті якого встановлено, що використання цифрових зображень географічних територій і розвиток геоінформаційних систем стають ключовими в сучасному сільському господарстві. Це сприяє ефективному управлінню посівними площами, аналізу та прогнозуванню врожайності, особливо в контексті зростаючої потреби в продовольстві на фоні збільшення світового населення. Розвиток технологій, надає нові можливості для інтенсифікації та оптимізації аграрного сектору. Також встановлено, що нейронні мережі є ефективними інструментами для прогнозування урожайності. Вони здатні моделювати складні нелінійні залежності в агрономічних даних, а також обробляти супутникові зображення та дані дистанційного зондування.

Виявлено, що жоден з сервісів або програмного забезпечення не поєднує в собі всі необхідні можливості з обробки і аналізу даних, які необхідні для управління посівними площами: знаходження аномалій, ідентифікація фенологічних змін, оцінювання врожайності. Показано, що важливим завданням є створення спеціалізованого програмного забезпечення, що дозволяло б завантажувати і працювати з архівами знімків великого обсягу і мало б вбудовані методи інтелектуальної обробки даних і розпізнавання образів.

В другому розділі описано використання аерокосмічних знімків та аналіз часових рядів зображень для визначення фенологічних показників та інших важливих індикатори росту та здоров'я рослин. Підкреслено важливість використання індексу NDVI як ключового показника для оцінки рослинного покриву і врожайності.

Розглянуто підхід до обробки та аналізу даних у контексті геоінформаційних систем. Великий обсяг географічних даних вимагає редукції для обробки засобами обчислювальної техніки. Розділ також охоплює структуру та організацію геоданих у векторних та растрових форматах, розкриваючи їх унікальні можливості та обмеження для представлення та аналізу географічних інформаційних моделей.

Також побудовано концептуальну модель дослідження геоінформаційної системи для моніторингу сільського господарства. Розробка системи поділена на чотири етапи: визначення мети, опис функціональності, реалізація та діагностування. Кожен етап включає кроки, які сприяють створенню ефективної системи для моніторингу та управління врожайністю сільськогосподарських культур. Встановлено важливість системного підходу до створення та використання геоінформаційних систем у сільському господарстві.

В третьому розділі описано математичну модель зв'язку між фенологічними показниками та врожайністю сільськогосподарських культур яка розглядає мультиспектральні зображення поля для прогнозування врожайності в динаміці. Розділення фенологічних показники на трендову, сезонну та випадкову складові спрямовані на забезпечення ефективного моніторингу врожайності. Модель включає кроки бінаризації зображення з використанням порогової функції, для визначення областей засаджених культурою і метод Оцу для вибору оптимального порогового значення.

Також описано процес створення та навчання гібридної нейронної мережі, яка інтегрує дані з зображень та інформацію про ґрунт для прогнозування урожайності. Архітектура мережі включає згорткові нейронні мережі (CNN) для обробки зображень та повністю з'єднані шари для аналізу даних про ґрунт. Ця інтеграція дозволяє мережі враховувати різноманітну інформацію, що збільшує її здатність точно прогнозувати урожайність. На другому етапі в мережі використовуються рекурентні нейронні мережі для аналізу послідовностей даних, що додає здатність враховувати часові залежності та контекст.

В четвертому розділі описано розробку інформаційної системи моніторингу урожайності, яка використовує дані з GIS для підвищення ефективності аграрного сектора. Описано модульну структуру системи, що включає модулі для збору даних, їх зберігання, обробки, візуалізації та аналізу, включаючи використання машинного навчання для прогнозування та оптимізації процесів.

Описано алгоритм для реалізації інтеграційної моделі штучного інтелекту для моніторингу врожайності на основі поєднання мультиспектральних зображень та геоінформаційних даних, який включає сім етапів: збору та підготовки даних, розробку нейронної мережі, її тестування та валідацію, оптимізацію, імплементацію в аграрні системи та подальший аналіз результатів.

Практичне значення одержаних результатів. Основні наукові положення дисертації доведені до рівня методичних узагальнень і прикладного інструментарію, що дає змогу здійснювати моніторинг врожайності .

Інформаційну систему моніторингу врожайності сільськогосподарських культур валідовано методам порівняльного аналізу. Порівняння прогнозів урожайності озимої пшениці, кукурудзи, та ячменю в Чернігівській області за 2019 рік з прогнозування за допомогою імітаційної моделі WOFOST та даними Державної служби статистики України про урожайність за 2019 рік показують, що модель моніторингу здатна давати достатньо точні прогнози урожайності. Встановлено, що врожайність значною мірою визначається розвитком рослин у перші 3 місяці після сходу, підкреслюючи важливість моніторингу в цей період. Отримані практичні результати підкреслюють потенціал та обмеження використання інформаційної технології моніторингу врожайності з використанням геоінформаційних систем.

Основні положення та результати дослідження впроваджено та застосовано в діяльності Yancheng Polytechnic College.

Ключові слова: моніторинг врожайності, нейромережа, машинне навчання, GIS, управління проєктами, біомоніторинг, інформаційний менеджмент, критична інфраструктура.

CONTENT

INTRODUCTION

As the world population grows, so does the need for food. Efficient agriculture can produce more food per unit area, providing food security for a growing number of people. Increasing efficiency allows you to reduce the use of water and chemical fertilizers, reducing the burden on the environment. Economic growth: Increasing agricultural productivity stimulates economic growth. Increasing the efficiency of agriculture is a key factor for achieving sustainable development, solving economic challenges and ensuring the well-being of the population at the global level.

Yield monitoring science plays a critical role in improving agricultural efficiency Yield monitoring allows farmers to understand how various factors, such as soil type, fertilizer application and water availability, affect crop productivity. This allows you to optimize the use of resources, reducing costs and environmental impact. Yield monitoring helps develop accurate forecasting models that can predict food production, aiding planning at national and global levels to prevent food crises. Given global climate change, yield monitoring is key in determining the impact of weather conditions on agricultural productivity and developing adaptation and mitigation strategies. Research contributes to the development and implementation of the latest technologies in agriculture, such as remote sensing, automated data collection, and precision agriculture, all of which together increase production efficiency.

Thus, scientific research on yield monitoring not only contributes to increasing the productivity and efficiency of agricultural practices, but also plays a key role in ensuring food security, sustainable development and economic well-being at the global level.

Connection of work with scientific programs, plans, topics. The dissertation work was carried out in accordance with the research plan of Taras Shevchenko National University of Kyiv within the framework of the topic "Information technologies of analysis and forecasting of processes, invariant to the subject area", No. 0123U101621. The applicant conducted data collection and processing.

This work solves an important task, namely: the development of information technology, which includes mathematical models, methods and procedures for monitoring productivity based on geoinformation data (scientific component), as well as the development of an information system for monitoring productivity, which allows to automate the collection, processing and use geoinformation data for yield forecasting (practical component).

The object of the study is yield monitoring process.

The subject of the study is models, methods and information technology of yield monitoring based on interaction with Geographic Information Systems.

Research methods. The conducted research is based on the methods of system analysis, technical analysis, artificial intelligence, big data processing and object-oriented programming.

The study aim is to develop models, methods and data processing procedures necessary for yield monitoring.

That achieve the goal, the following tasks must be solved:

1. Conduct a critical analysis of scientific research in the fields of yield monitoring and the capabilities of geoinformation systems.

2. Formalize the task of yield monitoring using geoinformation systems.

3. Analyze spatial data representation models in GIS using raster and vector models.

5. Build Conceptual Research Model of developing the geographical information system for agriculture.

6. Develop mathematical models to reflect the relationship between phenological indicators and the yield of agricultural crops.

7. Using the methods of artificial intelligence and machine learning, develop a mathematical model for the integration of data from multispectral images and soil information for yield prediction.

8. Develop an information system for monitoring productivity that uses GIS data to improve the efficiency of the agricultural sector.

9. To master the information system for monitoring the yield of agricultural crops using the methods of comparative analysis and simulation modeling.

Scientific novelty of the obtained results:

- *For the first time*, an integration model of artificial intelligence for yield monitoring has been developed, based on the combination of multispectral images and geoinformation data. This model integrates Convolutional Neural Networks and Recurrent Neural Networks to enhance the accuracy and sensitivity of monitoring.
- The mathematical model of the relationship between phenological indicators and crop yields *has been improved*. Unlike other models, this combined model includes an adaptive threshold method for determining the membership of crop pixels and identifying the trend and seasonal components of the phenological indicators' time series. This improvement enhances the accuracy of forecasting.
- The information technology for yield monitoring based on geoinformation data *has been improved*. The enhancement involves the use of a combined model of the relationship between phenological indicators and crop yields, as well as an integration model of artificial intelligence. Unlike other technologies, the developed technology takes into account a wider variety of data, which simplifies integration with Geographic Information Systems (GIS).
- The methods for representing and storing geoinformation data *have been further developed* in terms of correlating key properties of agricultural objects with aerospace images of the locality.
- Information technologies for project management *have been further developed* in the aspects of monitoring and forecasting yields and integration with geographic information systems.

In the first chapter, an analysis of scientific literature was carried out, as a result of which it was established that the use of digital images of geographical territories and the development of geoinformation systems are becoming key in modern agriculture. This contributes to the effective management of cultivated areas, analysis and forecasting of yields, especially in the context of the growing need for food against the background of the increase in the world population. The development of technologies provides new opportunities for intensification and optimization of the agricultural sector**.** It has also been established that neural networks are effective tools for predicting yield. They are capable of modeling complex non-linear relationships in agronomic data, as well as processing satellite images and remote sensing data.

It was found that none of the services or software combines all the necessary capabilities for data processing and analysis, which are necessary for the management of sown areas: finding anomalies, identifying phenological changes, and estimating yield. It is shown that an important task is the creation of specialized software that would allow downloading and working with archives of large volumes of images and would have built-in methods of intelligent data processing and pattern recognition.

The second chapter describes the use of aerial imagery and time series analysis of images to determine phenological indicators and other important indicators of plant growth and health. The importance of using the NDVI index as a key indicator for evaluating vegetation cover and yield is emphasized.

The approach to data processing and analysis in the context of geographic information systems is considered. A large amount of geographic data requires reduction for processing by means of computer technology. The chapter also covers the structure and organization of geodata in vector and raster formats, revealing their unique capabilities and limitations for representing and analyzing geographic information models.

A conceptual model of the research of the geoinformation system for agricultural monitoring has also been built. The development of the system is divided into four stages: definition of the goal, description of functionality, implementation and diagnostics. Each stage includes steps that contribute to the creation of an effective system for monitoring and managing crop yields. The importance of a systematic approach to the creation and use of geoinformation systems in agriculture has been established.

In the third chapter, a mathematical model of the relationship between phenological indicators and crop yield and biomonitoring are described, which considers multispectral images of the field to predict yield in dynamics. Separation of phenological indicators into trend, seasonal and random components is aimed at ensuring effective yield monitoring. The model includes steps of image binarization to determine areas with crops using a threshold function and the Otsu method to select the optimal threshold value.

The process of creating and training a hybrid neural network that integrates image data and soil information for yield prediction is also described. The network architecture includes convolutional neural networks (CNNs) for image processing and fully connected layers for ground data analysis. This integration allows the network to consider a variety of information, increasing its ability to accurately predict yield. In the second stage, the network uses recurrent neural networks to analyze data sequences, which adds the ability to take into account temporal dependencies and context.

The fourth chapter describes the development of the yield monitoring information system, which uses GIS data to improve the efficiency of the agricultural sector. The modular structure of the system is described, including modules for data collection, storage, processing, visualization and analysis, including the use of machine learning for forecasting and optimization of processes.

The algorithm for implementing an integration model of artificial intelligence for yield monitoring based on a combination of multispectral images and geoinformation data is described, which includes seven stages: data collection and preparation, development of a neural network, its testing and validation, optimization, implementation in agricultural systems and further analysis of the results.

Practical significance of the obtained results confirms that the developed methods, models, and information technology for crop yield monitoring and integration with geographic information systems are an important step in developing a theoretical and practical basis for ensuring sustainability and profitability of agricultural companies. The developed toolkit is of practical importance for farming companies, agroholdings, and agriculture as a whole. In the long term, the use of developed methods and models will positively impact the development of the agricultural sector in the People's Republic of China and Ukraine. The main scientific provisions of the thesis are brought to the level of methodical generalizations and applied tools, which makes it possible to monitor yield.

The information system for monitoring the yield of agricultural crops was validated by the methods of comparative analysis. Comparison of winter wheat, corn, and barley yield forecasts in Chernihiv region for 2019 from forecasting using the WOFOST simulation model and data from the State Statistics Service of Ukraine on yield for 2019 show that the monitoring model is able to provide fairly accurate yield forecasts. It was found that yield is largely determined by plant development in the first 3 months after emergence, emphasizing the importance of monitoring during this period. The obtained practical results emphasize the potential and limitations of the use of yield monitoring information technology using geoinformation systems.

The main provisions and results of the research have been implemented and applied in Yancheng's operations Polytechnic College.

Personal contribution of the acquirer. The applicant personally received the main provisions and results of the dissertation work. Digital is presented in work [1]. image analysis technologies for decision support systems in agriculture the personal contribution of the acquirer consists in the analysis of advantages and disadvantages of geographic information systems. Monitoring is considered in the individual work [2. and forecasting tools of the crop yield. In the work [3], it is conceptual model of geographic information system for agriculture The personal contribution of the recipient consists in the formalization of the three main tasks of the geoinformation system for agriculture. The development of yield is proposed in the individual work [4]. monitoring model based he analysis of surveys and images of the field by combining mathematical models of the relationship between phenological indicators and agricultural productivity. In a solo work [5], the acquirer suggested use of geoinformation systems for agricultural problems.

The materials from international conferences were also published, in which the provisions of the dissertation work are revealed in more details [6-11].

Approval of the results of the dissertation. The main results of the work were reported , discussed , and received a positive evaluation at IEEE conference " Smart Information Systems and Technologies" (SIST-2021), Astana , Republic of Kazakhstan , as well as at international conferences " Information technologies and interactions ", Kyiv (2018, 2019, 2021), "Project Management in the Development of Society ", Kyiv (2019), " Information Modeling Technologies, Systems and Complexes ", Chernivtsi (2019), " Technology Development Management ", Kyiv (2020).

Publications. Based he the dissertation materials, 11 scientific works have been published, including: 4 scientific articles in specialized publications of Ukraine, 1 article in a publication that is not included in the letter of the Ministry of Education and Culture, 6 materials of international conferences. The main results of the work were obtained by the author personally. Of the scientific works published in co-authorship, the dissertation research describes those provisions resulting from the author's staff work.

Structure and scope of work. The dissertation consists of an introduction, four chapters, chapter conclusions, main conclusions, a list of references and appendices. The total volume of the dissertation is 140 pages, of which the main part is 127 pages, including 26 figures, 8 tables, a bibliography of 93 titles and 2 appendices.

CHAPTER 1. THE PROBLEM OF MONITORING THE YIELD OF AGRICULTURAL CROPS

1.1. Geo-information technologies of cropland management

In recent decades, the use of digital images of geographical areas for agricultural purposes has gained relevance. The rapid development of this direction is facilitated by the development of wireless communication technologies, digital photography technologies, devices for displaying images, operating systems and services for saving and processing these images. These opportunities became a prerequisite for the intensive development of geoinformation systems and technologies in general.

The use of geoinformation systems is an important tool for crop area management, yield analysis, and future yield forecasting. By 2050 according to Food and Agriculture Organization of the world's population may grow to 9.6 billion [12]. This is a big burden on agriculture, because productivity should increase significantly in the coming years. Moreover, agriculture already uses most of the water reserves, and there is not enough arable land. That is why the relevance of the development of geoinformation technologies to support decision-making in the field of agriculture is beyond doubt [2].

It should be noted that in the last decade, the European Union actively finances a number of projects related to the informatization of agriculture. These include, in particular, Horizon 2020 research programs. In addition, many European companies are actively engaged in this direction: eCow, Connected Cow, Anemon, etc. [13]. The product of the American company ESRI called ArcGIS [14] can be called the first effective geoinformation system created for agriculture. Essentially, ArcGIS is a family of geoinformation software products used in land management, geodesy, and land management in general. ArcGIS products have a whole line of additional software modules for specific tasks, and there is also a separate ArcPad software product for portable computers. Due to its multifunctionality, the ArcGIS system can also be used for agricultural purposes, in particular, the management of sown areas to obtain the planned harvest of agricultural crops.

The paper [15] describes the hypothesis that the yield of agricultural crops can be directly determined by heterogeneities in field images. The processing of digital images of fields provides valuable information about the condition of agricultural crops, allows to assess the health of plants and predict the yield, terms, quantity and quality of the obtained products in the future.

The image of sown areas can be presented in the form of time series, which allows applying appropriate methods of their forecasting [16]. Methods and technologies for processing time series of digital images for decision-making in agriculture are described in works [17-19].

Crop yield management is part of management tasks at an agribusiness, which requires the application of new concepts of project and program management [20-22]. To analyze the yield, it is necessary to take into account external influencing factors, in particular, air quality in the area where agricultural crops are grown [23].

The use of geoinformation technologies for agricultural purposes may refer:

- collection of information on soil quality;

- collection of information on the need to apply fertilizers to specified agricultural areas;

- collection of information on plant diseases;

- collection of information on possible yield (for example, analysis of images of apple orchards during the flowering period of apple trees to estimate the yield of apple trees in the current season).

This information is necessary for effective management in order to obtain a larger harvest, forecast the price of the crop, prompt elimination of plant diseases, etc.

Today, the popular concept is based on Site Specific Crop Management (SSCM). An important task of SSCM is to increase the yield, that is, to ensure the quality and quantity of the obtained agricultural products. SSCM is implemented on the basis of GPS technology, which, through a network of satellites, determines the location of objects and the condition of plants in agricultural areas. Real Time Kinematic technology (RTK) allows you to determine objects on the area with great accuracy. After collecting information about the state of the soil, the SSCM technology allows you to manage the application of fertilizers, chemicals, herbicides, pesticides and other substances. The application of this technology is relevant not only for Ukraine, but also for China, where in some regions, farmlands produce crops several times a year. All this requires prompt management of the ripening process of agricultural crops. Equipment that uses a GPS module makes it possible to process large areas in a short time and minimizes human impact.

Operational information about the state of the field (presence of pests, plant diseases, soil condition) makes it possible to create yield maps. These maps make it possible to estimate future benefits from growing agricultural crops, because there can be significant differences in yield within the same field. This is influenced by both the condition of the soil and the slope of the field surface. Separately, soil maps can be formed and evaluated, allowing to evaluate the content of sand, clay, and peat in them. All this makes it possible to plan the places for planting plants within one field in a timely manner, as well as take into account the necessary proportions of fertilizers, which are determined differently for different types of soil. It is clear that the calculation requires not only information on the condition of the soil, but also the yield on these grants in previous periods, as well as information on the use of fertilizers in previous seasons.

To create maps of agricultural areas, drones equipped with a specialized camera can be used to create maps of the state of crops, taking into account infrared reflectance to determine the vegetation index normalized difference vegetation index (NDVI). It was found that on the basis of such pictures it is possible to assess the state of health of plants, the internal structure of a leaf, the content of chemical substances, etc.

Remote sensing of the Earth is also used for agricultural purposes. The information that can be obtained on the basis of this technology is the geophysical characteristics of the earth's surface. Also, remote sensing of the Earth allows obtaining up-to-date information on the movement of underground and surface waters. Such information can be used for irrigation and, in general, to form a comprehensive assessment of the environment for carrying out sowing operations, taking into account possible risks. However, remote sensing of the Earth is an expensive technology that depends on weather conditions, since the sensing is carried out using aircraft.

The cost of conducting remote sensing of the Earth can be reduced by the use of small aircraft that do not depend on cloudiness, because they fly significantly below the level of clouds and can quickly transmit information to the station for its operational processing [24-26].

For example, the company Ursula Agricultural uses small aerial vehicles to survey cultivated areas. With the help of such a study, areas of crops that require the use of pesticides are determined.

The works [27-30] describe the key features of sensors for assessing the yield of agricultural crops. At the same time, as a rule, productivity refers to two dimensions: a qualitative indicator and a quantitative indicator. Such sensors are Terra / Aqua MODIS, NOAA AVHRR, SPOT Vegetetion (VGT), Metop AVHRR, Meteosat MOBILE/SEVIRI, etc.

According to the calculations of the company PrecisionHawk [19], the analysis and scanning of areas of agricultural crops requires a large amount of time and resources. In addition, often the survey of areas is sometimes erroneous. To avoid errors in the analysis, long scanning is necessary - about 11 hours per 1 acre of crops. However, the use of advanced sensors and drones allows data to be scanned from 500 to 1,000 acres in less than a day.

Using drones, agricultural specialists can perform a number of important tasks related to the management of cultivated areas:

1. Measure the level of plant health, as well as determine the needs of plants in fertilizers, watering, etc.

2. To predict the yield estimate of agricultural crops that can be obtained in the current year.

3. To optimize the timing of harvesting, sowing and other activities.

4. Detect signs of plant diseases before they become inevitable.

5. Assess damages from showers, hail, wind or other adverse weather conditions.

Today, the task of automating the processes of processing digital images of aerial monitoring data is relevant, i.e. creating a geo-information surveillance system that combines aerial monitoring with data processing and analysis methods to obtain the state of cultivated areas and prompt decision-making regarding their management. This requires a clear reference to the geographical coordinates of the territory and the use of appropriate geo-information systems.

The use and development of image analysis methods, especially obtained by remote sensing of the earth's surface, is important for the tasks of monitoring the yield of agricultural crops, as well as the tasks of environmental monitoring, in particular the study of the state of the ocean, climate change, logging and the state of the forest, etc.

The urgency of the development of these methods and technologies is added by the food crisis, the consequence of which is a constant increase in the cost of food products. Crisis phenomena are observed also because in recent decades climatic changes have worsened and food shortages have increased due to a sharp increase in the number of the population. Consideration of the peculiarities of the growth of food shortages and related factors are described in the study [12]. That is why the tasks of management of cultivated areas and agriculture in general increasingly functions in conditions of uncertainty and risk, which requires the use of special research methods.

The development of geoinformation systems and technologies is associated with the improvement of imaging and image processing technologies, the appearance of high-quality maps, such as Google Maps, the growth of data transmission speed, the development of a new direction - the Internet of Things, which has a direct application in agriculture. The description of the evolution of the development of geoinformation technologies is carried out in [31].

Important tasks in the creation of geoinformation systems are:

- creating new territory maps and updating old maps;
- the task of saving and accessing digital images of maps;
- ensuring cross-platform compatibility of the system.

Geoinformation systems are integrated with:

- information collection systems;
- data storage systems;
- data processing systems;
- data display and visualization systems.

Geoinformation systems can also be part of decision support systems. Such systems are an effective tool for optimizing and automating the work of an agricultural enterprise, which allows reducing risks and uncertainty, as well as increasing profits. However, the development of analysis methods is a difficult task.

One of the classifications of geoinformation systems is as follows:

1. Geographical geoinformation systems. Such systems consist of subsystems of image processing, modeling, and analysis taking into account many factors. Such systems are used for planning, management of territories and for general tasks that require the involvement of geoinformation systems.

2. Specialized geoinformation systems. They are used for specific tasks in the field of land management, evaluation and analysis of territories, territorial management, finding a rational route, the task of extracting minerals, economic monitoring, in the field of transport, etc.

3. Tourist geoinformation systems.

4. Cadastral geoinformation systems. Such systems are used to account for land, forestry and water management.

5. Geoinformation systems for managing systems and processes. Such systems are intended for communication planning, traffic and operational management of other systems.

Geoinformation systems are based on methods whose main focus is:

1. Search, data interpolation, search by topic, data classification.
2. Location analysis.
3. Analysis of the territory.
4. The task of partitioning and finding the nearest neighbor.
5. Spatial analysis.
6. Measurements.

Functions of geographic information systems, as a rule, depend on the type of system and are determined in the process of creation. The main functions are:

- collection, presentation and preservation of digital information that reflects the spatial representation of objects;
- data conversion;
- averaging and summarization of data;
- data modeling;
- data analysis, including using additional sources of information;
- visualization of data in the form of maps, tables, graphs, diagrams.

In figure 1.1.1 the main tasks of geographic information systems in agriculture are shown.

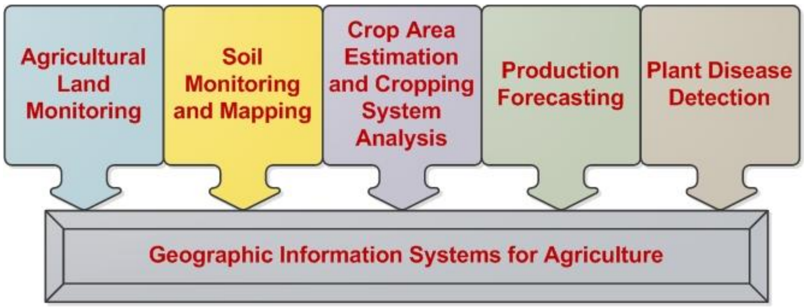


Figure 1.1.1. The main tasks of geoinformation systems in agriculture.

Agriculture is becoming smarter with the availability of modern technologies such as precision equipment, the Internet of Things (IoT), sensors and geo - positioning systems, unmanned aerial vehicles, robotics, and more. The concept of the Internet of Things in agriculture is gaining wide popularity due to the many advantages it offers. It allows farmers to collect timely geospatial information about soil plant requirements and prescribe and apply site-specific plant treatments to increase agricultural production and protect the environment. Precision farming involves high-tech tools that are more precise, economical and user-friendly. Recently Global Market Insights released a market report that estimates the global precision agriculture market will reach $10 billion by 2024 [32].

So, it can be concluded that agriculture now needs the use of new information technologies, in particular GIS and GPS for effective management of cultivated areas. The disadvantages of such systems are the high cost, the need to involve specialists in field image analysis. In addition, such technologies often do not allow determining all the necessary characteristics of the fields, which are a feature of a specific farm and specific agricultural crops that are grown.

Today, small aircraft or drones can take high-quality digital pictures. These pictures are input information for specialized software complexes or decision support systems that analyze images and give a conclusion about the condition of soils , agricultural crops, etc.

As you know, the images of cultivated areas obtained as a result of remote sensing of the earth's surface or by other methods are transformed into time series, which must be analyzed and, as a result, a rational decision should be made regarding the management of these areas. For this purpose, specialized software is needed, which at a high level can process both one-dimensional data and multidimensional time series [33]. Another shortcoming of modern geographic information systems and software packages for image processing and pattern recognition is that such systems do not provide all the necessary functions for processing time series, and also require storage of data arrays for image access capabilities, that is, they do not provide optimization for effective monitoring the condition of cultivated areas of agricultural crops [34].

WinDisp software was developed, in which the key needs of managers in agriculture were reflected and implemented. In the early 2000s, support for this software stopped, as the tools implemented in it did not meet the challenges of time and became obsolete [12]. Specialized mathematical and statistical packages and tools MatLab [35], the R environment [36], as well as image processing software GRASS [37], IDRIDI [38], ENVI [39] can be used to analyze agricultural land images for yield monitoring. MySQL or PostgreSQL spatial databases to store information can be also used. However, processing images of agricultural land requires special programming and analysis skills, that is, the involvement of relevant specialists.

The necessary tools for the analysis of ground sensing information, such as NDVI vegetation index [40], surface moisture, anomalies, etc.: MARS [41], USGS [42], GLAM [43], CropExplorer [44]. However, these platforms do not have the necessary functionality to process images of agricultural lands that meet the needs of crop area management and yield forecasting.

Some online services provide services related to image processing, but they are mostly used for specific tasks [45-47]. For example, the TIMESAT service [48] is quite effective, but its drawback is the lack of a correct graphical interface. Another service, TimeStats, has built-in tools for intelligent data analysis and a sufficient amount of archival image data, but its use requires the involvement of data analysis specialists and remote sensing experts. Means of rheological modulation [49, 50] provide effective work with time series of images, but their capabilities do not meet the needs of processing a large number of images.

According to the results of the study of modern means of processing time series of images for monitoring the yield of agricultural crops, it was found that none of the services or software combines the entire necessary package of data processing and analysis capabilities that are necessary for managing cultivated areas: finding anomalies, identifying phenological changes, yield assessment, etc. Therefore, an important task is the creation of specialized software that would allow downloading and working with archives of large volumes of images and would have implemented, built-in methods of intelligent data processing and pattern recognition.

An urgent task is the development of methods of analysis of such digital images for establishing and forecasting the yield level, identifying overdried areas soil, weed content, plant diseases, etc. Integration of these methods into geoinformation systems will allow creating a multifunctional decision support system in agriculture.

1.2. Overview of artificial intelligence models for yield forecasting

Research shows that neural networks can be effective tools for yield forecasting due to their ability to model complex non-linear relationships in data. This is especially true of studies that consider the use of convolutional and recurrent neural networks for the analysis of seasonal and temporal changes in agronomic data, the use of neural networks for the processing and analysis of satellite images and remote sensing data for yield forecasting. Research that integrates various machine learning techniques, including neural networks, to optimize resources and increase productivity. Articles comparing the performance of neural networks with traditional statistical methods or other machine learning algorithms in yield prediction tasks.

Research [51] emphasizes the importance of yield forecasting for crop market planning, crop insurance, harvest management and optimal nutrient management. Traditional yield forecasting approaches include, but are not limited to, extensive manual surveys or the use of remote sensing data. In view of the growing volume of data provided by remote sensing images, this approach is becoming more and more important for the problem of yield forecasting, which necessitates the need for more advanced methods for detecting the inherent spatio-temporal patterns of these data. Despite significant progress in the field through the use of deep learning techniques such as convolutional neural networks (CNNs), the use of convolutional long-term memory networks (ConvLSTM) for yield prediction has not been explored until now . The study proposes DeepYield, a combination framework that integrates ConvLSTM layers with a three-dimensional CNN (3DCNN) for more accurate and robust spatiotemporal feature detection. The models are trained using historical county-level yield data and MODIS land surface temperature (LST), surface reflectance (SR), and land cover (LC) data for 1,836 major soybean-growing counties in the United States. The prediction performance of the developed models is compared with competing approaches including decision trees, CNN+GP, and CNN-LSTM, and the results indicate that DeepYield significantly outperforms these techniques and also outperforms both ConvLSTM and 3DCNN.

The study [52] focused on predicting maize yield in Jilin Province, China, from 1962 to 2004, using climate conditions and fertilizer use as predictors. Multivariate linear regression (MLR) and nonlinear artificial neural network (ANN) models were used in the study. Yield was determined as a function of rainfall in July-August, rainfall in September and the amount of fertilizer applied. In the ANN model, fertilizers appeared to be the dominant predictor, which had a highly non-linear relationship with yield. Given the difficulty in obtaining fertilizer use data for corn, the study also tested the possibility of using previous year yield data instead of fertilizer data. Calculation of predictive skill scores using cross- validation and back-validation showed that the ANN models significantly outperformed the MLR and persistence models (i.e., the predicted yield is identical to the observed yield of the previous year).

Because the data were non-stationary, cross- validation proved to be less reliable than retrospective validation in assessing predictive skill. This study highlights the value of non-linear models, particularly artificial neural networks, in yield prediction, and points to the potential of using historical yield data as a proxy for fertilizer use data.

Research [53] emphasizes the complexity of yield forecasting, which is determined by many factors, including genotype, environmental conditions and their interaction. Accurate yield forecasting requires a deep understanding of the functional relationships between yield and these interrelated factors. To detect such connections, comprehensive datasets and powerful algorithms are needed.

Within Syngenta Crop Challenge 2018, Syngenta released several large datasets containing genotype and yield performance information for 2,267 corn hybrids planted in 2,247 locations between 2008 and 2016, and asked participants to predict yield performance in 2017. In the winning team 's research, a deep neural network (DNN)-based approach was developed that used advanced modeling and solution techniques. This model showed high prediction accuracy with a root mean square error (RMSE) of 12% of the mean yield and 50% of the standard deviation for the validation data set with forecasted meteorological data. With perfect weather data, the RMSE would drop to 11% of the mean yield and 46% of the standard deviation.

Feature selection was also carried out based on the trained DNN model, which made it possible to successfully reduce the dimensionality of the input space without significantly reducing the prediction accuracy. Computational results reported in the study showed that this model significantly outperformed other popular methods such as Lasso, Superficial Neural Networks (SNN), and Regression Tree (RT). The results also revealed that environmental factors have a greater effect on yield than genotype.

Paper [54] describes a study that used historical data on wheat yield and associated crop area, precipitation, and temperature. It shows that the use of statistics and artificial neural networks can provide high-quality wheat yield forecasting. However, the authors note that no comparison has been made between the results obtained from the spatial neural network and the results obtained from commonly used temporal neural network models in crop yield prediction.

The article presents the latest research results obtained using both spatial and temporal neural network models in yield forecasting. The simulation shows that the spatial neural network model is capable of predicting the yield of wheat relative to a given crop area with high accuracy compared to the temporal models NARNN and NARXNN. However, the high accuracy of the spatial NN model in yield forecasting is limited to only predicting yields within normal ranges.

The authors emphasize that users should be cautious when using NARNN or NARXNN models for yield prediction due to their inconsistency between training and prediction results.

Paper [55] focuses on accurate and timely corn yield forecasting in the US, which is the leading corn producer in the world, supplying more than 30% of global corn production. Accurate forecasting of maize yields is essential for commodity trade and global food security.

Traditional deep learning approaches generally require a large dataset for training and are prone to overtraining when the number of samples in the training set is relatively small. To address these limitations, the study developed a county-level corn yield forecasting model based on a Bayesian Neural Network (BNN) using several publicly available data sources, including time series of satellite products, serial climate observations, soil property maps, and historical records of corn yields.

Using data from previous years starting in 2001 to train the model, the BNN-developed model achieved an average coefficient of determination (R^2) of 0.77 for late-season forecasting across the entire US Corn Belt in the 2010-2019 test years and outperformed five other advanced machine learning models.

A detailed evaluation in three representative test years showed that the proposed BNN model could accurately estimate corn yield not only in normal years but also in abnormal years when extreme weather events occurred. In addition, the timeliness of the forecast was evaluated during the growing season, reaching an R of approximately 0.8 by mid-August, which is approximately two months before harvest.

Estimated uncertainty was also estimated, and more than 84% of the observed yield records were successfully included in the 95% confidence interval of the estimated yield distribution. The results also showed that the level of uncertainty steadily decreased over time and stabilized around the beginning of August. Uncertainties in yield forecasting were mainly caused by observational noise and also related to interannual and seasonal variability of environmental stress such as heat and water stress. This paper provides a solid basis for in-season yield prediction and highlights the need for a deeper understanding of the impact of environmental stress on agricultural productivity and yield estimation.

Studies [56, 57] have focused on the importance of timely and accurate monitoring of crop growth conditions and yield estimation for the agricultural industry and food security. They use different approaches to combine remotely sensed data and machine learning models to forecast wheat yields at the regional level.

Both studies integrate remotely sensed indices such as leaf area index (LAI) and vegetation temperature condition index (VTCI), which are closely related to crop growth conditions and plant water stress, into their yield estimation models.

The first study uses a back-propagation (BP) neural network and an improved BP neural network with a particle swarm optimization (IPSO) algorithm, while the second study uses a long-short-term memory (LSTM) model for yield prediction.

Both studies compare the performance of their models against other approaches, highlighting the benefits of using sophisticated machine learning models to provide more accurate yield predictions.

While the first study focuses more on the spatial aspect, using LAI and VTCI indices at different growth stages, the second study focuses on temporal relationships, using LSTM to recognize consistent patterns and relationships in the data.

Both studies highlight the importance of integrating different data sources and advanced machine learning technologies to improve yield estimation accuracy, which is key to informed agricultural management and food security.

Research [58] focused on the development of a dynamic yield forecasting system for three major grain crops (wheat, corn, and rice) in China using data for seven years (2013-2019). What is unique about this study is the use of multiple sources of environmental predictors, including climate, vegetation indices, and soil properties, to predict yield.

A random forest (RF) model was used to develop this forecasting system, which showed good performance in yield estimation of all three crops with a correlation coefficient (r) above 0.75 and normalized root mean square errors (nRMSE) below 18.0%. The results also showed that crop yield can be satisfactorily predicted one to three months before harvest.

The optimal lead time for yield prediction depended on the type of crop. In addition, the study found that the main predictors that affect yield vary between crops. In particular, for winter wheat, the main predictors were solar radiation and vegetation indices (especially at the stages from earing to milk maturity); for spring corn – vegetation indices (throughout the entire growing season) and drought (especially at the stages from emergence to cob setting); for summer corn, late and mid-season rice, the dominant predictor was soil moisture; for early rice - precipitation (especially at the stages from earing to flowering).

This study provides valuable guidance for practical yield forecasting and understanding yield response to environmental conditions across China. The methods used in this study can be easily adapted in other countries with available information on climate, soil and vegetation conditions.

Research [59] focused on the analysis of crop yields in karst areas in Guizhou Province, China, which suffers from soil degradation and low yields due to exposure of carbonate rocks. The main goal of the research is to better understand how environmental factors affect productivity in order to ensure a more sustainable use of natural resources for food production and development.

The study used four types of artificial neural networks to analyze and model spatial yield patterns for seven types of crops grown in Guizhou Province, exploring the relationships with meteorological, soil, irrigation and fertilizer factors.

The spatial classification results showed that most of the regions with high yield per unit area and total yield are located in the central-northern part of Guizhou.

The three artificial neural networks used to model spatial yield patterns all showed good correlation coefficients between simulated and actual yields. However, the backpropagation network showed the best performance in terms of accuracy and execution time.

Among the 13 investigated influencing factors, temperature (16.4%), radiation (15.3%), soil moisture (13.5%), fertilizer N (13.5%) and P (12.4%) had the greatest contribution to spatial distribution of productivity.

This study highlights the potential of applying neural networks to identify environmental controls on yield and model spatial patterns of yield, which can help local stakeholders realize sustainable development and crop production goals.

Research [60] focuses on forecasting wheat yields in China, which is key to regional trade and national food security. The main focus is on the integration of data from different sources and the application of machine learning methods to create a simple, timely and accurate crop yield forecasting model at the level of administrative units.

Developed a model framework for integrating climate data, remote sensing data, and soil data using the Google platform Earth Engine (GEE) for forecasting the yield of winter wheat.

The entire plant growth period was divided into four-time windows, and their respective predictive ability was evaluated using the main winter wheat production regions in China as an example.

Support Vector Machines (SVM), Gaussian Process Regression (GPR), and Random Forest (RF) were found to be the top three methods for yield prediction among the eight typical machine learning models tested.

The models can accurately predict yields 1-2 months before the harvest date at the county level in China with $R^2 > 0.75$ and a yield error of less than 10%.

Different agrozones and training time settings affect the accuracy of the prediction. The three models show better results when more information about the winter wheat growing season becomes available.

This study highlights the potential of using multi-source data and machine learning methods for yield prediction, which can be applied in other regions and for other crops.

Analyzing the findings from all the discussed studies, several key insights and trends can be identified in the field of crop yield forecasting using machine learning and multi-source data integration.

Successful yield forecasting requires the integration of data from various sources, including meteorological data, soil information, remote sensing data, vegetation indices, etc. This allows models to better understand and simulate the complex interactions between different factors that affect yield.

Various machine learning approaches, including neural networks, random forests, support vector machines, and long-term memory, have demonstrated their effectiveness in yield prediction. These techniques help discover complex implicit relationships in data and provide more accurate forecasting.

Consideration of crop yield in terms of spatial and temporal variables is important for understanding and forecasting. Different time windows and spatial patterns can significantly affect the accuracy of forecasts.

It is important not only to develop forecasting models, but also to carefully evaluate their accuracy and reliability using statistical metrics and validation methods. Attention should also be paid to the potential of retraining models.

Research shows that environmental factors such as weather, soil moisture, temperature and fertilizer use have a significant impact on yield. Understanding these influences can help improve forecasting models.

Taken together, these studies highlight the significant potential of applying machine learning and multi-source data analysis to improve yield forecasting, which can play a key role in improving agricultural productivity and food security.

1.3. Formulation of the task of monitoring the yield of agricultural crops based on geo-informational images

The density and health of agricultural crops can be monitored by analyzing images of the field surface that are acquired at certain time intervals. Such time series of images contain valuable information about chlorophyll activity, plant activity, and the possible presence of diseases or pests. The analysis of multispectral images of the field makes it possible to identify the necessary indicators that determine the growth of the crop, its ripening, etc. Therefore, it can be assumed that NDVI vegetation indices, or other similar indices, can be used to monitor the state of the crop and estimate the yield of agricultural crops in a local field. It should be said that productivity is also affected by other factors: soil quality, type and efficiency of management, ecological situation and weather conditions, etc.

Main tasks can be formulated as follow:

1. To describe the relationship between the values of vegetation indices, such as NDVI, and the yield of agricultural crops. To find out whether it is possible to estimate the yield quantitatively and qualitatively based on the calculation of the NDVI index.

2. To describe the factors affecting productivity, in particular the management factor, environmental factor, soil quality factor, etc.

3. To assess the possibility of using multispectral images for forecasting the yield of agricultural crops.

4. To propose an information technology for monitoring the yield of agricultural crops, which is based on the basis of the geoinformation system.

In order to complete the tasks, you need to answer the following questions:

1. Is there a correlation between the values of the vegetation index N DVI and the yield of agricultural crops?

2. Can the value of NDVI indicate the quality of crop management, land quality or ecological situation in some local areas?

NDVI, is it possible to predict the yield of agricultural crops and calculate the potential profitability from the realization of a potential harvest?

4. What should be the market value of the crop to cover the cost of seed, management, and water and human resources expended in the cultivation process?

5. What should be the functionality of the information system for monitoring the yield of agricultural crops?

So, on the basis of the set goals, the following hypotheses of the dissertation research can be formed:

1. There is a strong correlation between NDVI index values and crop yields. Formally, this means the presence of a functional dependency$g$ between yield V and NDVI index:

$$V = g(NDVI). \tag{1.3.1}$$

2. There is a relationship between the quality of land resources, management efficiency, other possible influencing factors and the NDVI index:

$$V = g(NDVI, Z, M), \tag{1.3.2}$$

where Z are indicators of the quality of land resources, M are indicators of management efficiency.

3. Multispectral field images can be used to predict yield. Spectral data is an integration of all factors affecting crop growth.

To predict the yield of a field on the basis of multispectral images, in particular, to identify quantitative and qualitative indicators of yield, possible plant diseases, etc., it is necessary to have a knowledge base with reference images that correspond to certain growth indicators of specified agricultural plants. By comparing the current image with images from the knowledge base to establish similarities, it is possible to draw a conclusion about:

- possible deviations from normal growth and ripening of the crop. In the event of a negative deviation, it requires the necessary solutions to correct the situation.
- quantitatively calculate the yield forecast.

However, this conclusion must be adjusted taking into account other factors that affect the growth and maturation of plants.

Therefore, the promising tasks of the research are:

1. Basic research of known methods of digital image analysis.
2. Develop new and modify existing methods of digital image analysis. The developed methods must take into account the needs of the agricultural enterprise and determine the maximum number of indicators for growing agricultural crops with maximum efficiency.
3. Integrate the developed methods into the geoinformation system, which automates the processes of operational decision-making in agriculture and monitors the yield of agricultural crops.
4. Verification of the developed system and methods of processing digital images in the work of a real agricultural enterprise.

Conclusions to chapter 1

The theoretical basis of the research was created to solve the task, which consists in creating an information technology for monitoring the yield of agricultural crops based on the analysis of multispectral images obtained by remote sensing or by other methods. The geoinformation system created on the basis of this technology should monitor and forecast yield by analyzing time series of satellite images to identify quantitative and qualitative indicators of yield, possible plant diseases, etc. This task is especially relevant in conditions of environmental uncertainty.

Hypotheses and main tasks of the dissertation research are defined. One of the claims that is planned to be tested is the claim that multispectral field images can be used to predict yield.

It was revealed that as a result of the food crisis, the consequence of which is a constant increase in the cost of food products, agriculture is increasingly operating in conditions of uncertainty and risk, which requires the use of special research methods.

According to the results, it was found that none of the services or software combines all the necessary data processing and analysis capabilities that are necessary for the management of sown areas: finding anomalies, identifying phenological changes, estimating yield, etc. An important task is the creation of specialized software that would allow downloading and working with archives of large volumes of images and would have built-in methods of intelligent data processing and pattern recognition.

CHAPTER 2. CONCEPTUAL MODEL OF RESEARCH DEVELOPMENT OF INFORMATION TECHNOLOGY FOR YIELD MONITORING

2.1. Analysis of time series of images for the tasks of monitoring the yield of agricultural crops

The system, which is designed to monitor the yield of agricultural crops, should take into account a large number of different growth indicators, perform statistical analysis and forecast the yield based on the monitoring results:

1. Meteorological indicators (monitoring of weather conditions).
2. Agrometeorological indicators (monitoring of crops).
3. Phenological indicators (phenological monitoring).

To ensure effective monitoring at all levels, it is necessary to provide for analysis high-quality aerial images that can be stored and analyzed in the form of time series.

Under yield monitoring, we will understand the system of monitoring and measuring the state of growth of agricultural crops, taking into account meteorological, agrometeorological, phenological and other indicators based on the analysis of time series images obtained as a result of photographing cultivated areas, with the aim of evaluating and forecasting the potential crop yield. It is also supposed to take into account the unevenness of the yield on different plots of sown areas, which can be visualized with the help of yield maps, on which the amount of products obtained after harvesting is marked in different colors. It should also be taken into account that yield monitoring should involve a two-dimensional assessment: the amount of products that will be potentially obtained and the quality of these products on the appropriate scale.

Phenological indicators of plant growth take into account the dependence between seasonal changes, weather conditions and climatic changes that affect the timing of crop ripening, the occurrence of diseases, the timing of the start of certain agricultural works, etc. The study of phenological indicators is a necessary element of quality management of sown areas.

Most of the studies related to the selection of phenological indicators in the analysis of images of cultivated areas by means of remote sensing of the earth's surface consist in the calculation of the normalized differential vegetation index (NDVI). This is a simple quantitative indicator of photosynthetically active biomass used to quantify plant cover:

$$NDVI = \frac{N-R}{N+R}, \tag{2.1.1}$$

where R is 630-690 nm, the visible red region of the spectrum, N is 760-900 nm, the reflective infrared region of the spectrum.

According to this formula, the density of vegetation (NDVI) at a certain point of the image is equal to the difference in the intensities of reflected light in the visible and infrared ranges, divided by the sum of their intensities [61, 62]. The general concept of detecting phenological indicators is based on the determination of critical points in the NDVI trajectory. Let

$$\beta = \{\beta_1, \beta_2, \dots, \beta_n\} \tag{2.1.2}$$

is a discrete time series of estimates NDVI recorded at specified intervals of time (day, week, two weeks, etc.) is $\beta_i$an estimate of NDVI at the i- th moment of time. Then, by plotting this time series on a graph, you can study how the vegetation index changes in dynamics, which allows you to track sudden changes in the level of vegetation in real time and make an appropriate decision to ensure effective crop management.

Consider the Breaks For Additive Seasonal and Trend method (BFAST), which can be used to identify long-term phenological changes in image time series. This method combines methods of detecting changes in the behavior of time series with methods of decomposition of series into components that determine trend changes, seasonal changes, and random components [63]. The work [64] describes the capabilities of the BFAST method to detect long-term phenological changes using a harmonic seasonal model. According to this method, the additive decomposition model of the time series of the image has the form:

$$Y_t = C_t + Q_t + r_t, \tag{2.1.3}$$

where $Y_t$is the time series data recorded at time t, $T_t$is the trend component, $S_t$is the seasonal component, $e_t$is the residual, random components, $t = \overline{1,n}$, and n is the number of observations or the number of elements of the image time series. The residual components are variations of the time series that deviate from the trend or seasonal components. In this model, it is assumed that the trend component is piecewise -linear), which means its assignment in the form:

$$C_t = a_i + t \cdot b_i, \tag{2.1.4}$$

where $i = \overline{1,m}$ are observation control points.

The seasonal component changes similarly. Changes can occur from one segment to another between control points. Also, for the seasonal component, control points can be defined differently. In [64], a harmonic model was defined for describing the seasonal component:

$$S_t = \sum_{k=1}^{K} \left(\alpha_{jk} \, sin(2\pi kt) + \beta_{jk} \, cos(2\pi kt) \right), \tag{2.1.5}$$

where α_{jk}, $\beta_{jk}$are model coefficients.

The described model has the following advantages, in contrast to the usual seasonal model:

1. The model is less sensitive to short-term changes and noise.

2. Many observations are not required to calculate the parameters of a multiple regression model.

3. The parameters $\alpha_{jk}$are β_{jk}calculated relatively easily.

In addition to highlighting phenological changes, methods that allow identification of meteorological and agrometeorological changes can be considered. Their content generally does not change. It is necessary to select time series of data (meteorological and agrometeorological indicators) that characterize the research area, carry out its decomposition, select seasonal and trend components, as well as evaluate the residual component. After carrying out this procedure, it is possible to calculate the forecast of both the seasonal component $\hat{Q}_t$ and directly the time series of data with a certain confidence interval $\hat{Y}_t$. The received forecast data is valuable information for planning and organizing agricultural work, in terms of monitoring the yield of agricultural crops. Two pictures of the same field are shown on figure 2.1.1 with an interval of 10 days.

Figure 2.1.1. Satellite images of the same field taken at 10-day intervals

Changes in the image are visualized without the use of software, only by viewing the image. However, for effective crop management, such undesirable changes must be identified before they become visible to the naked eye to take the necessary action and not lose the crop.

To highlight phenological, meteorological and agrometeorological indicators, it is necessary to build a so-called satellite time series of Satellite Image Time Series (SITS). Such time series are satellite images of a certain area taken at fixed moments in time. To build it, you need:

1. Take photographs of the territory with a certain time interval.

2. Perform recognition of changes in these photos using special image recognition methods.

3. Visualize a map of the territory, on which to display (preferably in color) the changes associated with the investigated feature (Fig. 2.1.2).

The collected information is analyzed and a conclusion is formed regarding the assessment of the yield of agricultural crops in the specified area.

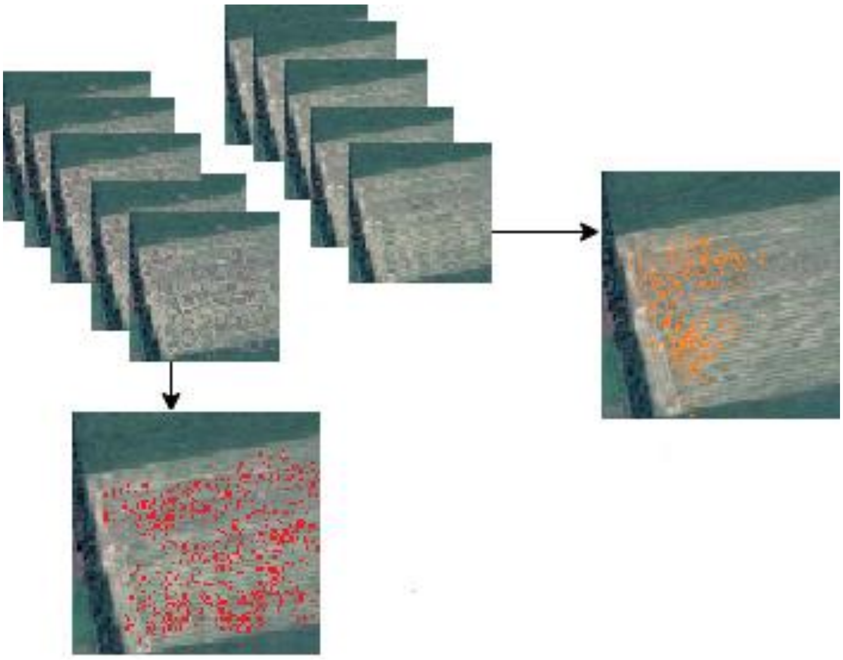


Figure 2.1.2. Satellite time series of images of the territory and color display of deviations from the norm by two indicators

It should be noted that an arbitrary image can be defined as some function $z(x, y)$, where x defines the abscissa coordinate, and y defines the ordinate coordinate on the plane. The value of the function $z(x, y)$is the intensity. In the case of a black and white image, the value of the function $z(x, y)$is the gray level. A digital image is a function $z(x, y)$whose arguments take a finite number of discrete values. If you divide the image into a finite number of pixels, the position of each pixel will be determined by x, y coordinates.

As a rule, black and white images are not used for yield monitoring, because color contains very valuable information, on the basis of which you can determine the NDVI indicator or identify possible plant diseases.

Color models mostly have three representations [65]:

1. RGB color mixing model.
2. CMYK model used in typography.
3. YIQ, YUV, YCrCb models used in video systems.

One of the ways to compare images with each other is the method of perceptual hashing. One of the simplest hash functions displays the average value of the low frequencies. In images, high frequencies provide detail, while low frequencies show structure. A large, detailed photo contains many high frequencies. There are no details in the very small image, so it consists entirely of low frequencies. The hashing algorithm consists of the following steps:

1. Reducing the image size. When reducing the image, high frequencies practically disappear. It is recommended to reduce the image to at least 8 by 8 pixels. Even if the image is not square, it still needs to be compressed to a square of minimum size.
2. The next step is to convert the image to shades of gray without color.
3. Calculation of the average value of all shades of all image pixels.
4. If the shade of the pixel is darker than the average shade, then it receives the value 0, i.e. black. If the shade of the image pixel is lighter than the average shade, then it gets the value 1, i.e. white. Thus, the image is converted to black and white.
5. Construction of a hash, i.e. conversion of image bits into one value. In the case of an 8 by 8 image, this would be a 64-bit value in pixels or bits.

The final hash will not change if the image is scaled, compressed or stretched. To compare images, similarity can be quantified using, for example, Hamming distance.

Another method is perceptual image hashing, which uses the discrete cosine transform of the image. The discrete cosine transform (DCT) is an orthogonal transform, a variant of the cosine transforms for a vector of real numbers. It is used in lossy compression algorithms, such as MPEG and JPEG. This transformation is closely related to the discrete Fourier transform and is a homomorphism of its vector space. Mathematically, the transformation can be performed by multiplying the vector by the transformation matrix. At the same time, the inverse transformation matrix is equal to the transposed matrix [66].

perceptual hashing algorithm consists of the following steps:

1. Reducing the image size. A perceptual hash is calculated based on a small image, such as 32 by 32.

2. Remove the color and perform a discrete cosine transformation, which divides the image into a set of frequencies and vectors.

3. In the upper left corner of the image obtained after the cosine transformation, only lower frequencies will be stored, so it is suggested to fix a part of the image for further analysis: the upper left corner, measuring 8 by 8.

4. Calculate the average value in the reduced block.

5. Assign 0 or 1 to the cosine transformation image, depending on whether the shade is darker or lighter.

6. After that, it is necessary to construct one perceptual hash value and compare it with the hashes of other images.

There are variations of the perceptual hashing algorithm that also improve performance. For example, images can be cropped before resizing. In this case, the loss of information around the main part of the image does not play a special role.

The disadvantage of the hashing method for the given task is the need to take into account high frequencies as much as possible. Also, the disadvantage is that the change in the color of the pixels of the image will not significantly affect the construction of the hash, and this change can signal negative processes occurring in the development of plants.

Another approach that can be useful when analyzing crop field images is image segmentation. Segmentation involves dividing an image into segments according to certain rules. Segmentation methods, taking into account high frequencies, make it possible to detect changes that have taken place on the site, new objects that have appeared on it, etc. Each pixel of the image is assigned a label, and it is assumed that pixels with the same label are visualized in the same way.

There are two main approaches to segmentation. The first is based on the division of the image by brightness levels. A sharp change in brightness when moving from one pixel to another can indicate the identity of the object. The second class of methods consists in partitioning the image into regions according to predefined criteria.

One of the missions of the research satellite EOS AM-1, which operates under the direction of NASA, is to photograph the Earth's surface in 36 spectral bands with a length from 0.4 μm to 14.4 μm and an extension from 250 m to 1 km. This mission is called Moderate Resolution Imaging Spectroradiometer (MODIS) and is generally intended for the study of processes (radiation, meteorological, etc. on the surface of the earth and oceans). This tool is also used in agriculture, in particular for calculating vegetation indices and soil moisture level.

The MODIS sensor has 36 spectral bands, seven of which are intended for the study of vegetation and the earth's surface: blue (459–479 nm), green (545–565 nm), red (620–670 nm), infrared (NIR1: 841–875 nm; NIR2: 1230–1250 nm), and short-wave infrared (SWIR1: 1628–1652 nm, SWIR2: 2105–2155 nm). MODIS Land Team Science provides a suite of standard MODIS products for users, including an 8-day composite MODIS Surface product Reflectance (MOD09A1). According to satellite data, vegetation indices are calculated using the following formulas:

$$NDVI = \frac{R_{NIR1} - R_{RED}}{R_{NIR1} + R_{RED}}, \qquad (2.1.6)$$

where NDVI is Normalized Difference Vegetation Index, R_{NIR1}– infrared band, R_{RED}– infrared band.

$$LWSI = \frac{R_{NIR1} - R_{SWIR}}{R_{NIR1} + R_{SWIR}} \qquad (2.1.7)$$

where LSWI is Land Surface Water Index, R_{NIR1}– infrared band, R_{SWIR}– short-wave infrared band,

$$EVI = \frac{5}{2} \cdot \frac{R_{NIR1} - R_{RED}}{1 + R_{NIR1} + 6R_{RED} - \frac{15}{2}R_{BLUE}}, \qquad (2.1.8)$$

where EVI is Enhanced Vegetation Index, R_{NIR1}– infrared band, R_{SWIR}– short-wave infrared band, R_{BLUE}– blue band [67].

So far, vegetation indices, and especially the NDVI index, are the most informative indicators of automated yield assessment. However, its use for this purpose has a number of peculiarities. First, all studies on the use of the NDVI index for yield estimation, as a rule, refer to local areas, a defined list of agricultural crops and were carried out at the regional level, which does not cover a sufficient representation of plants, regions and geographical features. Often, these studies use low-resolution images for analysis. Therefore, the results of studies in which the NDVI index is proposed as the only possible index in the calculation of crop yield estimates are not entirely correct. Since the quality of the images is low, weeds and other vegetation that cannot be filtered from the original image can be perceived as cultivated plants. The obtained results are then adjusted taking into account the final harvest.

In general, most studies indicate a high correlation between NDVI index data and the volume of the final harvest. However, these studies analyzed small areas and many of them used images obtained from ground platforms.

To monitor the yield of agricultural products, manufacturers offer many devices. One of them is Grain yield monitor (GYM). This is a device with sensors for calculating grain yield placed on the combine. GYM is part of precision agriculture, providing agricultural producers with tools to reduce costs, increase yields and improve efficiency. GYM is designed to measure grain harvest mass flow, moisture and speed to determine the total amount of harvested grain. In most cases this is now combined with a Global Positioning System to record yield and other variable field information. This enables the creation of a grain yield map that provides information on spatial variability and supports management decisions for producers.

A yield monitor is a device that records data that determines grain yield. Modern yield monitors provide operators with a user interface that displays grain yield, grain moisture, as well as a color-coded spatial map that displays grain yield of defined parts of the field [67].

Grain yield maps can be displayed on a monitor or through spatial data management software such as SMS or Apex. Yield maps are used in management decisions, such as fertilizer rates and seeding rates, etc. [68]. Yield maps are also used to make decisions about best management practices in terms of comparing crop varieties, fertilizer types and application rates, and pesticide use. These and other precision farming methods can be recorded as spatial maps and overlaid on grain yield maps for further analysis and decision making. For example, fig. 2.1.3. shows a yield map of a soybean field generated by Ag software Leader SMS.

Research on the relationship between remote sensing data and yield is relevant, as not many studies have been conducted in this direction. Research in this direction is especially important due to significant ecological uncertainty, which complicates decision-making on the management of cultivated areas.

As a result of the analysis of scientific research in the direction of establishing the relationship between yield and the results of the calculation of vegetation indices and other indicators of the growth activity of agricultural plants, it was found that most of them were performed either on local areas taking into account a large number of controlled parameters, or on large areas with a significant degree of generalization, as some fragments of the image in this case are difficult to study in detail.

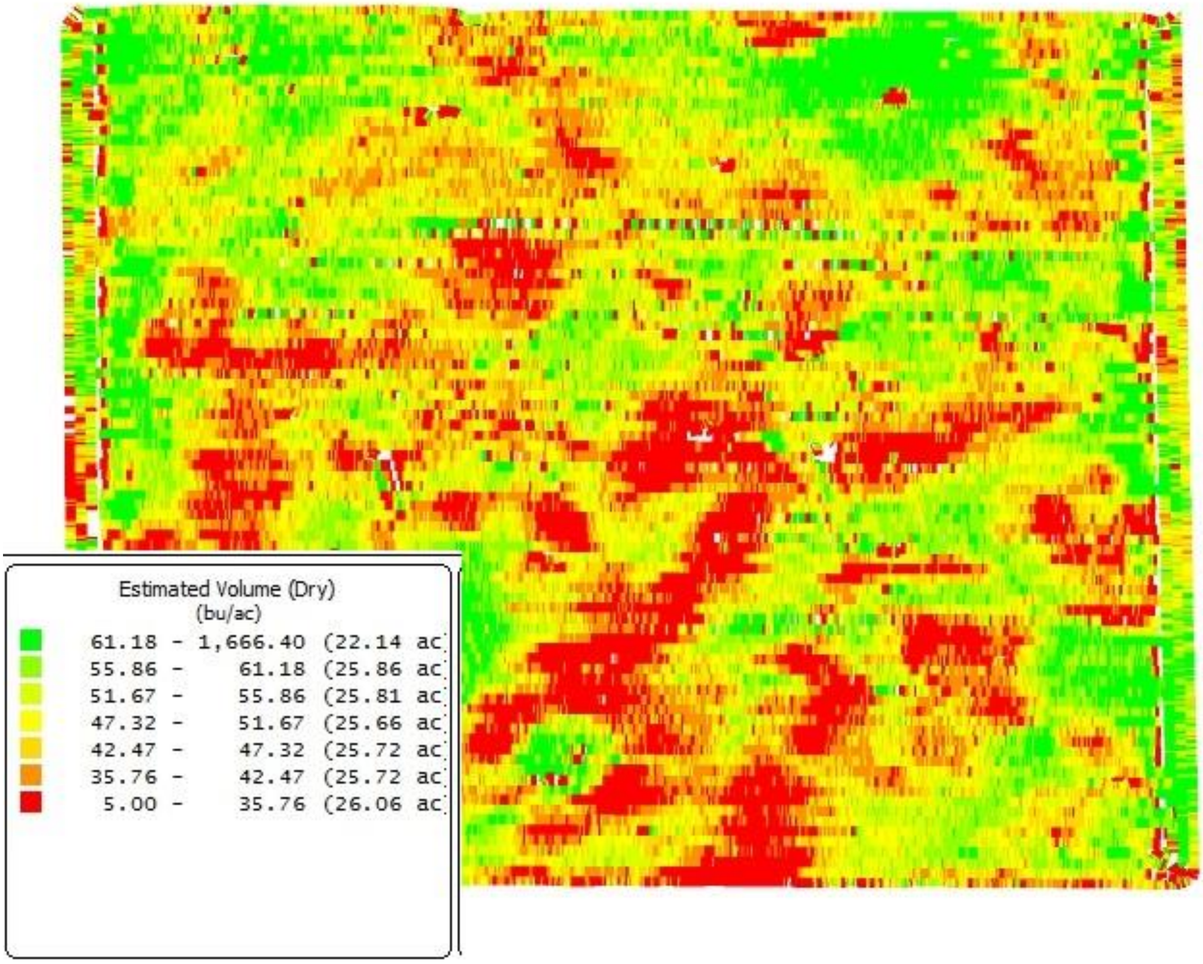


Figure 2.1.3. Soybean field yield map generated by Ag software Leader SMS.

It should be noted that the growth of agricultural plants is influenced by many factors, including soil quality, management quality, ecological situation, etc.

2.2. Representation of data in geoinformation systems

In the eternal search for harmony with the natural world, we seek a deeper understanding of the processes that occur in the universe, and the recording of these processes in various forms we call data. These data are nothing more than recorded evidence of phenomena occurring in the world around us.

In a geographical context, phenomena and phenomena appear in their infinite complexity and diversity, reminding us of the limitlessness of the natural world. Approaching any geographical object, we discover more of its details, and such immediacy of observation emphasizes that to fully understand the essence of the real world, we would need infinitely voluminous databases. But, since we have to adapt these data for processing with the help of computer technology, it is necessary to reduce them to acceptable sizes.

In the field of geographic information systems, the creation of a database can absorb a significant part of project resources. However, such a database is not just a repository of information; it is structured in a special way to optimally work with spatial data that can be represented through vector or raster models. Thus, the database in GIS is transformed into a unique representation or model of the real world, adapted for a specific spatial application.

Thus, spatially distributed data can be integrated into a database using vector or raster models. Each of these models provides a unique approach to the representation and analysis of geographic data, opening new opportunities for a deeper understanding of the complex interaction between various elements of the natural world.

In the field of raster modeling, a structure in the form of a matrix or grid of pixels serves as the basis for saving visual information. The determination of the spatial position of each pixel in the raster image is carried out by using one of its corners or the centroid. Image quality is directly proportional to the size of these pixels, each of which is associated with certain discrete attributes. Raster processing in geoinformation systems allows you to represent natural phenomena through individual elements of this matrix, where each pixel is a fundamental unit of information, the size of which can vary widely depending on the needs of the application.

The use of the raster model in geoinformation systems is manifested in such areas as processing of remote sensing data or the creation of digital terrain models.

On the other hand, the vector model manages spatial data through graphical objects – points, lines and contours associated with attribute data. The coordinates of these elements can be determined both in Cartesian and geographic coordinate systems, which allows for the transfer of topological relationships between objects.

Compared to vector models, raster models have a simpler data structure and spatial analysis methods, but their effectiveness is limited by memory requirements and the accuracy of detail representation. However, the flexibility of both models allows mutual conversion of data between raster and vector formats.

In geographic information systems, attribute data provide a link between a symbol and its geographic meaning, regardless of whether this symbol is an element of a raster matrix or a graphic object in a vector model. This relationship is implemented through a unique spatial object identifier, allowing the integration of non-spatial data that may exist in a variety of database formats.

In the depths of knowledge related to geographic information systems, we discover sophisticated mechanisms that ensure not only the preservation of coordinates and attributes of a geographic entity, but also allow studying the relationships between geographic objects. In this network of spatial data, we observe relationships defined by distance, proximity, neighborhood, as well as more complex interactions, such as internal or external location and intersection of objects.

The essence of geographic information is embodied through spatial objects associated with attributive data, which serve as the basis for data models, establishing rules for structuring both spatial and non-spatial attributes. In our practice, we prefer both vector and raster models to solve problems, depending on the specific requirements of the project, both types of models support attributive data that can be structured in the format of "flat" files, as well as hierarchical, network or relational databases data

In GIS, data is divided into spatial and non-spatial categories, and the boundary between them is often blurred because of their close interaction. Spatial objects are classified by complexity into elementary, composite, and complex, each of which has its own description structure that reflects their semantic and graphical attributes and the nature of interaction with other objects, providing a deep understanding of their nature and relationships in the geospatial context.

In the wisdom drawn from the depths of geoinformation knowledge, we consider the selection of entities to be represented in the model by means of point objects as a process that depends on the scale of the cartographic image and the characteristics of the area under investigation. On small-scale maps, settlements are represented by dots, while on large-scale maps they are revealed as site objects. The following criteria are decisive for displaying an entity as a point object: the importance of spatial location, the insignificance of metric dimensions, and the disproportionality of the size of the object to the scale of the model.

Point objects serve as the most elementary representatives of spatial entities. The coordinates of each point are embodied in the form of additional columns in the database, with each row representing a separate point, with all information about it recorded within this row.

Line objects are used to express entities that have a length but no width. These objects form network-like structures that may include, for example, transport routes or natural river systems. Nodes in such networks serve as critical points connecting lines or arcs and have attributes that determine their topological relationships.

In this intellectual journey through geoinformation spaces, we also note the more complex inter-location of entities, such as overlapping zones, internal or external placement of objects, and the possibility of objects crossing each other, where each layer of information can be transformed to achieve a deeper understanding of the interactions in geospatial context.

Contour boundaries can represent various natural phenomena, such as lakes, forests, large settlements, reservoirs, etc.

1. The essence is isolated zones that may overlap. Any point can be inside any number of objects. Objects may not completely cover the studied area.

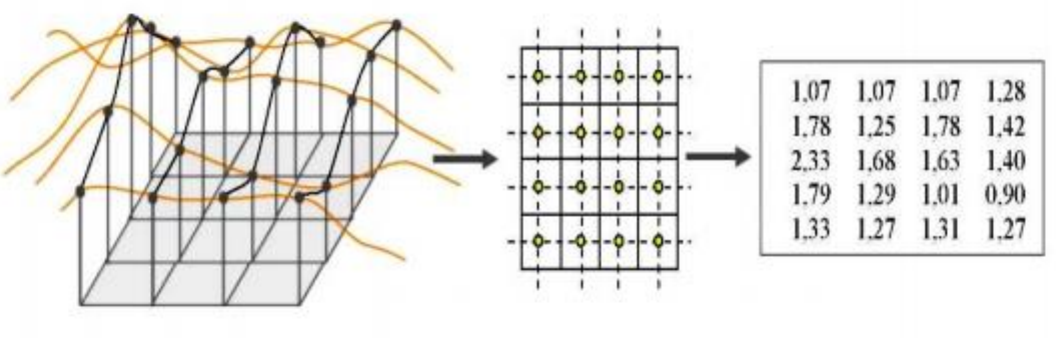
2. Any point must be in the middle of one object. The objects completely cover the studied area. Each boundary line separates two site objects. Site objects cannot intersect.

3. Any layer of the first type can be converted to a layer of the second type: each site object can now have any number of attributes.

Site objects can have “holes” that have a set of attributes different from the attributes of the main object. For example, there are islands on rivers.

Some entity cannot be accurately represented as discrete points, lines, or regions. Some essence is continuously changing in space. Therefore, there are objects that are best represented in GIS by continuous surfaces. Examples of continuous surfaces: relief. surfaces are represented as points, lines and areas.

The representation of surfaces in the form of points is called a digital model of the terrain and is based on sampling at regular intervals values from the studied surface. The result is a matrix of values, also called a raster, a grid, a grid. Many digital terrain models are created in this way and can simply be converted to a raster image for visualization (Fig. 2.2.1).



1,07	1,07	1,07	1,28
1,78	1,25	1,78	1,42
2,33	1,68	1,63	1,40
1,79	1,29	1,01	0.90
1,33	1,27	1,31	1,27

Figure 2.2.1. Surface raster model

The representation of surfaces as linear objects is identical to what we see on topographic maps and is based on the use of linear objects. Lines connect sample points that have the same attribute values (Fig. 2.2.2.)

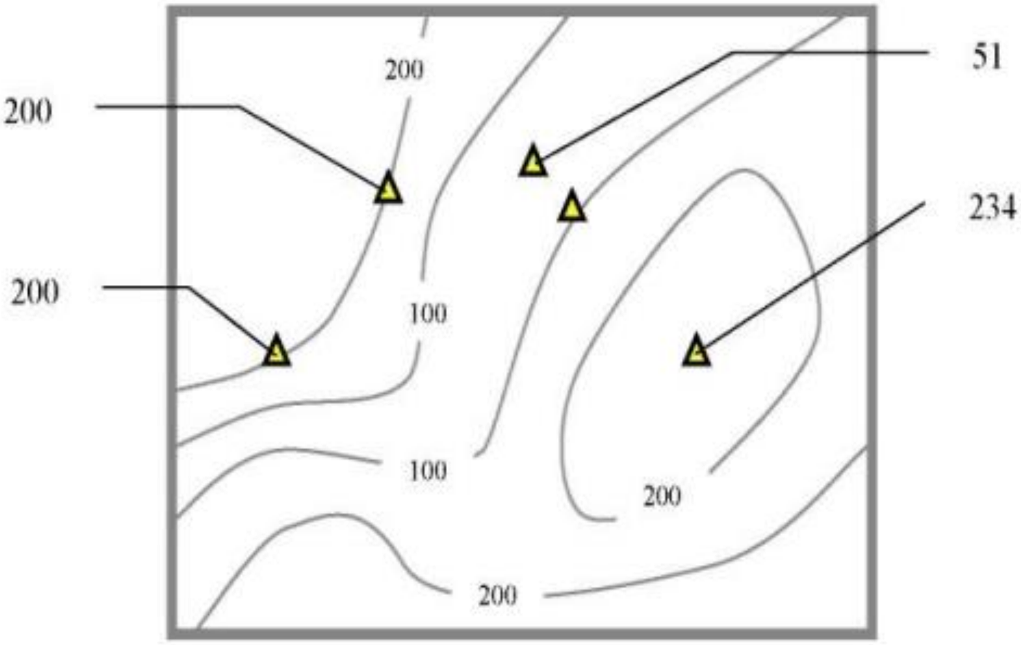


Figure 2.2.2. Linear object surface

Surfaces can be represented by platform objects, most often triangles, since this figure is always convex and lies in one plane. The representation of a surface by a set of triangles is called triangulation. The sample points are the vertices of the triangles; triangles completely cover the studied area. Sampling points are most often located in peaks and depressions, along the lines of ridges and lowlands. The result of modeling is nodes connected by arcs and triangles (Fig. 2.2.3).

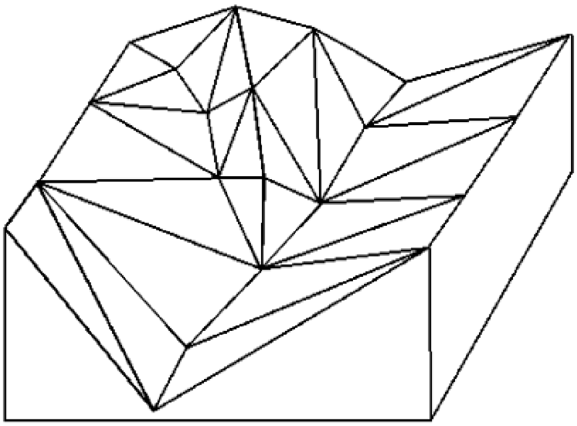


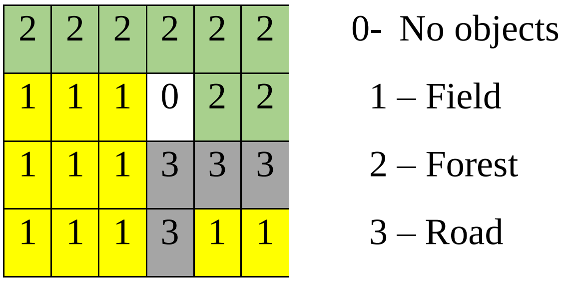
Figure 2.2.3. Surface 3D triangulation

A raster model can be thought of as a picture that draws the world around us through a regular grid of miniature squares, each of which carries a piece of geographic wisdom. This model, embodying simplicity and accessibility, turns the complex world into a rectangular array, where each element, like a pearl, stores information about a geographical phenomenon.

The extension of this raster picture is determined by the minimum size of its elements, which reproduce the smallest details of the earth's surface, usually these elements have a square shape. The orientation of this grid is determined by the angle between true north and the direction given by a row of these squares.

Within this raster model there are zones where each square has the same value, forming a certain layer, although not all maps include complete zones - sometimes they represent a continuously changing phenomenon. The information in each square indicates a specific geographic location, and usually, the coordinates of key points are known, allowing the exact location to be reproduced in the real world.

Each square in this large mosaic carries only one meaning, combining into layers of information that must perfectly relate to each other, creating a single whole with the same number of rows and columns to represent the same fragment of space.



2	2	2	2	2	2
1	1	1	0	2	2
1	1	1	3	3	3
1	1	1	3	1	1

0- No objects

1 – Field

2 – Forest

3 – Road

Figure 2.2.4. Object representation model

The types of values contained in the cells depend on both the type of geographic entity being modeled and the GIS software used. Different systems allow the use of different classes of values: integers, real numbers, string values. Most bitmap systems use only integers for cell values. Integer values are often used as codes identifying the cell class of the area covered (Fig. 2.2.4.)

2.3. Conceptual construction research model of geographical information system for agriculture

We will consider conceptual research model of geographical information system for agriculture as a way of describing it. The main task of this model is to display the features and key characteristics of the system. The model also provides an answer to the question of the purpose of geographic development information system for agriculture system and the expected results of its operation.

When developing research model authors followed the sequence of application of scientific approaches. At the first step, according to the conceptual approach, the statement is formulated, the main direction, sequence and architecture of the research are determined. The main reference points of the development of the information system have been determined. As a result of the conceptual approach, a conceptual model was obtained [3].

Then applied to the aspect approach. As a result, the most important aspects of the system are highlighted. The paper [2] highlights the features of using image analysis to determine the yield of crops. In the future, it is proposed to continue the research using a systematic approach. The main goal is to determine the nature of the connections between the elements of the system.

Development process geographical information system for agriculture it is proposed to be divided into four stages.

1. The stage of determining the goal. At this stage, it is necessary to determine and investigate the environment where the information system will function. Select all objects and subjects in this environment. Formulate the requirements of the final product. Determine the results expected from the launch of the information system and its benefits.

Management of the yield of agricultural crops is part of the management tasks at an agribusiness, which requires the application of new concepts of project and program management

2. Functionality description stage. This stage consists in the formalization of the research task. Development of a mathematical model for describing the problem. The choice of methods that make it possible to solve the research problem in the best possible way.

3. Implementation stage. At this stage, the methods and models obtained at the previous stage should be implemented in the form of software. Also, the specifics of the state problem require the use of special hardware. Therefore, at the stage of implementation, it is necessary to configure it.

4. Stage of diagnosis. At this stage, criteria and quantitative indicators should be identified by which the results of geographic functioning can be evaluated information system for agriculture. Also, at the fourth stage, the system is tested.

The considered four stages are presented as 4 components of Conceptual research model (Fig. 2.3.1). Let's consider each component of the model in more detail.

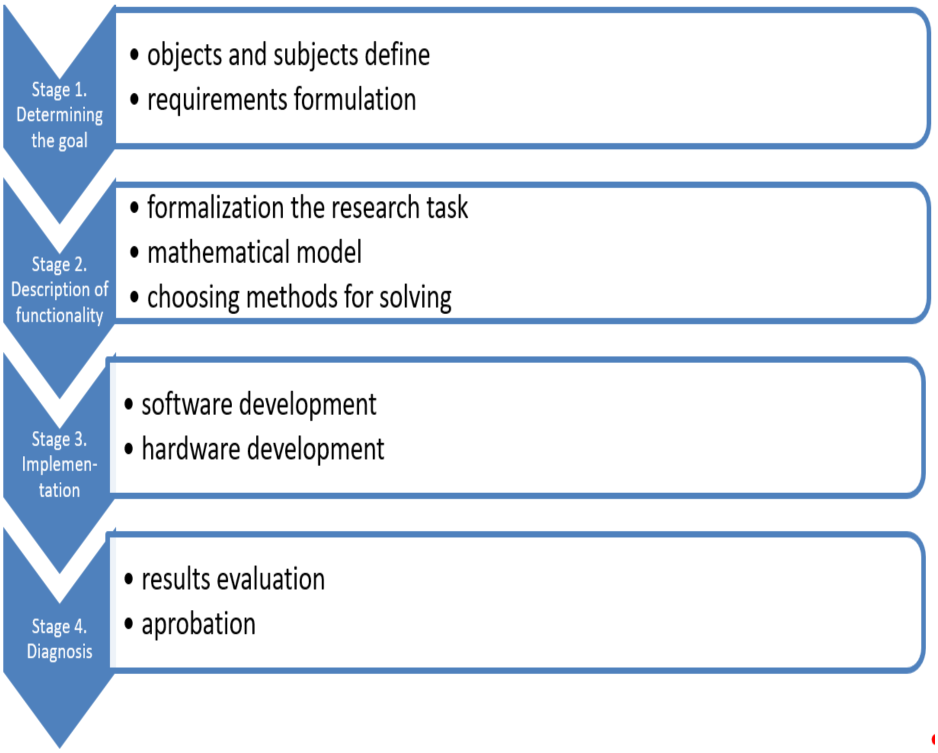


Figure 2.3.1. Conceptual research model

1. At the initial stage, the objective of the geographic information system (GIS) for agriculture should be to gather and display information essential for managerial decision-making. Specifically, for cultivated area management, this should encompass data on:

a. Plant Health: monitoring the overall health and condition of the plants, identifying issues such as disease or nutrient deficiencies.

b. Plant Needs: information on the requirements for fertilizers, water, and other agricultural inputs to optimize plant growth and yield.

c. Harvest Forecast: predictions concerning the quantity and quality of the agricultural crop yield expected in the current year.

d. Harvest Readiness: assessing the appropriate timing for harvest, digging, and other related agricultural processes.

e. Pest Spread: detecting and monitoring the signs of pest infestations to enable timely and effective interventions.

f. Emergency Damage: identifying and assessing damage caused by flooding, hail, wind, or other emergency situations affecting agriculture.

To be effective, the geoinformation system must fulfill the following essential requirements:

a. Map Utilization. The system should have the capability to incorporate new territorial maps and update existing ones to reflect current conditions accurately.

b. Digital Imagery. The system should support the storage and retrieval of digital images of the territory, allowing for their comparison with maps for enhanced analysis and decision-making.

c. System Integration. It must be able to integrate seamlessly with other systems for data collection, storage, processing, and visualization to provide a comprehensive tool for agricultural management.

d. Cross-Platform Compatibility. The system should be accessible and functional across different platforms, ensuring that it can be used on various devices and operating systems, thereby enhancing its usability and accessibility for all users involved in agricultural management.

Also, at the first stage, the following subjects and objects of the system environment were selected:

2. Territory on the Earth's surface. The territory is the main object of observation. The territory is divided into a finite set of plots. Each of them is considered as a valuable resource for the user's activities. The plot has a certain set of characteristics. The result of activity in the specified area depends on the specific value of these characteristics at each moment of time. Site characteristics are dynamic values, that is, they change over time. The characteristics of each site are determined by its location on the territory. Environmental, biological factors and the results of activities on the site in previous periods influence the value of the characteristics of the land plot. Estimated values of land plot characteristics are usually known a priori and can be obtained from the cadastral register, agrarian databases, etc.
3. Land user. It can be an agricultural enterprise, a private person or the state. This facility uses land plots and other types of resources to generate food products.
4. Environmental monitoring systems. For these purposes, you can use drones, which are equipped with a specialized camera to create maps of the state of crops, taking into account infrared reflection. Remote sensing using aviation is also used.

The methodology of creating information resources, designing and filling the information base is based on the system principle of organizing the information environment, the use of unified territorial and branch bases and a spatially oriented system of data collection, accumulation, processing and presentation. Therefore, system-wide principles of database formation are outlined, requirements for the structure, composition and differentiation of data are regulated, models of their organization, forms of presentation, main sources of information and ways of their improvement are defined [1]. The organization of the information base is connected with the use of system-wide principles, the main of which are:

- the principle of territorial and sectoral organization of information - the source information belongs to a specific territory with its inherent specificity of natural conditions and economic activity; it is focused on obtaining and modifying knowledge for specific functional tasks;
- the principle of hierarchy and multi-level structure with arrangement and functional subordination of elements of the whole from higher to lower; at the same time, each level specializes in the performance of a certain range of functions - at higher levels of detail, integration functions are performed (mostly), at lower levels - differentiation.
- the principle of inclusion, which is a direct consequence of the principle of hierarchy, assumes that the requirements for the creation, operation, and development of GIS of each level are determined by a more complex system of a higher rank.
- the principles of complex display of objects, development and block organization of the software and information complex (PIC) and its information base are oriented towards the step-by-step development and implementation of each of the modules (blocks) against the background of 54 general single concepts of system organization, building up and improvement of the components of the PIK STZ;
- the principle of system unity consists in the fact that at all stages of the creation, operation and development of the PIC STZ, the integrity of the system is ensured by connections between its subsystems;
- the principle of thematic (or subject) and system organization of data: thematic organization of information involves its grouping according to the characteristics of the main components of natural and agro-remedial systems, system – the differentiation of thematic data according to their purpose and use in computer software;

- the principle of structural and functional specialization (differentiation) of both the tasks to be solved and the object composition of the information base with the allocation of structural and functional units (subsystems, modules, classification groups, etc.);
- the principle of differentiated description of relations between objects and the objects themselves and their state, achieved thanks to the creation of reference and regulatory information subsystems; data banks of fixed parameter values, subsystems of territorial binding of objects, etc.;
- the principle of combining the database and the knowledge base in the structure of the PIC STZ;
- the principle of compatibility and unity of the information bases of various applied GIS, which are implemented thanks to the geographic information entered into their composition (geographical model of the territory, territorial binding of objects, specialized fund of cartographic information, etc.); in addition, compatibility is ensured by the unity of information-search languages and mathematical support, the commonality of the organizational structure, a single order of information collection and processing, unification of documentation and information coding

At the second stage of the development of the information system, mathematical models and methods that ensure the functioning of the information system should be determined. Let's consider the mathematical model of monitoring the yield of agricultural crops.

Under yield monitoring, we will understand the system of monitoring and measuring the state of growth of agricultural crops, taking into account meteorological, agrometeorological, phenological and other indicators based on the analysis of time series images obtained as a result of photographing cultivated areas, with the aim of evaluating and forecasting the potential crop yield. It is also supposed to take into account the unevenness of the yield on different plots of sown areas, which can be visualized with the help of yield maps, on which the amount of products obtained after harvesting is marked in different colors. It should also be taken into account that yield monitoring should involve a two-dimensional assessment: the amount of products that will be potentially obtained and the quality of these products on the appropriate scale. To ensure effective monitoring at all levels, it is necessary to provide for analysis high-quality aerial images that can be stored and analyzed in the form of time series.

The digital image is a finite matrix of pixels m*n. Therefore, we consider the image as a function of F(x, y), where x defines the coordinate of the abscissa $x \in X=\{x_1,x_2,..., x_m\}$, and y is the coordinate of the ordinates of the plane $y \in Y =\{y_1,y_2,...,y_n\}$. The function value is a set of intensities $<r_1,r_2,...,r_w>$, where r_q is a real number that determines the light reflection intensity in a given spectrum range, and w is the number of spectrum ranges . That calculate the reflection intensity of a plot in a given spectrum, it is necessary that find the sum of the intensities of the corresponding component in all pixels belonging that the plot. Let D be some area of the image corresponding that the area where the agricultural crops are grown. Then the plot can be defined as $D \subset X\times Y$, and the intensity of reflection of the plot in the given range of the spectrum is calculated by the formula:

$$R_q = \sum_x \sum_y^{(x,y)\in D} r_q(x,y), \tag{2.3.1}$$

then the index can be regarded as a function of the image B(F(x,y)). But plant development is a dynamic process, that means it changes over time. Therefore, the indexmust be regarded as a function of the image and the time B(F(x,y), t). Let

$$\beta = \{\beta_1, \beta_2, \cdots, \beta_n\},\ \beta_i = B(F(x,y),t), \tag{2.3.2}$$

where β is a discrete NDVI time series fixed at specified times $t_1, t_2,..., t_n$, and β_i are NDVI at time t_i and t is then shown that there is a functional relationship between land quality, management efficiency , NDVI and site yield :

$$V=g(\beta, Z, M), \tag{2.3.3}$$

where β is the NDVI time series from sowing that harvest Z is land quality indicators MS management parameters (irrigation, fertilizer, pesticide, etc.). The following three main GISA objectives can thus be formulated (2.3.1). Calculation of NDVI time series metrics based he snapshots that are obtained at regular intervals. The problem is solved by calculations by formulas (2.3.1)-(2.3.3) based the forecasting at times t_{n+1}, t_{n+2},... the problem is solved with the help of certain forecasting models, in particular [9, 13] proposing that use the BFAST method.

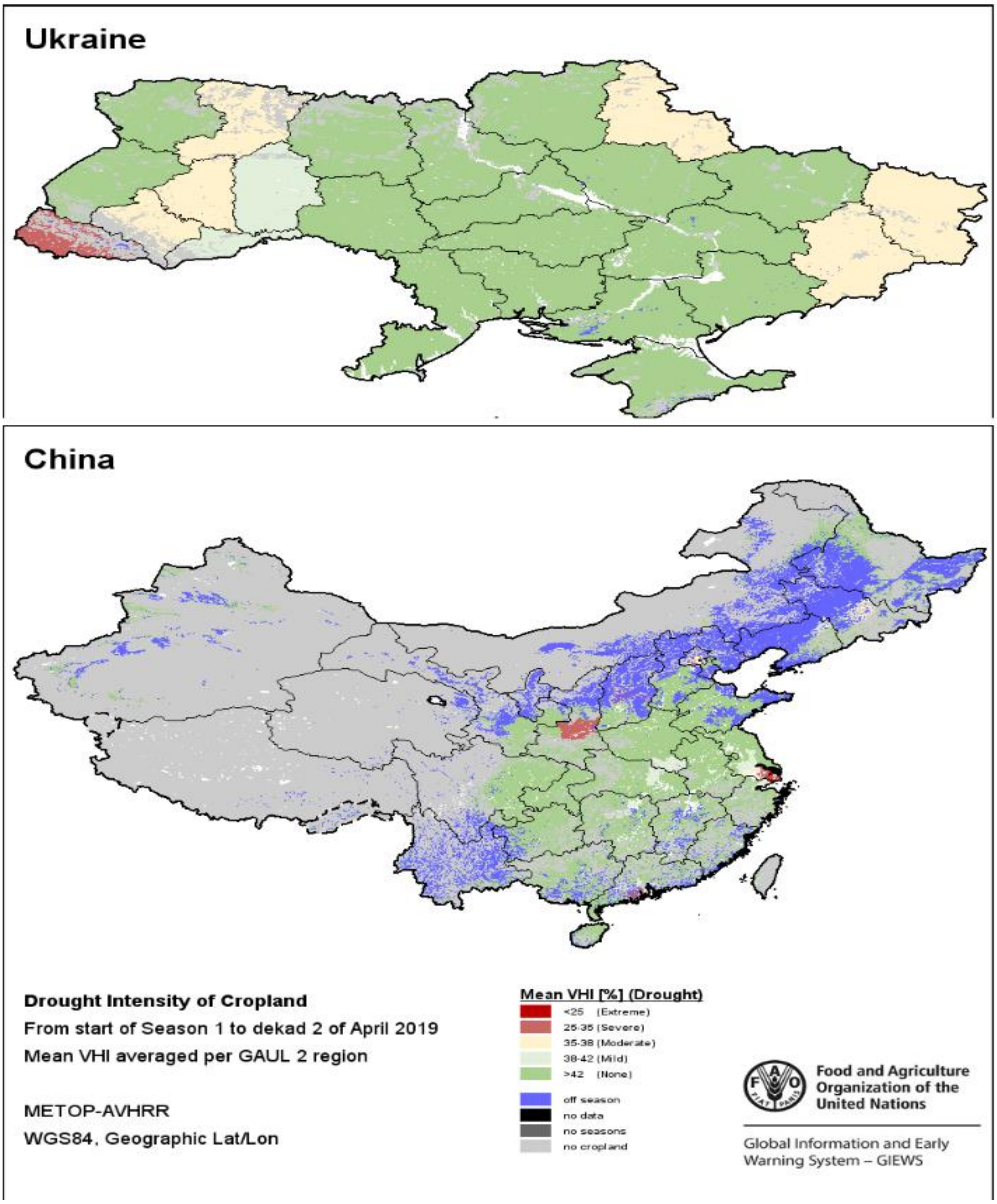


Figure 2.3.2. Example of agricultural drought intensity maps of Ukraine and China for first decade of April 2019

Calculation of land quality indicators Z based on a priori land information. For example, Earth Observation (see fig. 2.3.2) provides information he the latest 36-month period of seasonal, vegetation and precipitation indicators [14]. GISA can also be part of the Decision Support Systems (DSS). The main task of DSS is the task of maximizing productivity

$$V \to \max , \tag{2.3.4}$$

$$H(M) = 0 \tag{2.3.5}$$

where H is a function that specifies the control parameters

The main phenological indicator in the analysis of images of sown areas is the normalized differential vegetation index. This is a quantitative indicator of photosynthetically active biomass used for the quantitative assessment of plant cover, which is calculated according to the formula (2.1.1)

The input data for the information system is the plot image. The image can be represented as some matrix of size $X \times Y$. In the image, certain pixels show a culture, others show soil that is not a culture (Fig. 2.3.2). The intensity for each image has an intensity distribution close to the normal distribution [2]. Therefore, a threshold function can be introduced:

$$\delta(x,y) = \begin{cases} 1, & if \ \beta(x,y) \geq B \\ 0, & if \ \beta(x,y) < B\text{'} \end{cases} \tag{2.3.6}$$

where x, y are the coordinates of a pixel in the image, β is the NDVI value at a certain pixel, and B is some threshold value. Marking Then the productivity of the plot is described by the following model:

$$\beta(x,y) = \delta(x,y)p(x,y) + (1 - \delta(x,y))\overline{p}(x,y),$$

where is $p(x,y)$ the density distribution of the culture in the image, $\overline{p}(x,y)$ is the density distribution of the non-culture. Yield can be estimated by finding the two-dimensional integral of the density distribution of the crop.

The forecasting model based on phenological indicators is based on the determination of critical points in the NDVI trajectory. Consecutive measurements of the NDVI indicator for one area at defined time intervals (day, week, fortnight, etc.) form a discrete time series of estimates.

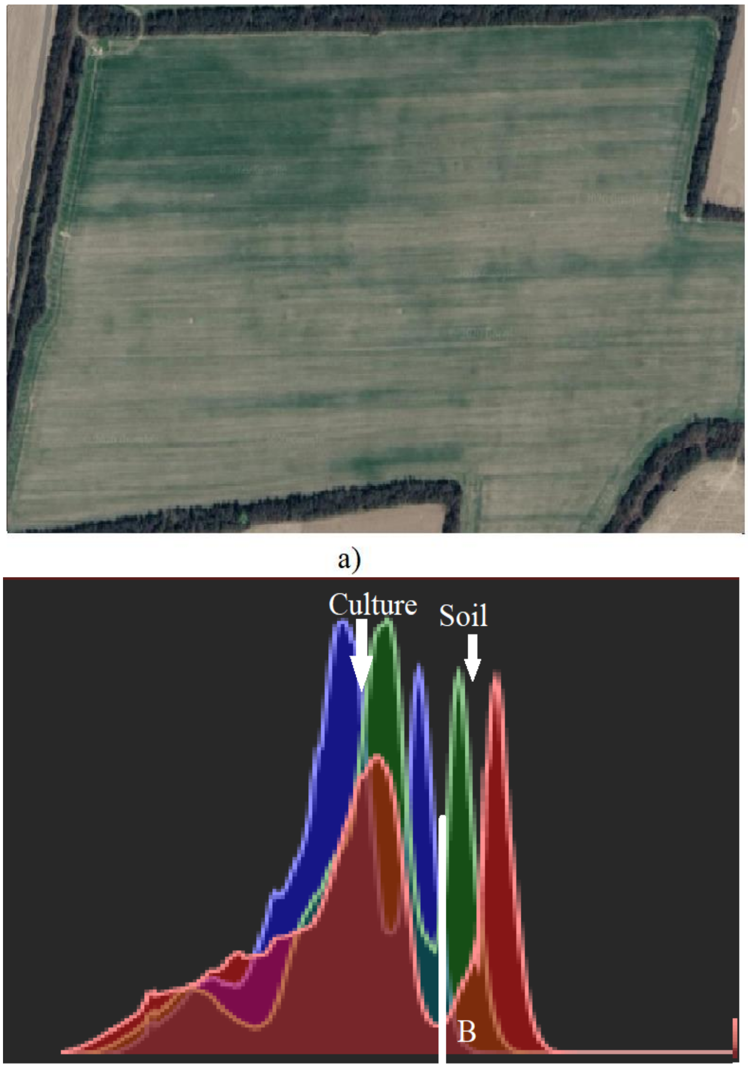


Figure 2.3.3. a) crop field satellite image

b) frequencies histogram with threshold value B.

Using forecasting methods for time series [12], forecast NDVI values for the following time periods. And using the yield forecasting model (1) - estimate yield in the future.

Mathematical and statistical packages and tools such as MatLab [35], the R environment [36], as well as image processing software GRASS [37], IDRIDI [38], ENVI, can be used for the analysis of agricultural land images for yield monitoring [39]. Also, to store information, it is not enough to use traditional databases such as

MySQL or PostgreSQL, so it is necessary to use special spatial databases such as MLPQ/GIS [69].

At the fourth stage, it is necessary to determine the criteria and indicators of the functioning of the geoinformation system. These indicators should make it possible to quantitatively assess the result of the activity. It is a diagnostic tool that allows you to assess the usefulness of GIS in decision-making and improve it in the following seasons. This issue needs a deeper study.

Conclusions to chapter 2

The tasks of geoinformation technologies, the classification of geoinformation systems and the main methods that are at their core are described. Considered concepts, which consist in ensuring the quality and quantity of the obtained agricultural products, as well as in ensuring the effective management of cultivated areas. Also considered are some software, services and devices that allow you to estimate the yield of agricultural products and process time series of images.

Hashing, perceptual hashing, and segmentation methods used for processing time series images of agricultural fields are described. The definition of yield monitoring has been formed, under which we will understand the system of monitoring and measuring the state of growth of agricultural crops, taking into account meteorological, agrometeorological, phenological and other indicators based on the analysis of time series images obtained as a result of photographing sown areas, with the aim of evaluating and forecasting the potential crop yield .

The presentation of spatial data in GIS using raster and vector models is considered. Conceptual was built as a result of research Research Model of developing the geographical information system for agriculture. This model includes four stages that completely determine the outline of the research. At the first stage, the objects and subjects in the environment where the geoinformation system functions are identified, the requirements for the final product are formulated and the expected results are determined.

The presentation of spatial data in GIS using raster and vector models is considered. In exploring the natural world, geographic information systems (GIS) serve as tools to collect and interpret Earth's phenomena, transforming complex data into models that mimic reality for informed decision-making. GIS combines raster and vector data to represent geographical features, where raster models capture visual data through pixels and vector models provide detailed representations via points, lines, and polygons. This integration allows for a nuanced analysis of spatial relationships and patterns, essential for understanding geographic contexts. By balancing the strengths of raster and vector models, GIS facilitates a comprehensive approach to visualizing and analyzing spatial information, supporting various applications in environmental management and planning.

CHAPTER 3. MATHEMATICAL MODELS OF YIELD MONITORING

3.1. Mathematical model for biomonitoring for project management in agrarian sphere.

Biomonitoring is a scientific technique used to assess human or environmental health by measuring the presence and concentration of substances in biological specimens. This method is widely used in environmental science, toxicology, public health, and occupational health. Here are some key aspects of biomonitoring:

Human Biomonitoring involves measuring chemicals, their metabolites, or biological markers in human tissues or fluids, such as blood, urine, hair, or breast milk. The objective is to understand the levels of exposure to environmental chemicals and their potential impacts on health. For example, measuring the level of lead in blood can help assess exposure to this toxic metal and its potential health risks.

Environmental Biomonitoring focuses on assessing the health of an ecosystem by examining the organisms living in it. By analyzing the chemicals present in wildlife, plants, or microorganisms, scientists can infer the levels of pollution in an environment. For instance, the concentration of pesticides in fish can indicate the level of these substances in aquatic ecosystems.

Biomonitoring is crucial in various fields. In public health, it helps in the early detection of exposure to harmful chemicals, enabling timely intervention. In occupational health, it aids in monitoring workers' exposure to hazardous substances. In environmental protection, it assists in tracking pollution levels and the effectiveness of regulatory measures.

Challenges. Biomonitoring comes with its set of challenges. Interpreting data can be complex, as the presence of a chemical does not always indicate harm. It requires understanding the toxicology of substances, exposure routes, and individual susceptibility. Additionally, ethical considerations, particularly in human biomonitoring, are vital to ensure participants' privacy and consent.

Biomonitoring is a powerful tool that bridges the gap between environmental exposures and potential health outcomes, providing essential data for informed decision-making in public health and environmental management.

To build the model, consider the assumption on which the research is based: there is a causal relationship between the quality of land resources, management decisions, the values of phenological indicators, and the yield of agricultural crops. The main phenological indicator used in the analysis of images of cultivated areas is the normalized differential vegetation index (NDVI). This is a quantitative indicator of photosynthetically active biomass used for quantitative assessment of vegetation cover, which is calculated according to the formula (2.1.1).

It is assumed that there is a functional relationship between yield and NDVI. That is, the yield V can be expressed as some function of g:

$$V = g(\beta, Z, M, e, t), \tag{3.1.2}$$

where β is NDVI value, Z is quality indicators of land resources, M is management efficiency indicators, e are other factors that also affect yield and t is time. Other factors include weather conditions, natural disasters, epidemics and random processes that cannot be predicted and are beyond our control.

In the realm of project management within corporate agriculture, particularly concerning crop yield, several management efficiency indicators play a critical role in guiding decisions and strategies. These indicators offer a lens through which the effectiveness of various agricultural practices and project outcomes can be assessed, thereby informing management on areas of success and those requiring improvement.

One crucial indicator is the yield per acre, which measures the output in terms of crop volume or weight produced per unit area. This indicator helps managers gauge the effectiveness of their agricultural practices and inputs like fertilizers, water, and pest control measures. By comparing these figures across different periods or against industry benchmarks, managers can identify trends, forecast future performance, and make informed decisions to enhance productivity. Cost efficiency is another vital metric, encompassing the analysis of input costs relative to the output value. This includes the cost of seeds, labor, machinery, and other inputs against the revenue generated from the crop yield. Efficient management aims to optimize this ratio, ensuring that the agricultural operations are not only productive but also economically viable. Another significant aspect is the time efficiency concerning the agricultural project's lifecycle, including planting, growing, and harvesting phases. Time-related metrics assess whether the crop's growth and harvesting align with planned schedules, which is crucial for maximizing yield and quality, especially in crops sensitive to seasonal and market fluctuations. Additionally, the use of technology and innovation metrics, such as the adoption of precision agriculture techniques, can offer insights into the project's modernity and sustainability. The integration of advanced technologies like satellite imagery, drones, and data analytics into agricultural practices can significantly enhance decision-making, resource allocation, and ultimately, crop yield. Lastly, sustainability indicators, including soil health, water usage, and carbon footprint, reflect the long-term viability of agricultural practices. These metrics not only align with the growing emphasis on sustainable agriculture but also ensure that crop production can be maintained or increased without compromising future resources.

By monitoring and analyzing these management efficiency indicators, corporate agriculture can refine its project management approaches, driving improvements in crop yield, cost-effectiveness, and sustainability, ultimately contributing to the sector's overall success and resilience.

As a corollary, it is hypothesized that multispectral field images can be used to predict yield.

We will build a model for monitoring the yield of agricultural crops based on the analysis of geodata and site images. For this, we consider that according to the BFAST method, the additive decomposition model of time series forecasting of phenological indicators has the form:

$$\beta_t = C_t + Q_t + e_t, \quad (3.1.3)$$

where β_t is the time series data, that is, the value of the phenological indicator recorded at time t, C_t is the trend component, Q_t is the seasonal component, e_t is the random component, $t = \overline{1, n}$, $n$is the number of observations, that is, the number of elements of the time series.

Taking into account (3.1.2) and (3.1.3), the model for monitoring the yield of agricultural crops based on the analysis of geodata and images of the site will look like this:

$$V = C(\beta, Z, M, t) + Q(\beta, Z, M, t) + e(t), \quad (3.1.4)$$

where C defines the trend component, Q defines the seasonal component, and e defines random components. Trend and seasonal components do not depend on random factors and therefore can be predicted using appropriate methods.

To predict the yield of agricultural crops using the developed model, it is necessary to have a knowledge base with reference images that correspond to certain growth indicators of specified agricultural plants. By comparing the current image with images from the knowledge base to establish similarities, it is possible to draw a conclusion about possible deviations from normal growth and ripening of the crop. In the event of a negative deviation, it requires the necessary solutions to correct the situation. Also, the developed model can be used to predict the amount of harvest. However, this forecast must be adjusted taking into account other factors that affect the growth and maturation of plants.

Development of the method of analysis of multispectral images taking into account geoinformation data

The input data for crop yield assessment is an image of the plot, which can be represented as some matrix of size $X \times Y$.

Agricultural culture in the field does not grow evenly. Therefore, it is necessary to consider the concept of the density of crop distribution as the ratio of the area where the crop grows to the total area of the field. Part of the image pixels corresponds to the part of the field where the crop grows. The other pixels correspond to the part of the field where the crop does not grow. Let's introduce the threshold function:

$$\delta(x,y) = \begin{cases} 1, if \ \beta(x,y) \geq B \\ 0, \ if \ \beta(x,y) < B' \end{cases} \qquad (3.1.5),$$

where *x* and *y* are the coordinates of the pixel in the image, *β(x,y)* is the NDVI value of the image pixel with the corresponding coordinates, and B is some threshold value.

The thresholding function defines an image thresholding operation that transforms an input color image into a black and white image. For the binarization of monochrome images, Otsu's method [71] is quite simple, but effective. This method is used to automatically find the binarization threshold based on the analysis of the shape of the intensity frequency histogram.

For the practical implementation of the Otsu method, the restriction that each pixel of the image may not contain an arbitrary intensity value, but only one of a predefined discrete set of values, is important. Specifically, the SENTINEL mission provides data using 12 bits for each band. Accordingly, the intensity of each of the ranges can take one of 4096 values. Then the intensity frequency histogram *H(i)* is the fraction of image pixels with intensity equal to $i = \overline{0, I}$, and *I* is the maximum intensity value. For the mission SENTINEL *I =4095.* For older systems that use 8 bits per image channel, 256 intensity values are available, so *I =255.*

Let the desired threshold value *0≤B≤I,* which binarizes the image into two classes q1 and q2. Then, the probability that some pixel of the image belongs to each of the classes depends on the threshold value B:

$$q_1(B) = \sum_{i=0}^{B} H(i), \qquad (3.1.7)$$

$$q_2(B) = \sum_{i=B+1}^{I} H(i). \qquad (3.1.8)$$

Then the mathematical expectation is calculated according to the formulas:

$$\mu_1(B) = \sum_{i=0}^{B} \frac{iH(i)}{q_1(B)}, \qquad (3.1.9)$$

$$\mu_2(B) = \sum_{i=B+1}^{I} \frac{iH(i)}{q_2(B)}, \qquad (3.1.10)$$

$$\mu(B) = \sum_{i=0}^{I} iH(i). \qquad (3.1.11)$$

And the variance is calculated according to the formulas:

$$\sigma_1{}^2(B) = \sum_{i=0}^{B} \frac{(1-\mu_1(B))^2 H(i)}{q_1(B)}, \qquad (3.1.12)$$

$$\sigma_2{}^2(B) = \sum_{i=B+1}^{I} \frac{(1-\mu_2(B))^2 H(i)}{q_2(B)}, \qquad (3.1.13)$$

Consider the problem of maximizing interclass dispersion:

$$\sigma^2(B) = q_1(B)(\mu_1(B) - \mu(B))^2 + q_2(B)(\mu_2(B) - \mu(B))^2 \to max \qquad (3.1.14)$$

Given that $q_1(B)+q_2(B)=1$, and $q_1(B)\mu_1(B)+q_2(B)\mu_1(B)=\mu(B)$, we obtain the equivalent maximization problem:

$$q_1(B)q_2(B)(\mu_1(B) - \mu_2(B))^2 \to max \qquad (3.1.15)$$

to find the threshold value B as a solution to problem (3.1.15) by performing a direct enumeration of all threshold values.

The following model can be used to estimate crop yield on the site:

$$\beta(x,y) = \delta(x,y)p(x,y) + (1-\delta(x,y))\bar{p}(x,y), \qquad (3.1.16)$$

where $p(x,y)$ is the density of the distribution of culture on the part of the plot corresponding to the pixel on the image with coordinates x and y, and $\bar{p}(x,y)$ is the density of the distribution of non-culture on the part of the plot corresponding to the pixel on the image with coordinates x and y.

Yield can be estimated by finding the two-dimensional integral of the density distribution of the crop. Taking into account the discreteness of the image, the integral can be calculated as the sum:

$$V = \sum_{x=0}^{X} \sum_{y=0}^{Y} p(x,y) \qquad (3.1.17)$$

where V is an estimate of field productivity.

Filling the database is a significant part of the GIS project implementation efforts. Spatially distributed data can be represented in the database using vector or raster data models. The raster model of spatial data assumes that geographic information is stored as a matrix. One of its corners, usually the upper left or the center, is used to link each element of the matrix to spatial coordinates. The dimensions of the matrix depend on the resolution of the data collection devices and the dimensions of the investigated field. Thus, the research satellite EOS AM-1, which operates under the leadership of NASA, photographs the Earth's surface with a resolution of 250 m to 1 km. One pixel of the image corresponds to a square with a side from 250 meters to 1 kilometer. Each cell of the matrix can store one or more attributes, for example intensity in different frequencies of the spectrum.

A vector model of spatial data involves the representation of graphical data using lines and regions corresponding to certain attributes. As a rule, lines are specified as broken by the coordinates of their vertices. The coordinates of the vertices are given as Cartesian coordinates in some rectangular coordinate system, for example, in the Gauss-Kruger projection, or geographic coordinates of latitude and longitude.

A raster model of spatial data has a simpler data structure than a vector model, but requires more computer memory. Data from a raster model can be converted to a vector model and vice versa.

For the task of monitoring the yield of crops on the site, the most important task is to establish the boundaries of the site and link the available data to the site. A priori soil and weather data are also important. The presence of plot boundaries makes it possible to enter the membership function:

$$\omega(x,y) = \begin{cases} 1, \ if \text{ pixel}(x,y) \text{ is inside,} \\ \ \ 0, \text{ if } (x,y) \text{ is outside,} \end{cases} \qquad (3.1.18),$$

which allows you to take into account only the part of the image that belongs to the plot when predicting the yield. Then the yield estimate (17) is calculated taking into account the membership function according to the formula:

$$V = \sum_{x=0}^{X} \sum_{y=0}^{Y} \omega(x,y) p(x,y). \qquad (3.1.19)$$

Operational information about the state of the field (presence of pests, plant diseases, soil condition) makes it possible to create yield maps. These maps make it possible to estimate future benefits from growing agricultural crops, because there can be significant differences in yield within the same field. This is influenced by both the condition of the soil and the slope of the field surface. Separately, soil maps can be formed and evaluated, allowing to evaluate the content of sand, clay, and peat in them. All this makes it possible to plan the places for planting plants within one field in a timely manner, as well as take into account the necessary proportions of fertilizers, which are determined differently for different types of soil. It is clear that the calculation requires not only information about the condition of the soil, but also the yield on these soils in previous periods, as well as information about the use of fertilizers in previous seasons.

Application of the crop yield monitoring model based on the analysis of geodata and plot images for forecasting the yield of agricultural crops

Let us present the trend component of the crop yield monitoring model (3.1.3) as a linear-piecewise function [3]:

$$C(\beta, Z, M, t) = a_i + b_i t, \tag{3.1.20}$$

where $r_{i-1} < t \leq r_i$, $i = \overline{1, m}$are observation control points.

We will present the seasonal component of the crop yield monitoring model (4) linear harmonic regression:

$$Q(\beta, Z, M, t) = \sum_{k=1}^{K} \chi_k \left(\cos\gamma_k \sin\left(\frac{2\pi kt}{\lambda}\right) + \sin\gamma_k \cos\left(\frac{2\pi kt}{\lambda}\right)\right), \tag{3.1.21}$$

where amplitude χ_k and phase γ_k are unknown, and frequency λ is known. For Ukraine, as a rule, the frequency is 1 year. A frequency of 6 months is typical for China.

When using the agricultural crop yield monitoring model based on the analysis of geodata and site images to predict the yield in this study, the random component is not considered and is considered equal to 0. Considering (3.1.20) and (3.1.21), the agricultural crop yield monitoring model (3.1.3) can be written as:

$$V = a_i + b_i t + \sum_{k=1}^{K} \chi_k \left(\cos\gamma_k \sin\left(\frac{2\pi kt}{\lambda}\right) + \sin\gamma_k \cos\left(\frac{2\pi kt}{\lambda}\right)\right). \tag{3.1.22}$$

So, let's consider the algorithm for using the crop yield monitoring model based on the analysis of geodata and site images for yield forecasting:

Figure 3.1.1. NDVI calculation example.

1. Use available data in GIS about plot boundaries to calculate the membership function (3.1.18).

2. process a series of pictures of the site in different spectra. To do this, for each picture:

    2.1. Find the NDVI value (Fig.3.1.1) for each pixel according to formula (2.1.1);

    2.2. Construct a histogram of NDVI intensity frequencies and find the threshold value for the purpose Otsu method by formulas (3.1.7)–(3.1.15);

    2.3 Using the found value to calculate the threshold function (3.1.5), and find the value of the density of the crop distribution on the site as a solution of linear equations (3.1.16);

    2.4. Calculate yield estimates taking into account the membership function according to formula (3.1.19). Obtain a time series of yield estimates.

3. Using the time series of yield estimates, we find the coefficients of the trend component of the crop yield monitoring model (3.1.20) by the linear regression method.

4. Using the time series of yield estimates, we find the amplitudes and phases of the seasonal component of the crop yield monitoring model (3.1.21).

5. Find the value of the yield forecast according to formula (3.1.22) by substituting t equal to the harvest time.

3.2. Development of an integration model of artificial intelligence for yield monitoring based on GIS data.

The development of an integration model of artificial intelligence for yield monitoring is relevant, as it allows accurate forecasting of yield, optimizing the use of resources and increasing the productivity of the agricultural sector. This model helps adaptation agricultural practices before changes climate and volatility of growing conditions, providing sustainable and efficient agricultural production.

Convolutional neural networks (CNN) for image analysis and fully connected layers for structured data analysis can be used to create a neural network that combines images with soil and yield data [72].

Neural network architecture contains the image input layer. Image input layer must match the size of input images received from drones or satellites. Next are convolutional layers (CNN) for images. Several convolutional layers that detect features at different levels of abstraction in images. Also, pooling layers to reduce the dimensionality of output data from convolutional layers and increase computational efficiency [73].

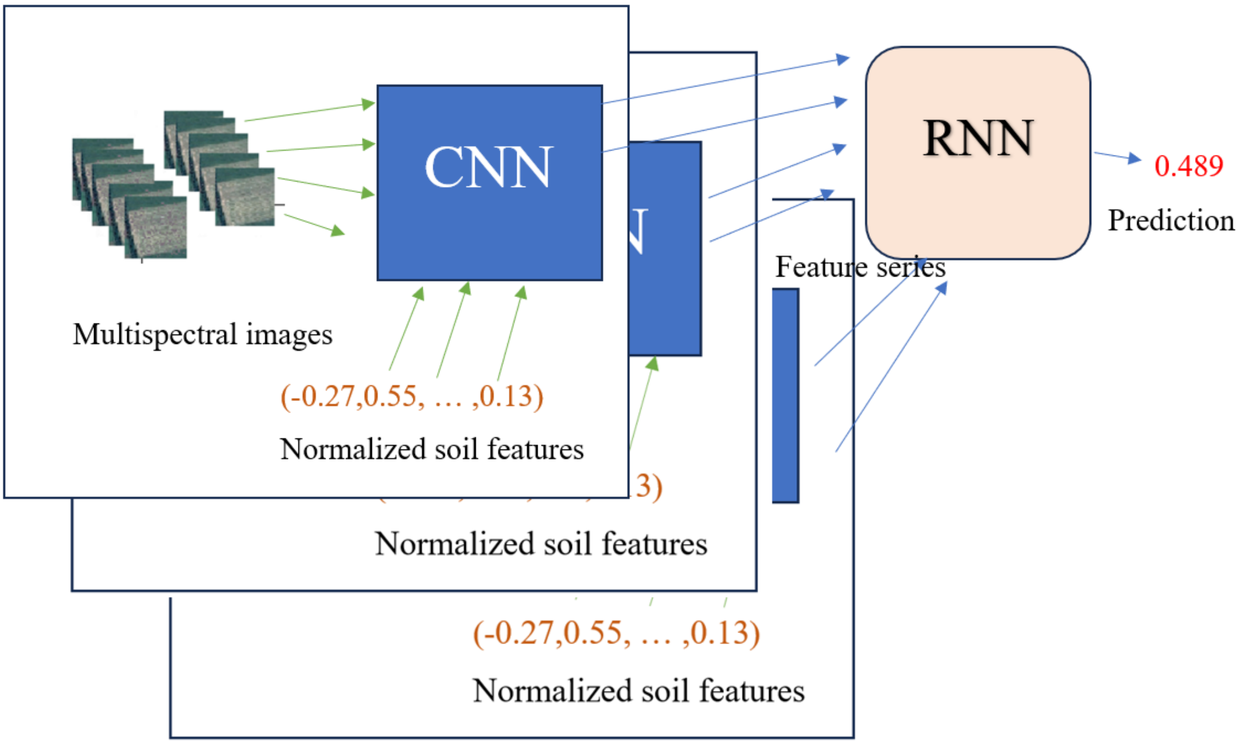


Figure 3.2.1. Neural network architecture

Also, the neural network contains a separate input layer for soil data, each neuron of which corresponds to a separate characteristic of the soil being analyzed. This layer is followed by one or more fully connected layers that process input soil data [74].

The output data from the last convolutional layer and the fully connected layer for the soil data are combined into a single vector. This combined vector is fed to additional fully connected layers for further processing (fig 3.2.2).

Recurrent neural networks (RNNs) are particularly suitable for processing data sequences due to their ability to retain information from previous steps that is used in processing the current input signal [75]. This makes RNNs ideal for time series analysis, as well as for analyzing sequences of feature vectors derived from CNNs during image processing.

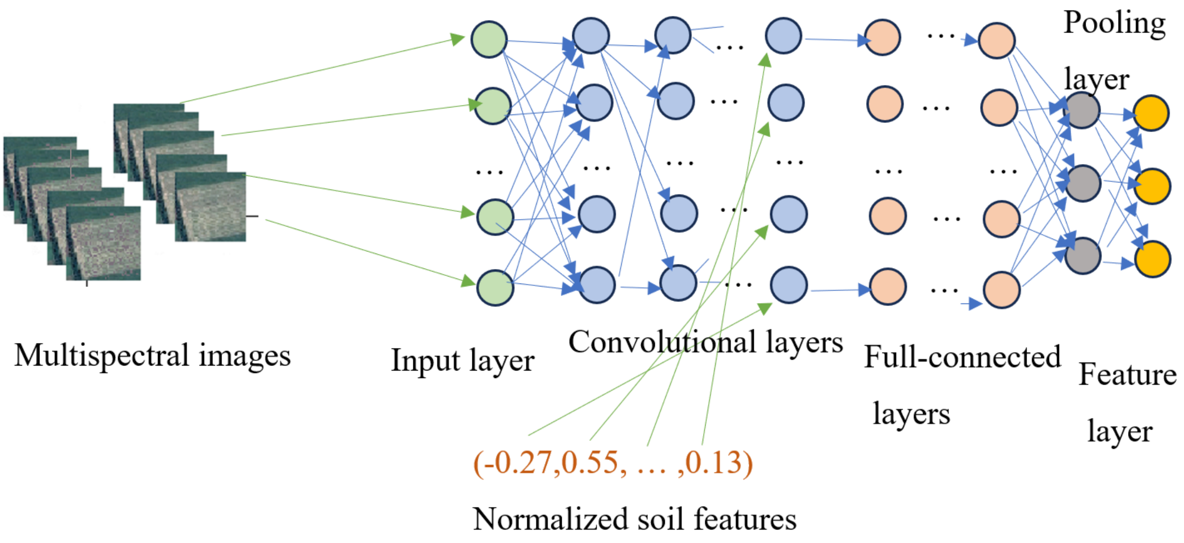


Figure 3.2.2. CNN architecture

RNNs consist of several layers: input, hidden, and output (Fig 3.2.3). The output from each step can be converted to the desired format. For the considered problem, this is a numerical forecast of productivity. Features of RNNs include context dependence. An RNN is able to take into account information from previous steps, allowing it to analyze input data in context. Be aware that the problem of missing or exploding gradients may occur when training RNNs, especially when working with long sequences. This can make network training more difficult.

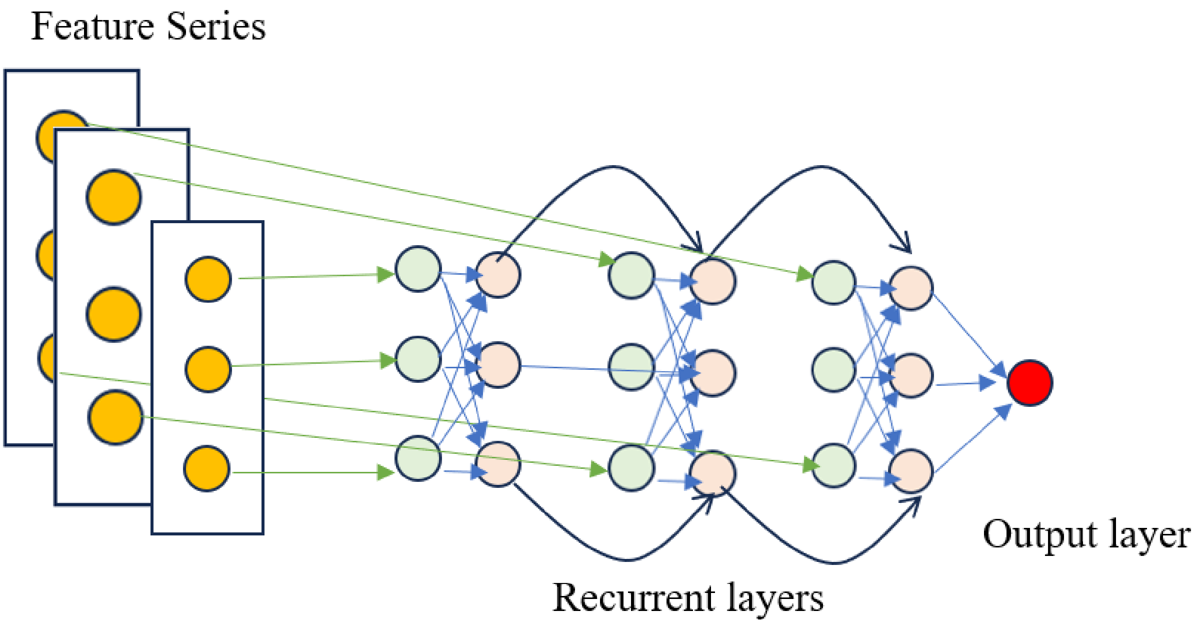


Figure 3.2.3. RNN architecture

RNN Improvements: To overcome some of the limitations of the basic RNN, improved architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), which better cope with long-term dependence and problems of vanishing gradients [76, 77].

The use of RNN to analyze sequences of feature vectors allows integrating spatial information obtained from individual objects or images with temporal dependence presented in the form of sequences, providing deep data analysis.

The last fully connected layer is used for yield prediction. The last layer contains a single neuron whose value corresponds to the yield forecast. Yield is the target variable.

In neural networks, activation functions play a key role, they help the model learn from complex data by introducing non-linearity into the learning process, allowing the network to learn and model complex tasks such as classification and regression.

Normally, activation functions are not used for input layers, as they serve to pass input data directly to subsequent layers.

For convolutional layers, it is advisable to use rectified linear unit (ReLU) or its variations, such as Leaky ReLU or Parametric ReLU , as these features help avoid the problem of vanishing gradients, speeding up learning convergence and providing efficient feature detection in images [78].

Rectified linear unit, is one of the most popular activation functions used in neural networks. Formally, the definition of ReLU can be presented as follows:

$$f(x) = \begin{cases} x, if\ x > 0 \\ 0, if\ x \leq 0 \end{cases} \qquad (3.2.1)$$

This means that if the input x is positive, the function returns x; if the input is negative, the function returns 0.

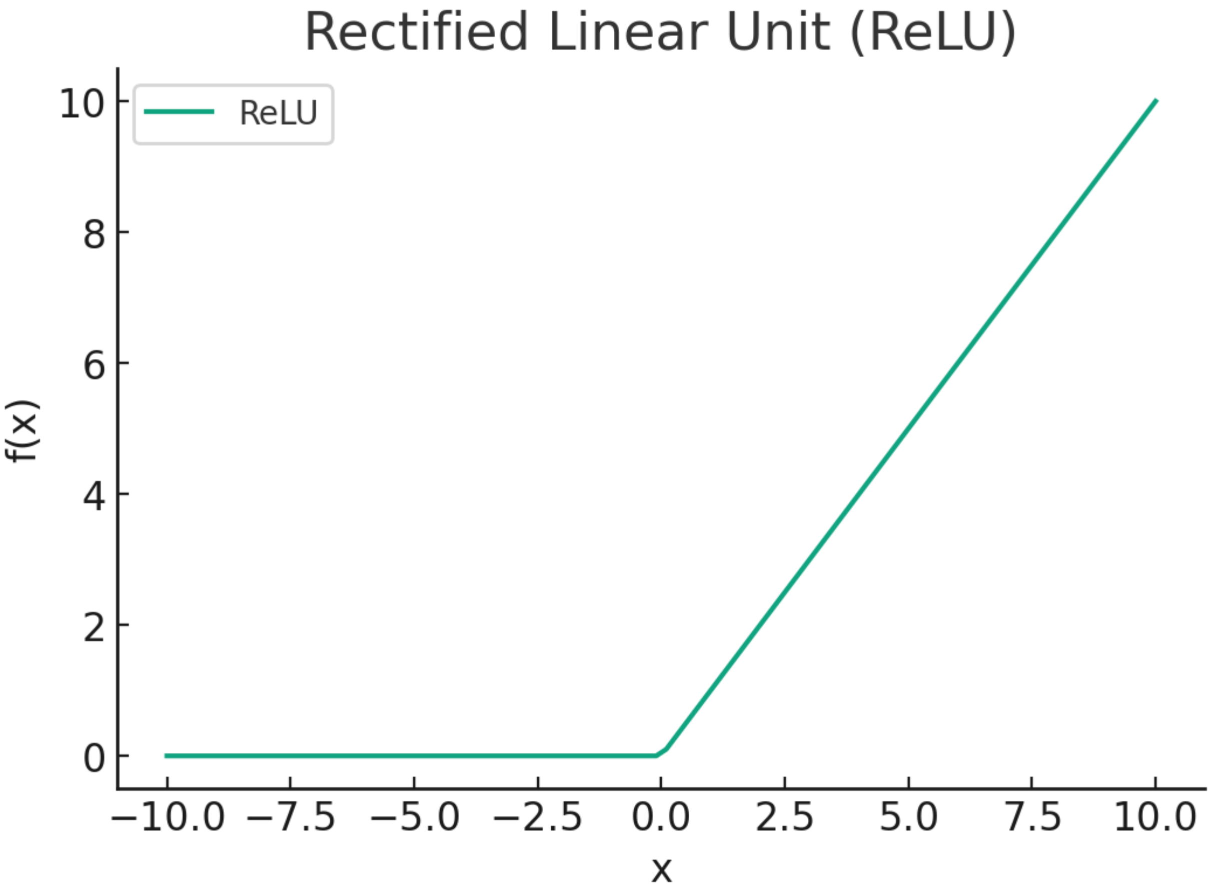


Figure 3.2.4. Rectified Linear Unit chart

Although ReLU looks very simply, it is non-linear, allowing neural networks to learn on complex and non-linearly separable data.

The feature of this function is its computational efficiency: ReLU is less computationally complex compared to other activation functions such as sigmoid or hyperbolic tangent, as it requires fewer mathematical operations.

In deep networks with a large number of layers, ReLU helps avoid the vanishing gradient problem, where the gradients become too small for effective deep layer learning.

The main limitation of ReLU is the problem of "dead" neurons. If the neuron outputs negative values, ReLU will set these outputs to zero, which can lead to a situation where the neurons become "dead" and do not fire regardless of the input, since the gradient for negative values will also be zero.

Leaky ReLU is a modification of the standard ReLU activation function designed to eliminate the problem of "dead" neurons that can occur when using ReLU [79]. In standard ReLU, all negative inputs are zeroed, which can lead to loss of gradients during backpropagation. In Leaky ReLU, instead of setting negative inputs to zero, they are multiplied by a small positive number α, less than 1. Formula for Leaky ReLU looks like this:

$$f(x) = \begin{cases} x, if\ x > 0 \\ \alpha x, if\ x \leq 0 \end{cases} \tag{3.2.2}$$

where α is a small coefficient that allows a small part of the negative signal to pass through.

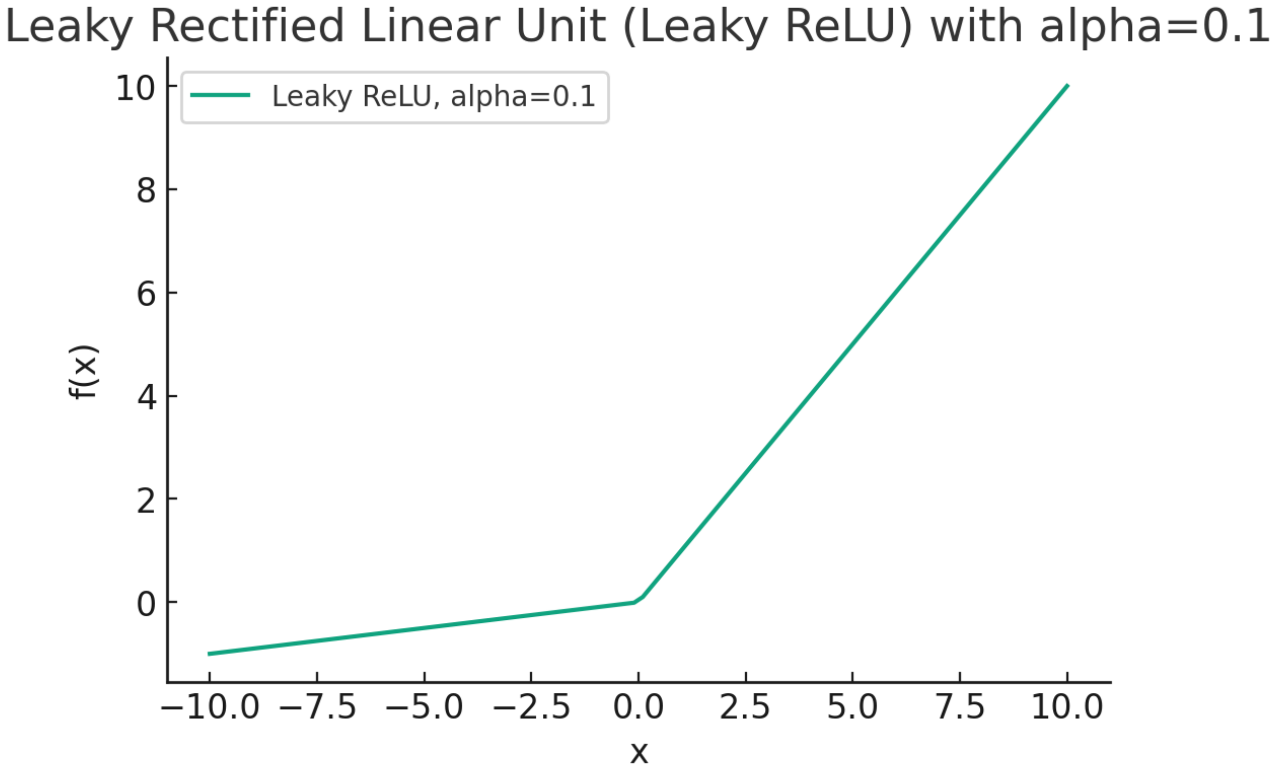


Figure 3.2.5. Leaky rectified linear unit chart

Parametric ReLU (PReLU) extends the idea of Leaky ReLU , making the coefficient α a learned parameter. This means that instead of fixing α to a constant small value, the network itself determines the optimal value of α during the learning process. This can add flexibility to the model, allowing it to dynamically adapt to different features of the data.

Exponential Linear Unit (ELU) is another variation of ReLU that introduces an exponential function for negative inputs instead of a linear one [80]. This is done to reduce the effect of dead neurons, allowing some gradients to "flow" even with negative inputs. The ELU function is defined as:

$$f(x) = \begin{cases} x, if\ x > 0 \\ \alpha(e^x - 1), if\ x \leq 0 \end{cases} \tag{3.2.3}$$

where α is a constant that controls the value of the slope for negative inputs.

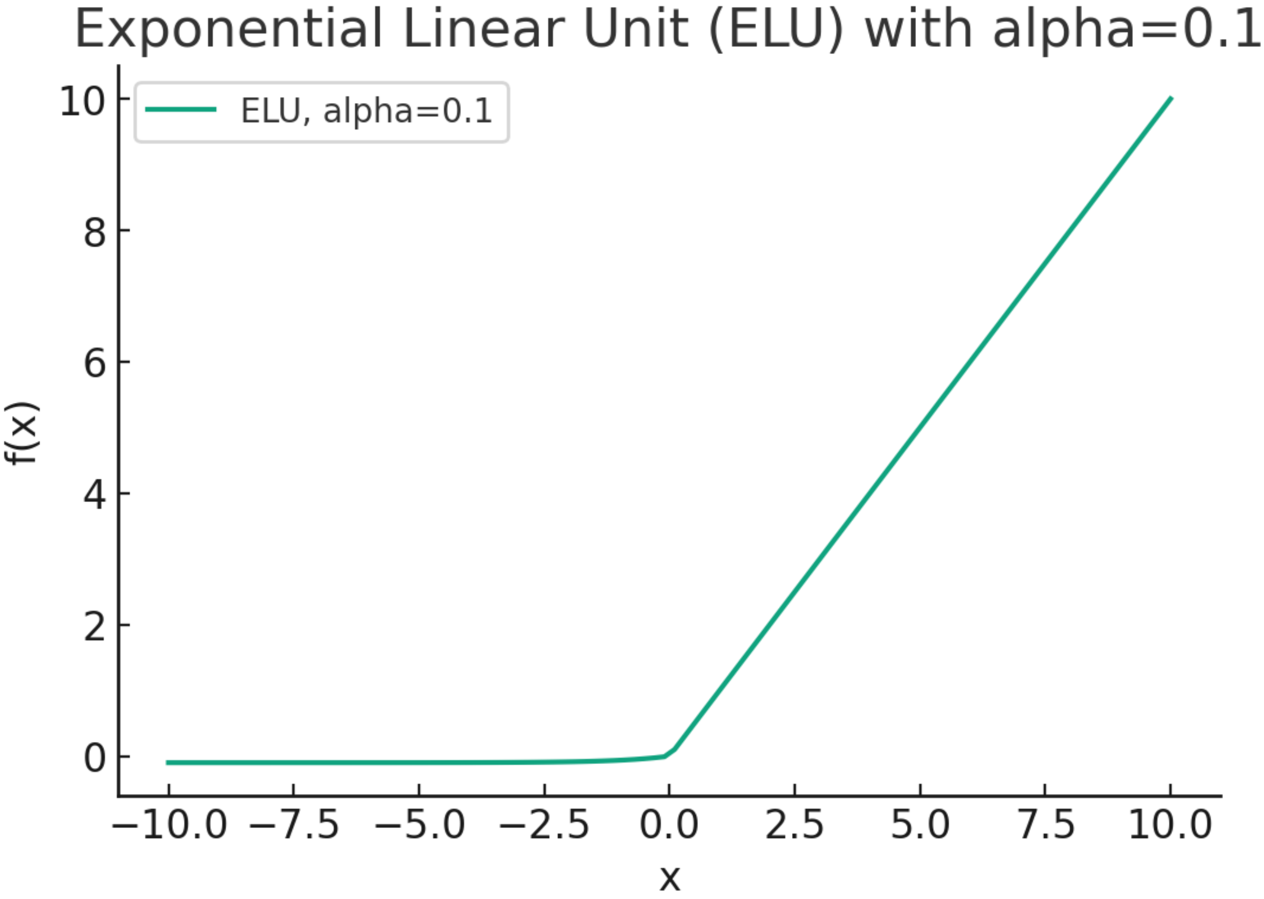


Figure 3.2.6. Exponential linear unit chart

These variations of ReLU aim to improve the learning process of neural networks by providing better processing of negative signals and increasing the ability of the network to learn from complex data without losing important information due to nulling of negative activations.

For fully connected layers used to input soil data as an activation function, ReLU can be used to ensure non-linearity and avoid the problem of missing gradients.

After combining features from different data types, one can continue to use ReLU or other non-linear activation functions for further fully connected layers. Although ReLU and its variants are a popular choice for activation functions in many layers of neural networks, there are scenarios were using the hyperbolic tangent (tanh) may be more appropriate, especially after combining features from different data types.

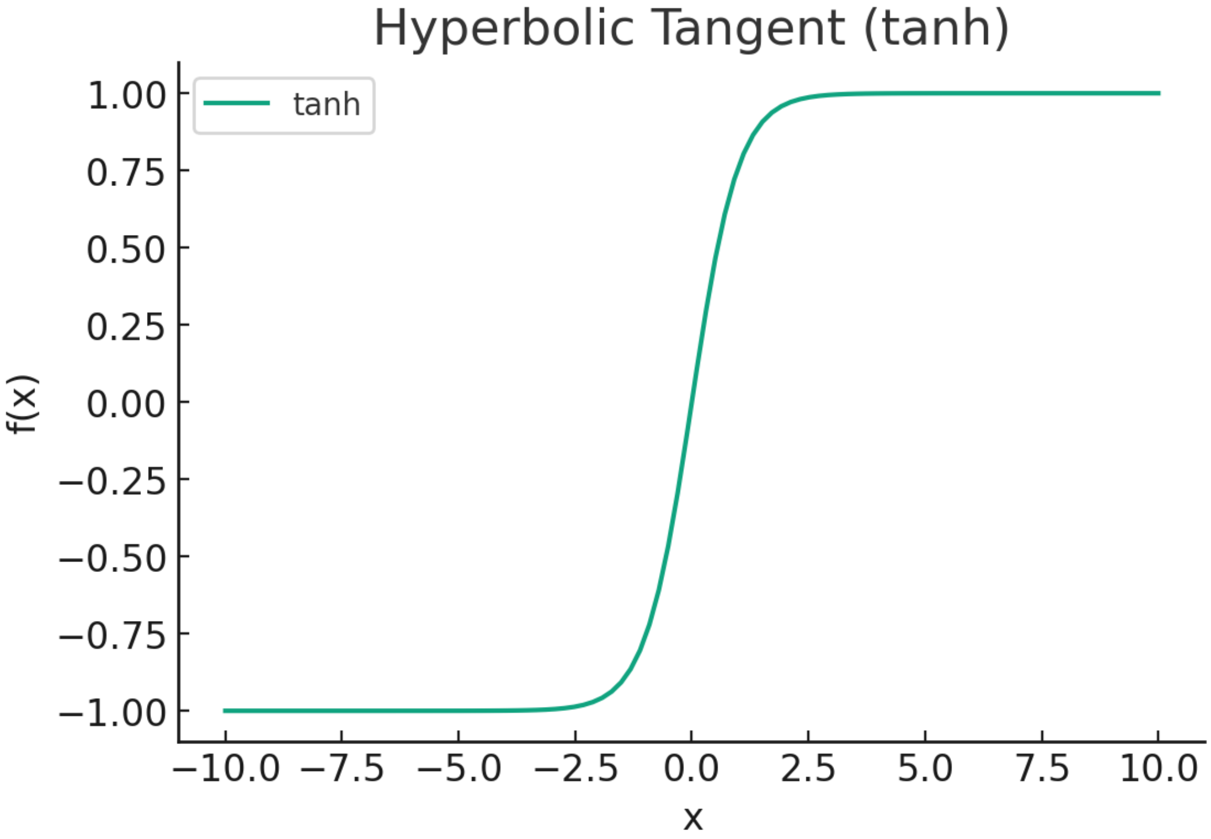


Figure 3.2.7. Tangent hyperbolic activation function chart

tanh activation function is defined as:

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3.2.4}$$

It converts an input value to an output in the range from -1 to 1, which can provide some advantages after combining features [81] :

1. Compared to ReLU , which outputs only non-negative values, tanh outputs an output that is centered around zero. This can lead to better convergence during training, as the average activations per layer will be closer to zero, making weight updating easier.

2. Since tanh has pronounced non-linearity, it can be useful in models where it is necessary to model complex relationships between data, especially when these relationships are not well expressed due to positive values of the outputs, as is the case with ReLU.

3. If the combined features have important information that is expressed as both positive and negative values, tanh allows you to preserve this information through its output, which is distributed symmetrically about zero.

Therefore, in contexts where it is important to preserve information about negative values or when faster convergence needs to be achieved, using tanh after merging heterogeneous data may be more efficient compared to ReLU.

The activation function for the output layer depends on the type of problem. Since our task belongs to regression, we can use linear activation.

Variable values between -1 and 1: Normalizing the data to this range can improve the learning speed and stability of the neural network because it helps avoid problems with weights that are too large or too small, which can cause gradients to explode or disappear. It also helps neurons with activation functions that are sensitive to the scale of the input, such as ReLU, which has a non-zero gradient only for positive values, work better. Normalization of the input data ensures that the activation of the neurons will be effective over the entire range of input values.

When the input has values over a wider range, this can cause neurons to fire only for very high input values, which can make the learning process more difficult or slow. Using normalized data helps to avoid this problem by ensuring that the neurons will respond to the input data more uniformly.

In addition, many optimization algorithms used to train neural networks assume that all input features have the same scale. When the data is normalized, it promotes faster and more stable convergence during optimization, as the magnitude gradients of the weight updates will be more consistent throughout the model.

Thus, normalizing the data to the range from -1 to 1 not only improves the learning process, but also helps improve the overall efficiency and stability of neural networks.

It is important to take into account a number of soil parameters for predicting productivity, as they affect the health of plants and, accordingly, the quantity and quality of the harvest. Key indicators of the soil:

Table 3.2.1. Soil features

Variable	Indicator
S_1	pH
S_2	The content of organic substances
S_3	Nitrogen content (N)
S_4	Phosphorus content (P),
S_5	Potassium content (K),
S_6	Calcium content (Ca),
S_7	Magnesium content (Mg)
S_8	Soil moisture
S_9	Soil density

Let consider indicators roles in detail: pH indicates the acidity or alkalinity of the soil. Plants have specific pH requirements, and a deviation can affect their uptake of nutrients. Organic matter is an important component of soil that contributes to its structure and ability to hold water and nutrients. The content of key elements such as nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg) and others are necessary for plant health and growth. The ability of the soil to retain water is important for providing plants with the necessary moisture. Soil density affects the root system of plants and their ability to penetrate the soil for water and nutrients. Understanding these indicators and their optimization can significantly improve yield forecasts and help in the rational use of resources.

3.3. Methods of teaching learning neural networks.

Let explore the foundational algorithms that guide these learning processes, dissect the mechanics of backpropagation, and delve into various training techniques that determine how a network adapts and evolves. By contextualizing these methods within the broader landscape of artificial intelligence, we provide a framework for readers to appreciate how neural networks can be taught to perform an array of complex tasks, from image recognition to natural language processing. Through this exploration, the chapter seeks to equip readers with a deep understanding of the mechanisms that enable neural networks to derive meaningful insights from data, setting a solid groundwork for further exploration and application in the field of machine learning.

Hybrid neural network training includes the following stages:

1. Data preprocessing includes normalization, resizing, data augmentation, etc. to ensure effective and efficient network training. The preparation of multispectral images for training neural networks for yield prediction involves several key steps that ensure that the input data are as informative and relevant as possible. Multispectral images obtained from drones or satellites provide a unique opportunity to detect and analyze various characteristics of vegetation and soil.

2. Pre-processing can include distortion correction, noise removal, and atmospheric correction. Background removal such as soil or water can also be done at this stage.

3. Vegetation indices such as NDVI are calculated for each image pixel. These indices make it possible to assess the health and biomass of vegetation, which is critically important for predicting productivity.

4. Each multispectral image needs to be tied to a specific field, for which geospatial data is used. This makes it possible to accurately match vegetation indices with specific plots of land on which crops are grown. For neural network training, i.e. linking each image or its segment to a specific yield value. This require the collection of historical yield data for the fields concerned.

5. The dataset is divided into training, validation and test samples. This separation helps to evaluate the model's ability to generalize learning to new data.

6. The preparation of multispectral images is a key step in building effective yield forecasting models, as it determines how well the model can interpret the data and make accurate predictions.

A neural network training process that involves several steps using first a Convolutional Neural Network (CNN) for image analysis and then an Recurrent Neural Network (RNN) for analyzing sequences of CNN-derived feature vectors is an example of deep learning using hybrid architectures.

CNN uses filters or kernels that "pass" the input image to create feature maps. These filters are capable of detecting edges, corners, textures and other characteristics in an image. During training, the weights of these filters are optimized to minimize the difference between the actual and predicted network outputs.

pooling operation reduces the dimensionality of each feature map while preserving important information. This helps to reduce the number of parameters and computational complexity of the network.

pooling operation (such as max pooling) reduces the dimensionality of each feature map while preserving important information. This helps to reduce the number of parameters and computational complexity of the network.

Training includes the following stages:

1. Forward propagation. In forward propagation, the input data (image) is passed through convolutional, activation, and pooling layers, and then through fully connected layers to find the value at the output layer. Each convolutional layer uses

a set of filters or kernels that are applied to an image or feature map, creating feature maps.

2. Calculation of loss function. After forward propagation, the output of the network is compared to the expected result using a loss function. The loss function determines how well the model performs its task by measuring the discrepancy between actual and predicted outputs.

3. Backpropagation. Using backpropagation, the gradients of the loss function are calculated with respect to each parameter (weights and biases) in the network [82]. These gradients are used to update the network parameters in order to minimize the loss function.

4. Optimization of updating weights based on calculated gradients. In this study, stochastic gradient descent (SGD) was used [83].

Stochastic gradient descent is one of the main optimization methods widely used in machine learning and deep learning to minimize the loss function. The difference between SGD and traditional gradient descent is that SGD updates model parameters (e.g., weights in the neural network) not based on the entire data set, but using only one or a few training examples at each optimization step. This makes SGD much more computationally efficient, especially for large datasets.

Parameter update: In SGD, model parameters are updated for each training example or mini-batch of training examples using the following rule [84]:

$$\theta = \theta - \eta \cdot \nabla \theta \, J(\theta; w\,(i)\,, v\,(i)) \qquad (3.3.1)$$

where θ are model parameters,

η is the learning rate,

J is the loss function,

w(i), v (i) are training examples.

The learning rate is a hyperparameter that controls how much the model parameters change in response to an estimate of the loss gradient. The correct choice of learning speed is critical for learning efficiency.

CNN training is an iterative process, where at each iteration the model gradually improves its ability to extract important features from the data and use that information to perform a specific task, such as image classification or object recognition.

Training a recurrent neural network on feature vectors derived from a convolutional neural network is an essential approach in tasks where sequential data needs to be analyzed. This is particularly relevant for yield forecasting, where a sequence of images from the field can provide important information about changes in crop growth and health.

After training the CNN, the feature vectors for each image in the sequence are extracted. These vectors represent the high-level features of the image, such as textures, colors, shapes, etc., that CNN has identified as important. These vectors serve as inputs to the RNN, where each vector corresponds to one time step in the sequence.

Recurrent Neural Networks are unique in that they can process sequential data while retaining information about the previous context in their internal state, making them ideal for tasks where temporal dependencies need to be taken into account, such as language processing, time series forecasting, or, as in your case, image sequence analysis for yield prediction.

Each element of the sequence (the feature vector obtained from the CNN for each image) is fed to the input of the RNN in turn. In the context of images, each vector can represent certain characteristics of the frame, for example, the state of the plants in the field at a certain time.

An RNN has an internal state that is updated at each step of sequence processing. This state functions as a "memory" of the network, allowing it to take into account information from previous steps when processing the current feature vector.

At each step, based on the current input vector and internal state, the RNN produces an output. For prediction problems, this output can be an intermediate prediction or part of the information for the next steps.

After the output value is computed, the internal state of the network is updated in preparation for processing the next element in the sequence. This update ensures that the network "remembers" information from previous steps and can use it to produce more accurate outputs.

Learning recurrent neural networks (RNNs) differs from learning traditional feedforward networks due to their ability to store information from previous steps (time points). Therefore, special methods should be used for RNN training:

1. Backpropagation Through Time (BPTT) is an extension of backpropagation adapted for RNN [85]. With BPTT, errors are propagated back through each time step, updating the weights based on their impact on the error at each time step. This means that errors from the network output are propagated back not only through the layers of the network, but also across time steps, allowing the weights to be updated according to their impact on predictions throughout the sequence.

When applying BPTT, the Gradient technique is important Clipping, which helps to avoid the problem of explosive gradients in RNN [86]. This technique consists in limiting the magnitude of the gradient to a certain threshold value if it exceeds this threshold. This prevents excessively large updates to the weights, which can destroy the network's training. One of the key advantages of RNN is its ability to take context into account in data, which is very important for sequences where each element depends on the previous ones. In the context of yield prediction, this means that RNN can take into account both the current state of plants and their developmental history, allowing for more accurate predictions.

Conclusions to chapter 3

A combined mathematical model was developed to reflect the relationship between phenological parameters and the yield of agricultural crops. This model combines the threshold adaptive method of determining the region belonging to the image of crop sowing and the method of forecasting the values of the time series of phenological indicators based on the selection of its trend, seasonal and random components. The Otsu method was adapted to determine the adaptive threshold value.

The process of creating and training a hybrid neural network that integrates image data and soil information for yield prediction is also described. It has been established that the task requires the use of a combined type neural network for effectively finding a solution. A network architecture is proposed, which for receiving input data includes convolutional neural networks for image processing and fully connected layers for soil data analysis. 9 parameters of the soil, which are important for the task, have been determined. This integration allows the network to consider a variety of information, increasing its ability to accurately predict yield. The output layer of the first stage contains neurons that store the integrated seed values.

In the second stage, the network uses recurrent neural networks to analyze data sequences, which adds the ability to take into account temporal dependencies and context. Proposed procedures of two-stage training with a teacher. The backpropagation method combined with stochastic gradient descent to update the weights was used to train the CNN. At the second stage, the Backpropagation method was applied Through Time in conjunction with the Gradient technique Clipping.

CHAPTER 4. INFORMATION SYSTEM FOR MONITORING CROP YIELDS USING GEOINFORMATION SYSTEMS

4.1 Development of an information system for monitoring crop yields using geoinformation systems.

The development of a GIS-based yield monitoring information system is critical for several reasons that underline the need for such technological solutions in today's agricultural sector. By accurately analyzing yield data, weak areas and potential for improvement in specific areas can be identified. This allows farmers to make informed decisions about fertilizer application, irrigation and other agronomic measures to increase productivity.

A GIS-based information system helps optimize the use of resources, including water, fertilizers, fuel for agricultural machinery, etc. This reduces costs and increases the efficiency of agricultural activity. GIS data analysis can promote more sustainable farming practices, reducing environmental impact. For example, precision farming can reduce the excessive use of fertilizers, which prevents water pollution.

The system provides the ability to forecast yields, allowing agricultural producers to plan marketing, logistics and sales of products more efficiently. Real-time monitoring allows prompt detection of problems, such as pests or plant diseases, and quick response to them, reducing possible losses. The information system allows you to document all interventions and changes in the fields, providing a detailed history for analyzing trends and planning future actions.

Thus, the development and implementation of a yield monitoring information system based on GIS not only contributes to increasing the efficiency and productivity of the agricultural sector, but also supports sustainable agriculture and ecologically sound resource management.

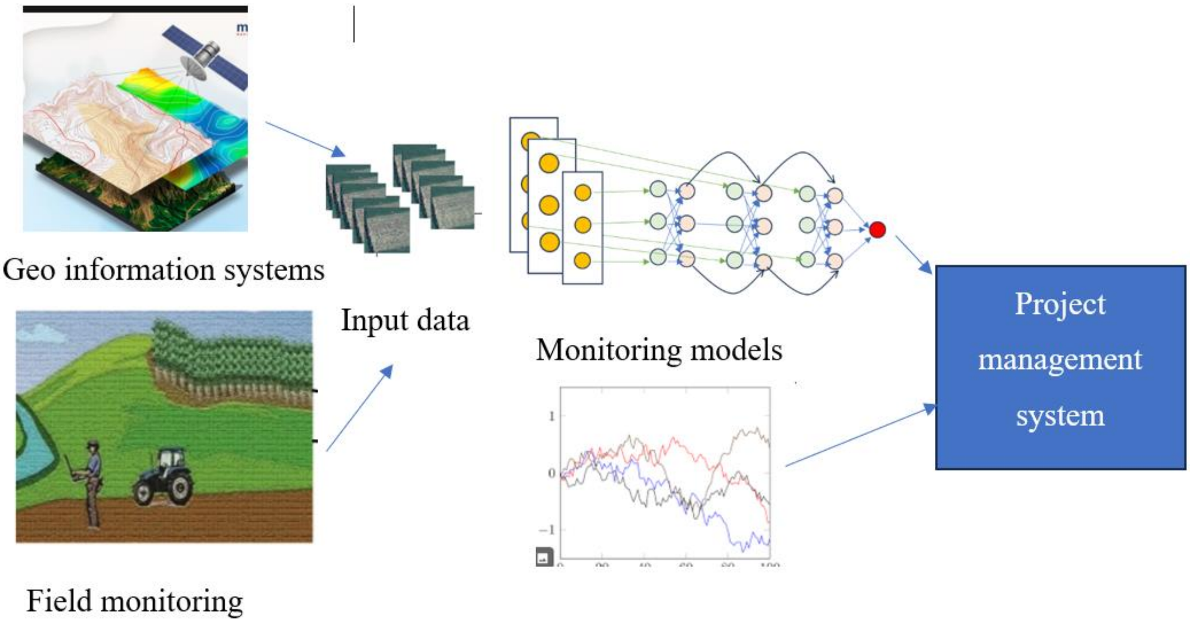


Figure 4.1.1 Conceptual diagram of the information system

A conceptual diagram of a modular yield monitoring information system, based on the principles of geographic information systems (GIS), covers a comprehensive integration of data and technology. This allows you to effectively collect, process, analyze and visualize information about the state of agricultural land. This system serves as a decision support tool aimed at optimizing agricultural processes, including cultivation, harvesting and resource management.

The developed information system has a modular structure and includes 5 modules:

1. The data collection module is responsible for integrating data from various sources, including satellite images, drones, ground sensors, etc. This module provides collection of data on the condition of crops, soil, climatic conditions and other important parameters.

2. The data storage module centralizes the storage and management of all collected data. It provides integration, storage and availability of data for further analysis and visualization.

3. The data processing module performs data analysis and processing functions, using machine learning algorithms, statistical analysis, and geospatial computing to detect patterns, estimate yields, and identify potential field problems.

4. The GIS module provides visualization and spatial analysis of data, allowing users to see information in the context of geographic maps, diagrams, heat maps, etc.

5. The machine learning module is responsible for developing models and building neural network architectures, training models based on historical and current data, evaluating and optimizing models, based on which model parameters can be optimized to improve their performance. Trained models are integrated into the data processing engine to perform real-time prediction and analysis.

Also, the information system can be supplemented with a decision-making module, the functionality of which allows you to use analytical data to develop recommendations and strategies for optimizing agronomic processes, planning the use of resources and managing productivity.

The data collection module in a GIS-based yield monitoring information system is fundamental because it is responsible for collecting input data that will be analyzed and transformed into useful information for monitoring. Special attention in this module is given to the integration of multispectral images from the SENTINEL-2B satellite and yield maps.

The SENTINEL-2B satellite provides high-resolution multispectral images that cover different spectral bands. These images can provide information about the state of vegetation, soil moisture, temperature, and other factors.

Once collected, the images are processed to correct possible errors, calibrate the spectral data, and convert it into an analyzable format. This includes correction of atmospheric and other effects on the image. On the basis of multispectral images, indices such as NDVI (formula 2.1.1) can be extracted, which allows to assess the condition and health of vegetation. This helps to determine, for example, if there are signs of stress in the plants or potential yield problems.

Yield maps are created based on data collected by ground sensors and equipment on agricultural machinery. These maps show the diversity of yields within a single field or region.

Integrating yield maps with multispectral imagery allows for a better understanding of how different factors affect yield. For example, it is possible to analyze how changes in the NDVI index affect yields in specific regions. The integrated data can be analyzed in a spatial context to identify spatial patterns, such as areas of high or low yield in fields. By integrating multispectral imagery from SENTINEL-2B with crop yield maps, the data acquisition module allows for a multidimensional view of farmland conditions, facilitating accurate and efficient yield monitoring and management.

Classic databases that do not have specialized spatial functions are often unable to efficiently process geospatial data, which is critical for information systems that use GIS.

Here are some key benefits that justify the need to use an extension like PostGIS for working with geospatial data:

1. Classic databases do not support spatial data types such as points, lines, polygons and other geometric types. PostGIS introduces these data types, allowing you to accurately describe geospatial objects within the database.
2. Analyzing geospatial data requires specialized functions that are not available in standard databases. PostGIS provides a large number of functions for spatial analysis, such as calculating distances, defining intersections of objects, creating buffer zones, etc.
3. Fast access to geospatial data requires special indexing methods, such as spatial indexes. PostGIS uses such indexes for efficient organization and quick retrieval of spatial data.
4. PostGIS complies with Open standards Geospatial Consortium (OGC), ensuring its compatibility with a wide range of other geospatial tools and

applications. This is critical for integration and data sharing between different systems.

5. Large geospatial data sets require a database to be able to scale efficiently. PostGIS , being an extension of PostgreSQL , inherits its advantages in terms of reliability, performance and scalability.

PostGIS is one of the most popular spatial extensions for the PostgreSQL relational database management system [87]. It allows users to store, query and manage spatial data within the database. Here's more about the key features and capabilities of PostGIS :

PostGIS contains spatial data types, such as GEOMETRY and GEOGRAPHY, which allow you to store a variety of geometric objects, including points, lines, polygons, and their collections.

To improve query efficiency, PostGIS uses spatial indexes, in particular R-tree indexes, which significantly speed up searching and accessing spatial data.

PostGIS provides a wide set of spatial functions for performing various operations, such as determining the distance between objects, finding intersecting objects, analyzing buffer zones, calculating the area and perimeter of polygons, etc.

Users can perform complex spatial queries using SQL syntax that allows spatial analysis to be integrated directly into the database.

PostGIS allows you to perform geometric transformations such as union, intersection, difference and other geometric operations.

PostGIS can integrate with various other tools and platforms used for geospatial analysis, visualization and cartography, such as QGIS, ArcGIS, as well as with various programming languages that support work with PostgreSQL databases.

PostGIS adheres to Open standards Geospatial Consortium (OGC), which ensures its compatibility with a wide range of other geospatial tools and data formats.

Users can extend the functionality of PostGIS using additional plugins and extensions that allow adding new features or integrating PostGIS with other systems and tools.

These capabilities make PostGIS an extremely powerful tool for anyone working with geospatial data, allowing you to efficiently store, process, and analyze spatial information within a relational database.

The data analysis module was successfully developed using the Python programming language, which is a leader in the field of data analysis due to its high performance, flexibility and support for a large number of specialized libraries.

The developed module includes comprehensive tools for data processing and analysis, integrating a number of powerful Python libraries for optimal efficiency. It provides deep analytical insight and supports a wide range of analytical operations. The following libraries were used in the development:

1. Pandas: Used for efficient processing and analysis of data in the form of tables. The module integrates Pandas for data manipulation and transformation, supporting a wide range of data formats.
2. NumPy: Used to optimize high-performance mathematical operations on arrays. The module uses NumPy to process numerical data with high speed and efficiency.
3. Matplotlib and Seaborn: These libraries are used for data visualization, allowing you to create graphs and other types of visualizations to analyze and present results.
4. Scikit-learn: A library applied to machine learning, the module includes Scikit-learn functionality to build predictive models and perform classification, regression, clustering, and dimensionality reduction algorithms.

This modular architecture ensures the flexibility, scalability and efficiency of the GIS-based yield monitoring system, allowing the integration of new technologies and data analysis techniques, as well as adapting to the changing needs of the agricultural sector.

4.2. Implementation of the integration model of artificial intelligence for yield monitoring

To implement the integration model of artificial intelligence for yield monitoring, an algorithm was developed to ensure precise, efficient, and data-driven decision-making in agriculture. This algorithm includes various stages to systematically process and analyze vast amounts of data, ranging from soil conditions to weather patterns, enabling farmers to predict and enhance crop yields effectively. By integrating diverse data sources and applying machine learning techniques, the algorithm helps in identifying patterns and anomalies that might not be apparent through traditional methods, thus optimizing resource allocation and maximizing agricultural productivity.

To implement the integration model of artificial intelligence for yield monitoring, an algorithm was developed, which includes the following stages:

Stage 1: Data collection.

Stage 2: Pre-processing and data analysis.

Stage 3: Development and training of the neural network.

Stage 4: Testing and Validation.

Stage 5: Refinement and optimization.

Stage 6: Implementation and monitoring.

Stage 7: Analysis and reporting.

At the first stage of the research, we initiated the process of data collection, which is a fundamental stage. To plan the shooting, you should determine the areas for shooting, ensuring a variety of agronomic conditions and stages of crop growth. Schedule shooting for different weather conditions to collect a representative data set. Drones with high-quality imaging cameras that provide sufficient image resolution to identify important field details should be used for filming.

Data obtained from drones can be supplemented with satellite images for a broader overview of the studied areas. Soil analysis is also an important step. For its implementation, you should choose areas for collecting soil samples, ensuring geographic and agronomic diversity. Plan sampling to match aerial survey locations. Standardized soil sampling should be performed documenting the exact location and other important parameters of the site. The samples should then be analyzed for key nutrients, structure, pH and other important characteristics.

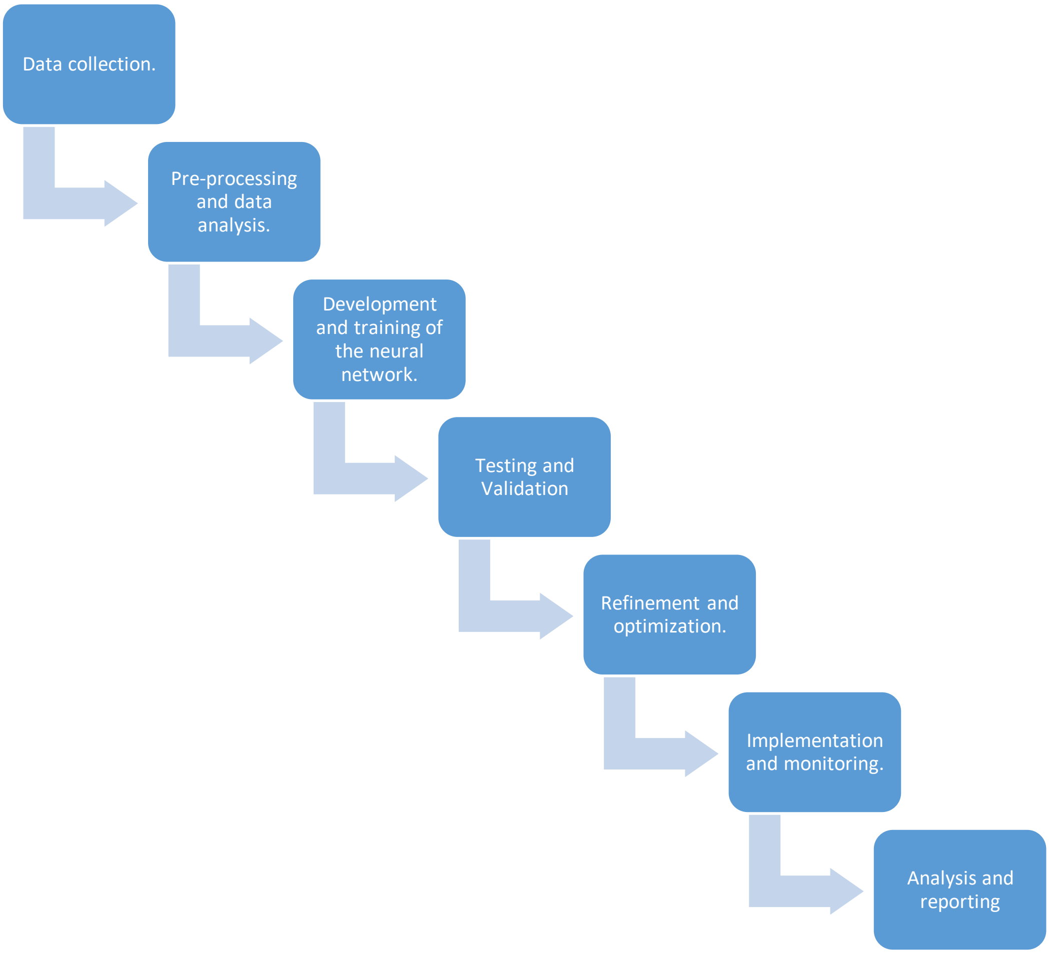


Figure 4.2.1. the integration model of artificial intelligence for yield monitoring implementation algorithm

For the completeness of the analysis, the yield information from each studied plot was systematized. These data form the basis for verifying the predictive accuracy of the developed neural network model, allowing to evaluate its ability to accurately predict potential yield based on the analysis of collected images and soil characteristics.

This integrated approach to data collection provides a solid foundation for the next phase of research, which consists of developing and training a neural network capable of effectively predicting yield based on the information obtained.

In order to ensure further processing and analysis of the data, pre-processing and analytical research of the collected data should be performed.

In the process of processing images obtained during aerial photography or from satellite data, it is critical to correct any distortions that may affect the accuracy of the analysis. First, we focus on correcting perspective and lighting. Perspective distortions can occur as a result of the non-uniform location of the camera relative to the surface of the earth, so their correction allows to ensure the uniformity of the geometric proportions of objects on the ground. Illumination, which can vary depending on the time of day and meteorological conditions, also needs to be normalized to ensure consistency of color and intensity, which avoids errors in further analysis.

After correcting the main parameters of the image, we proceed to segmentation, which is a key stage for the selection of objects of interest, such as vegetation, water bodies, soil areas, and others. Segmentation helps to determine the structural and textural characteristics of the surface, which is important for the analysis of the condition of the land. The use of machine learning or deep learning algorithms in this context allows to significantly increase the accuracy and efficiency of segmentation, providing high recognition ability and separation of objects of interest in images.

Analysis of soil data plays a critical role in determining relationships between soil characteristics and potential yield, which is a key aspect for forecasting in agronomy. The initial step in this analysis is the conversion of laboratory soil test results from traditional analog formats to digital ones, which allows for the integration of these data into broader information systems for further processing and analysis.

Soil data is collected using different techniques and may have different measurement scales, standardization is an important step that ensures data uniformity and compatibility. This means that all indicators must be reduced to a single scale, which allows them to be compared and combined in further analysis without the risk of distortions due to scale or measurement units.

Dataset preparation is a fundamental step that precedes neural network training, ensuring that the model is trained on relevant, accurate, and complete data. The main focus of this phase is the integration of the various data collected during the study into a single structured database that will be used for analysis and modeling.

First of all, it is necessary to combine the images obtained from drones or satellites with the corresponding results of soil analysis and yield data. This process requires an exact match between the location of the image and the geographic data of the soil samples, as well as the yield, to ensure the correctness of the subsequent analysis. Such integration requires care and detail, as inconsistencies in the data can lead to inaccuracies in the conclusions drawn from the analysis.

After creating an integrated data set, the next step is to divide it into different samples: training, validation and test. The training sample is used to train the model, the validation sample is used to fine-tune the hyperparameters and evaluate the portability of the model, and the test sample is used to evaluate its final performance. Allocation should ensure that each sample is representative represents the total population of data, avoiding biases or biases that could distort model training or evaluation results.

This approach to data preparation ensures that the neural network will be trained on reliable and high-quality data, which is critical for the accuracy and reliability of subsequent predictions.

The third stage in the research involves the development and training of a neural network, which is key to achieving the goal of yield prediction based on the analysis of images and soil data. At the beginning of this stage, there is the task of choosing the neural network architecture that best suits the specifics of our data and research goals. Convolutional Neural Networks (CNNs) are the optimal choice for image processing because they effectively detect and exploit spatial hierarchies in data, allowing for efficient field analysis and recognition of vegetation features, soil conditions, and other important attributes. For the analysis of soil data, where dependencies may be functional rather than spatial, fully connected networks can be useful because they are able to model complex nonlinear relationships between different soil characteristics.

After choosing the architecture, the next step is to train the model. This process involves feeding a training data set to the network, which adapts its weights and parameters to minimize the discrepancy between predicted and actual yield values. An important part of training is validation, which allows you to assess how well the model generalizes to data not seen during training, which occurs by fine-tuning the model's hyperparameters based on the results obtained from the validation data set.

The testing and validation stage is critical to verify the effectiveness and reliability of the developed neural network. This step helps to determine how well the model is able to predict yield based on the provided images and soil data, and whether it can be effectively applied in real-world conditions.

Model evaluation consists of using a test data set that was not used during training to test the model's performance. Comparing the yield values predicted by the model with the actual data allows you to assess the accuracy of the forecast. Using metrics such as mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2) help quantify how close model predictions are to true values and provide a means to compare the performance of different models or configurations of the same model.

Portability testing requires the model to be able to generalize what it learned on data that was not part of the training or test sets. This means that the model should perform well not only on data from one region or crop type, but also on data from other regions or for other crops. Such testing helps to understand how effective the model can be in different agronomic conditions.

Error analysis focuses on cases where the model makes significant errors. Detailed analysis of such cases is important to identify patterns or specific conditions under which the model fails to make an accurate prediction. Understanding the causes of these errors can point to ways to improve the model, such as by refining the architecture, changing data processing methods, or providing the model with additional data to train on.

The stage of refinement and optimization is an important step in the refinement of a neural network, which allows you to maximize its performance and prediction accuracy. This stage includes an iterative process of improving the model, based on the analysis of its previous results and identified shortcomings.

Model optimization involves making changes to the model's architecture or training parameters to improve its accuracy. This may include changing the number of neurons in the layers, adapting the learning rate, or changing the loss function. Optimization may also involve making changes to data reprocessing or using more sophisticated techniques to combat overtraining, such as regularization.

fine-tuning is the process of fine-tuning an already trained model to specific conditions or requirements. This may involve re-tuning the model weights on a new data set that has similar but not identical characteristics to the data on which the model was trained. Fine-tuning allows you to adapt the model to new conditions without having to start training from scratch, preserving the overall generalizability of the model while improving its performance in specific scenarios.

This stage not only increases the effectiveness of the model in the context of the tasks, but also helps to make the model more flexible and adaptable to various conditions, which is critical for its application in real agricultural conditions.

Further stages are beyond the scope of this study. At the sixth stage, it is necessary to integrate the developed model into agricultural management systems.

This integration will help support informed decision-making aimed at optimizing field cultivation, crop planning, and other important aspects of agronomic management. In addition, regular monitoring of the model's performance in real-world conditions is planned to ensure its relevance and accuracy, which may include updating the model as necessary by integrating new data.

At the seventh stage, the results of using the model were analyzed in order to understand its impact on productivity and efficiency of agricultural practices. Based on this analysis, reports will be prepared that will not only summarize the performance of the model, but also highlight key trends, successes, and identify potential areas for further research or improvement of the model.

The learning process of neural networks and the construction of corresponding graphs play a key role in understanding the effectiveness and dynamics of model training. Let's take a closer look at this process:

The learning process of neural networks begins with the initialization of the network weights. Random initialization was used.

Data are then iteratively fed to the input of the network, and using operations such as convolution (in CNNs) or recurrent transitions (in RNNs), they are passed through the network to obtain a prediction.

After receiving the output of the network, the loss function is calculated, which shows how much the forecast differs from the actual yield values.

Using a loss function, gradients are calculated for each of the network parameters, showing how the weights should be changed to reduce the error. The network weights are updated using the SGD optimization algorithm using the calculated gradients.

This process is repeated for each batch of training data, and subsequently for each epoch (a pass through the entire training data set).

Graphs are important for visualizing the learning process. They help track how loss and accuracy change with each epoch.

To train the crop yield monitoring model, based on the analysis of geodata and plot images, a comparison of yield forecasts of three crops: winter wheat, corn, and barley, harvested as of October 1, 2019, in the Chernihiv region was carried out. The reason for choosing the period 2018-2019 for forecasting is the lack of data on crops for 2020-2022 on the Public Cadastral Map [88]. All available multispectral images received from the SENTINEL-2B satellite [89] for the period from October 1 were used as input data 2018 to October 1, 2019. The data of the State Statistics Service of Ukraine [90] on the yield by region for 2019 is considered to be the exact value of yield.

Data of the State Statistics Service of Ukraine on yield by region and multispectral images obtained from the SENTINEL-2B satellite for 2011-2018 were used to find the parameters of the crop yield monitoring model based on the analysis of geodata and images.

The Sentinel-2B satellite, part of the European Space Agency's Copernicus program, provides detailed optical images of Earth's surface, contributing significantly to Earth observation and monitoring. The satellite carries a multispectral instrument (MSI) that captures data in 13 spectral bands with varying spatial resolutions. Here are the key characteristics of the Sentinel-2B image resolution and dataset:

Spectral Bands and Resolution:

Band 1 (Coastal aerosol): 60 meters
Band 2 (Blue): 10 meters
Band 3 (Green): 10 meters
Band 4 (Red): 10 meters
Band 5 (Vegetation red edge): 20 meters
Band 6 (Vegetation red edge): 20 meters
Band 7 (Vegetation red edge): 20 meters
Band 8 (NIR): 10 meters
Band 8a (Narrow NIR): 20 meters
Band 9 (Water vapor): 60 meters

Band 10 (SWIR - Cirrus): 60 meters
Band 11 (SWIR): 20 meters
Band 12 (SWIR): 20 meters

Sentinel-2B has a swath width of 290 km, which allows for significant coverage of the Earth's surface and aids in frequent revisit times.

The combined operation of Sentinel-2A and Sentinel-2B satellites provides a revisit time of 5 days at the equator and more frequent in higher latitudes, enabling timely monitoring of changes on the Earth’s surface.

Sentinel-2B data is freely available to the public, making it an invaluable resource for applications in agriculture, forestry, land use/land cover mapping, environmental monitoring, and disaster management.

The imagery provided by Sentinel-2B is of high radiometric quality, with 12-bit radiometric resolution allowing for the detection of many levels of data within each spectral band.

These characteristics make Sentinel-2B a robust tool for monitoring Earth's surface, providing critical data for environmental management, agricultural monitoring, and other applications.

When describing a Sentinel-2B dataset for one year, it's essential to consider the comprehensive and multi-dimensional nature of the data collected over this period. Here's an in-depth look at what such a dataset encompasses:

Temporal Coverage: Over the course of a year, Sentinel-2B provides frequent revisits to each point on Earth's surface, offering a temporal resolution that captures seasonal variations, changes in vegetation, and alterations in land use. The combined operation with Sentinel-2A enhances this frequency, ensuring detailed time-series data.

Spatial Coverage: The dataset would encompass images from all around the globe, with the satellite covering each point on Earth approximately every five days (in conjunction with Sentinel-2A). This results in multiple images of the same location over the year, valuable for tracking changes.

Each image in the dataset contains data across 13 spectral bands, offering detailed information that is not just limited to the visible spectrum. This multi-spectral data is crucial for various analyses, like vegetation health, water body monitoring, and urban development studies. The dataset includes high-resolution images, with spatial resolutions of 10, 20, and 60 meters depending on the spectral band. This allows for detailed observations and analyses of features on the Earth's surface A one-year dataset from Sentinel-2B constitutes a significant volume of data, given the satellite's high revisit rate and the detailed spectral information. Effective data management and processing systems are necessary to handle, store, and analyze this large amount of data.

The dataset supports a wide range of applications, from agricultural monitoring and forest management to urban planning and environmental conservation. Users can track seasonal changes, monitor crop health, detect deforestation, and observe urban expansion, among other uses.

The data from Sentinel-2B is processed into different levels, providing users with options ranging from raw data to processed images. The European Space Agency (ESA) ensures that the data is accessible to various users, from scientists to policymakers, facilitating its use in research, planning, and decision-making processes. Throughout the year, the dataset maintains high standards of data quality, with continuous calibration and validation to ensure accuracy and reliability of the information provided. In summary, a one-year dataset from Sentinel-2B offers a rich and multi-faceted view of the Earth's surface, providing invaluable insights for a myriad of applications and contributing significantly to our understanding and management of the planet's resources and changes

In the loss graph (Figure 4.2.2), you can see how the loss function decreases with each epoch, which indicates that the network is learning and the prediction error is decreasing.

The dataset was split into two distinct parts: training and testing, to evaluate the model's ability to interpret and predict based on spatial and temporal features captured in the satellite images.

The dataset comprised multispectral images from the Sentinel-2B satellite, covering the areas over one year. The images encompassed 13 spectral bands with different resolutions, providing comprehensive coverage of the Earth's surface features.

Training Set (80%): 80% of the dataset was allocated for training the CNN. This portion included a diverse range of images across different seasons and geographical locations to ensure the model learned a variety of patterns.

Testing Set (20%): The remaining 20% was reserved for testing the model. This set was carefully chosen to represent the diversity of the training set but consisted of images that the model had not previously seen.

The CNN was designed with several convolutional layers, pooling layers, and fully connected layers. Specific attention was given to the depth of the network and the size of the filters to optimize the model for high-resolution satellite image analysis. Preprocessing: Images were normalized to have values between 0 and 1, and augmentation techniques were applied to enhance the diversity of the training set. A batch size of 32 was chosen to balance the computational efficiency and model convergence rate. The SVG optimizer was used for its adaptive learning rate capabilities, which helped in faster convergence. Cross-entropy loss function was utilized to compute the difference between the predicted and actual outputs. The model was trained for 50 epochs, with early stopping implemented based on the validation loss to prevent overfitting. The CNN achieved an accuracy of 98% on the training set, indicating strong learning capabilities. On the testing set, the model demonstrated an accuracy of 95%, showcasing its ability to generalize well to unseen data.

The confusion matrix and other metrics like precision, recall, and F1-score were also calculated to provide a comprehensive view of the model's performance across different classes.

The high accuracy on the testing set suggests that the CNN effectively captured the spatial and temporal features inherent in the satellite images. However, there were challenges in classifying images with subtle changes or similarities, indicating areas for future improvement, possibly through deeper architectures or more sophisticated data augmentation techniques.

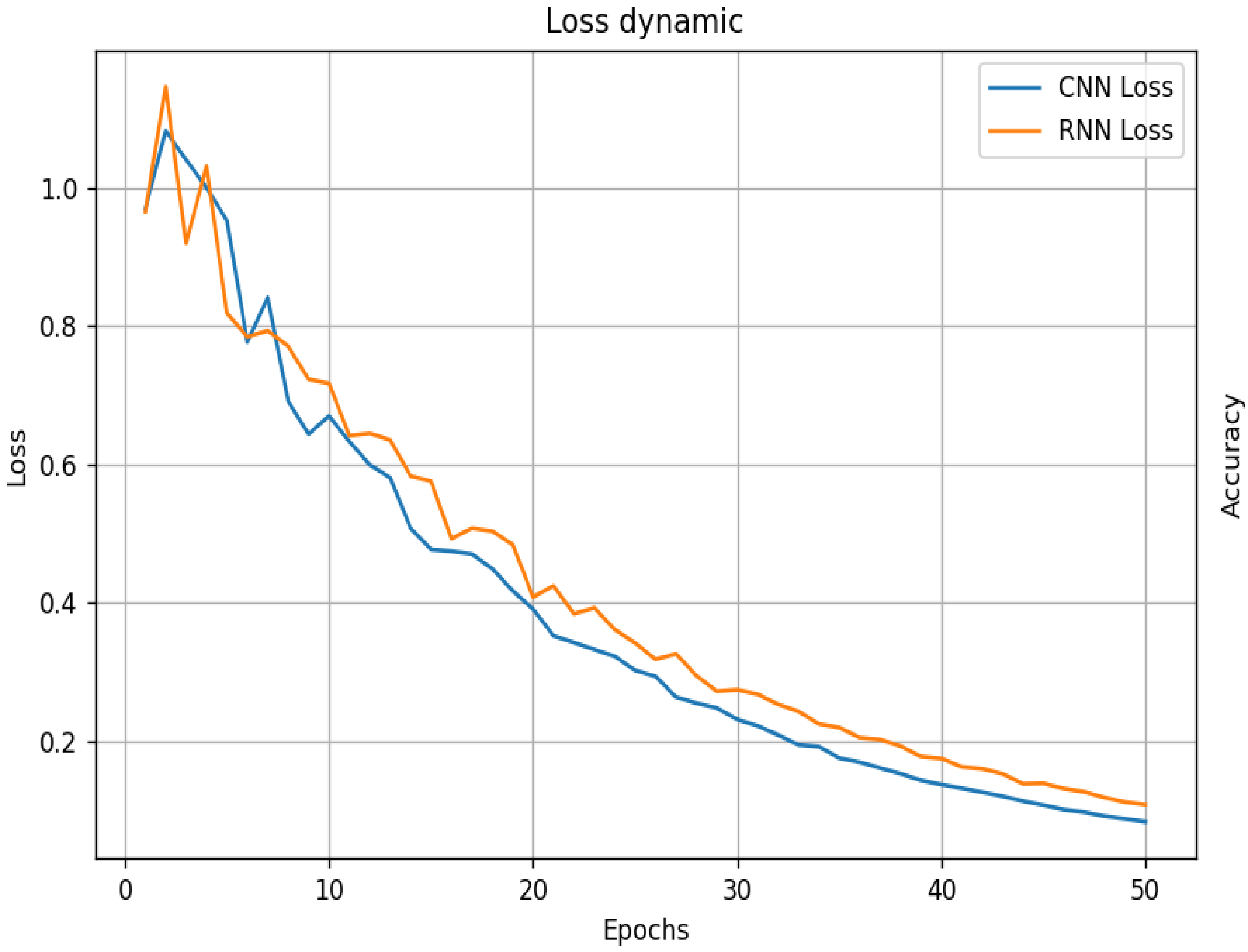


Figure 4.2.2 Loss graph

The training of a CNN with a one-year Sentinel-2B dataset split into training and testing sets demonstrated the potential of deep learning in interpreting satellite imagery

The accuracy graph (Figure 4.2.3) shows how the percentage of correct answers of the model on the training or validation data set increases with each epoch.

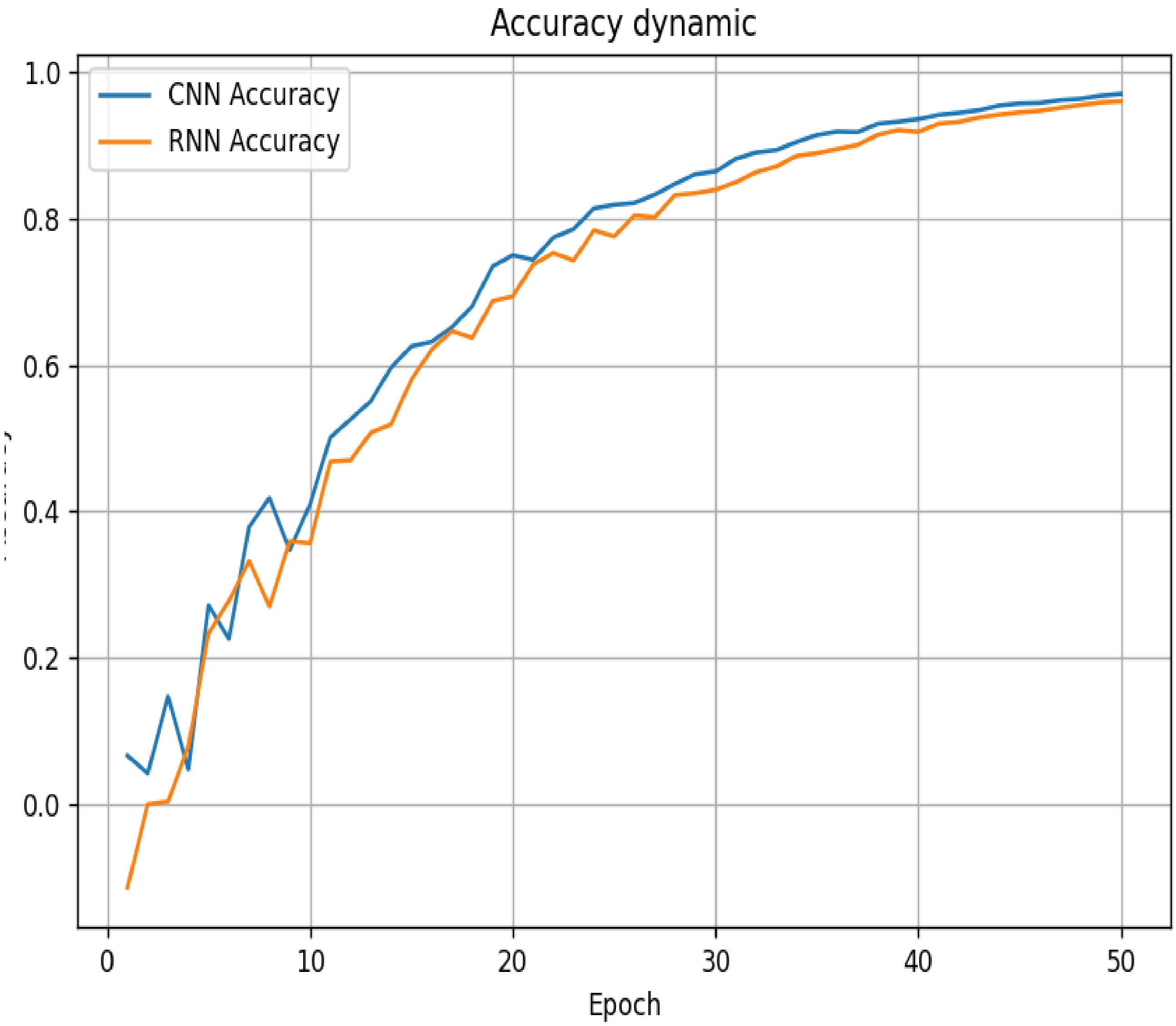


Figure 4.2.3 Accuracy graph

To understand the learning dynamics, but can also indicate problems such as overfitting or underfitting, allowing the developer to make timely adjustments to the learning process. The forecasting results are given in Table 4.2.1. The results make it possible to show that the proposed model for monitoring the yield of agricultural crops based on the analysis of geodata and site images provides a sufficiently accurate forecast. It can also be concluded that the yield is largely determined by the development of plants in the first months after emergence, therefore monitoring the condition of plants during this period is the most important.

Table 4.2.1. Crop yield forecast of Chernihiv region for 2019.

Forecasting method	Productivity, c/ha			Relative error, %	
	Winter wheat	Corn ore	Barley	Avg.	Max.
Yield according to DSSU data	48.2	80.0	41.2	0	0
Study results	49.9	81.1	41.8	2.12	3.53

From the conducted experiment, where we observed changes in loss and accuracy for a convolutional neural network (CNN) and a recurrent neural network (RNN) over 50 epochs, we can conclude that both models demonstrate positive learning dynamics. Assuming that the accuracy has reached a stable level with a deviation of only 2%, this may indicate the effectiveness of the selected training process.

4.3 Model for biomonitoring for project management in agrarian sphere validation.

To verify the crop yield monitoring model, based on the analysis of geodata and site images, a comparison of yield forecasts of three crops: winter wheat, corn, and barley, harvested as of October 1, 2019 in the Chernihiv region, was carried out. The reason for choosing the 2018-2019 period for forecasting is the lack of crop data for 2020-2022 on the Public Cadastral Map [86]. Forecasting was carried out in three ways.

The first method of forecasting consists in the use of a developed model for monitoring the yield of agricultural crops based on the analysis of geodata and images. All available multispectral images obtained from the SENTINEL-2B satellite [89] for the period from October 1, 2018 to October 1, 2019 were used as input data.

The second method of forecasting is to use the developed model, if all available multispectral images obtained from the SENTINEL-2B satellite for the period from March 1 to June 1, 2019 are used as input data.

The third method of forecasting is based on the use of the WOFOST simulation model [91], calculations were made using the WOFOST Control software Center 2.1. The data of the State Statistics Service of Ukraine [90] on the yield by region for 2019 is considered to be the exact value of yield.

Data of the State Statistics Service of Ukraine on yield by region and multispectral images obtained from the SENTINEL-2B satellite for 2011-2018 were used to find the parameters of the crop yield monitoring model based on the analysis of geodata and images.

Table 4.3.1. Crop yield forecast of Chernihiv region for 2019.

Forecasting method	Productivity, c/ha			Relative error, %	
	Winter wheat	Corn ore	Barley	Avg.	Max.
Yield according to DSSU data	48.2	80.0	41.2	0	0
WOFOST	50.3	82.1	42.8	3.62	4.35
Monitoring model, image for 12 months.	46.8	78.9	39.3	2.96	4.61
Monitoring model, image for 3 months.	46.2	79.1	38.7	4.51	6.06

The forecasting results are given in Table 4.2. The results make it possible to show that the proposed model for monitoring the yield of agricultural crops based on the analysis of geodata and site images provides a sufficiently accurate forecast. It can also be concluded that the yield is largely determined by the development of plants in the first months after emergence, therefore monitoring the condition of plants during this period is the most important.

Кадастровий номер: 7420887600:03:000:0007
Тип власності: Приватна власність
Цільове призначення: 01.01 Для ведення товарного сільськогосподарського виробництва для ведення товарного сільськогосподарського виробництва
Площа: 2.8791 га

Район: Борзнянський
КОАТУУ: 7420887600
Зона: 03
Квартал: 000

Область: Чернігівська

Figure 4.3.1. Setting the cadastral boundaries of plots on the portal of the Public Cadastral Map

A significant limitation of the proposed model is the low resolution of images. Thus, the research satellite EOS AM-1, which operates under the leadership of NASA, photographs the Earth's surface with a resolution of 250 m to 1 km. More modern satellites, such as SENTINEL-2B, photograph the earth's surface with a resolution of 20 m. However, this resolution is not sufficient to separate individual plants. The low-resolution leads to an error when weeds and other vegetation can be mistaken for agricultural crops. In order to reduce the error, the proposed model should be used to monitor the yield of sufficient crops of agricultural crops in a sufficiently large area.

Also, in the proposed model for monitoring the yield of agricultural crops based on the analysis of geodata and images, the influence of management decisions regarding the application of fertilizers, weeding, etc. is insufficiently taken into account. The data from the Public Cadastral Map [88] were used to establish the boundaries of the plots. This freely accessible tool provides data not only on the borders and owners of land plots (Fig. 4.3.1), but also allows to identify the plots on which winter and spring sowing of 19 agricultural crops were carried out (Fig. 4.3.2).

Кадастровий номер: 7420887600:03:000:0007
Тип власності: Приватна власність
Цільове призначення: 01.01 Для ведення товарного сільськогосподарського виробництва для ведення товарного сільськогосподарського виробництва
Площа: 2.8791 га
Замовити Витяг про земельну ділянку
Інформація про право власності та речові права
Запит на отримання документації із землеустрою

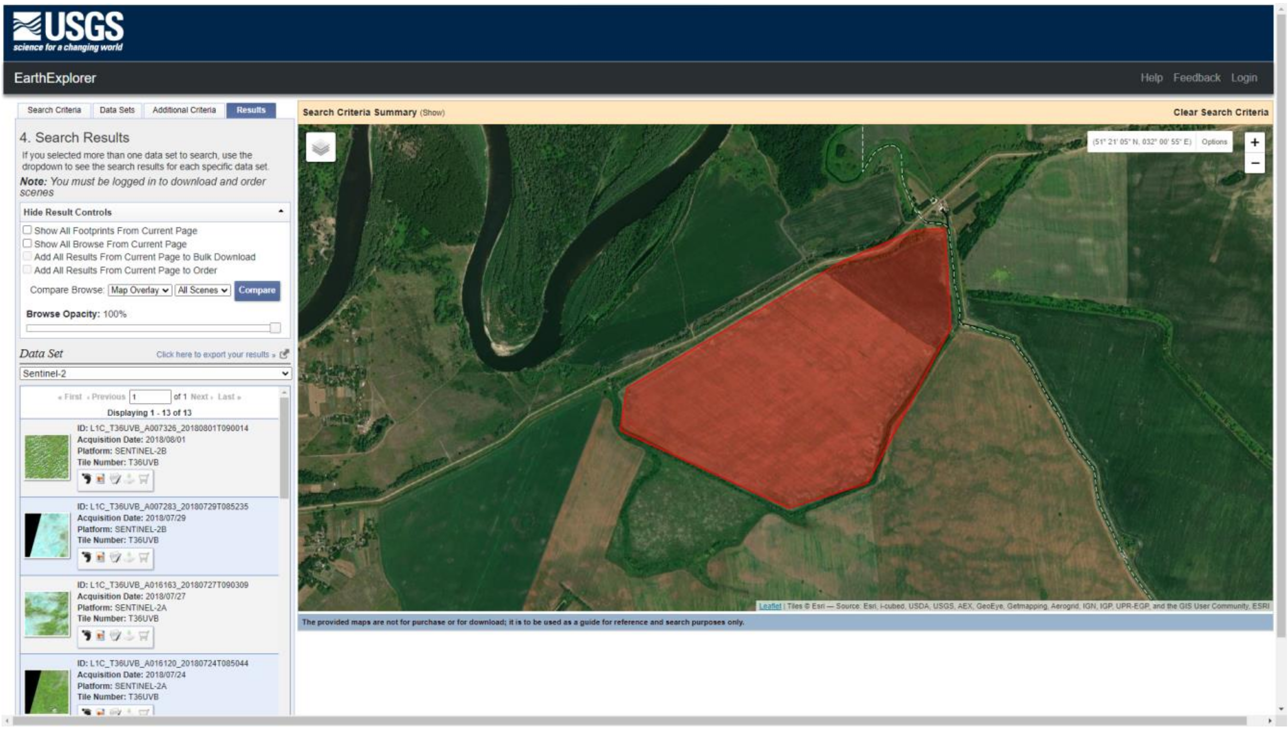
Район: Борзнянський
КОАТУУ: 7420887600
Зона: 03
Квартал: 000

Область: Чернігівська

Figure 4.3.2. Identification of winter and spring crops of plots on the portal of the Public Cadastral Map

Multispectral images were obtained from the SENTINEL-2B [89] satellite for 2015-2019. The first satellite of the optical earth monitoring mission was launched in June 2015, and the second satellite of the mission was launched in March 2017. The satellite provides images of the terrain taken in 13 bands from 443 to 2190 nm with a resolution of 20 m. The frequency of the images is from 2 to 5 days. To obtain data, open datasets posted on the USGS portal [92] were used (Fig. 4.3.3).

To determine the accuracy of yield forecasting data, data from the State Statistics Service of Ukraine [90] on the volume of production, yield, and harvested areas of agricultural crops by their types by region for 2011-2019 were used.



4. Search Results

If you selected more than one data set to search, use the dropdown to see the search results for each specific data set.

Note: *You must be logged in to download and order scenes*

☐ Show All Footprints From Current Page
☐ Show All Browse From Current Page
☐ Add All Results From Current Page to Bulk Download
☐ Add All Results From Current Page to Order

Compare Browse: Map Overlay, All Scenes

Browse Opacity: 100%

Data Set: Sentinel-2

Page 1 of 1

Displaying 1 - 13 of 13

ID: L1C_T36UVB_A007326_20180801T090014
Acquisition Date: 2018/08/01
Platform: SENTINEL-2B
Tile Number: T36UVB

ID: L1C_T36UVB_A007283_20180729T085235
Acquisition Date: 2018/07/29
Platform: SENTINEL-2B
Tile Number: T36UVB

ID: L1C_T36UVB_A016163_20180727T090309
Acquisition Date: 2018/07/27
Platform: SENTINEL-2A
Tile Number: T36UVB

ID: L1C_T36UVB_A016120_20180724T085044
Acquisition Date: 2018/07/24
Platform: SENTINEL-2A
Tile Number: T36UVB

The provided maps are not for purchase or for download; it is to be used as a guide for reference and search purposes only.

Figure 4.3.3. SENTINEL-2B mission data on the USGS portal.

Also, to validate the developed model, yield forecasting for the same plots was carried out using a simulation model for quantitative analysis of growth and production of annual field crops World Food Studies (WOFOST) version 7.1.7. Calculations were made using the WOFOST Control software Center 2.1[91].

Conclusions to chapter 4

An information system for yield monitoring has been developed, which uses GIS data to improve the efficiency of the agricultural sector. The modular structure of the system is described, including modules for data collection, storage, processing, visualization and analysis, including the use of machine learning for forecasting and optimization of processes.

The algorithm for implementing an integration model of artificial intelligence for yield monitoring based on a combination of multispectral images and geoinformation data, which includes seven stages: data collection and preparation, development of a neural network, its testing and validation, optimization, implementation in agrarian systems and further analysis of the results allows effective implement an artificial intelligence model for yield monitoring

The information system for monitoring the yield of agricultural crops was validated by the methods of comparative analysis. Comparison of winter wheat, corn, and barley yield forecasts in Chernihiv region for 2019 from forecasting using the WOFOST simulation model and data from the State Statistics Service of Ukraine on yield for 2019 show that the monitoring model is able to provide fairly accurate yield forecasts. It was found that yield is largely determined by plant development in the first 3 months after emergence, emphasizing the importance of monitoring during this period. The obtained practical results emphasize the potential and limitations of the use of yield monitoring information technology using geoinformation systems.

GENERAL CONCLUSIONS

1. Based on the analysis of the literature, the results revealed that none of the services or software combines all the necessary data processing and analysis capabilities that are necessary for the management of sown areas: finding anomalies, identifying phenological changes, estimating yield, etc. An important task is the creation of specialized software that would allow downloading and working with archives of large volumes of images and would have built-in methods of intelligent data processing and pattern recognition.

2. The theoretical basis of the research was created to solve the given task, which consists in the creation of information technology for monitoring the yield of agricultural crops based on the analysis of multispectral images obtained by remote sensing or by other methods. The geoinformation system created on the basis of this technology should monitor and forecast yield by analyzing time series of satellite images to identify quantitative and qualitative indicators of yield, possible plant diseases, etc. This task is especially relevant in conditions of environmental uncertainty. The tasks of geoinformation technologies, the classification of geoinformation systems, and the main methods underlying them are described. Considered concepts, which consist in ensuring the quality and quantity of the obtained agricultural products, as well as in ensuring the effective management of cultivated areas. Also considered are some software, services and devices that allow you to estimate the yield of agricultural products and process time series of images. Hashing, perceptual hashing, and segmentation methods used for processing time series images of agricultural fields are described. The definition of yield monitoring has been formed, under which we will understand the system of monitoring and measuring the state of growth of agricultural crops, taking into account meteorological, agrometeorological, phenological and other indicators based on the analysis of time series images obtained as a result of photographing sown areas, with the aim of evaluating and forecasting the potential crop yield.

3. The presentation of spatial data in GIS using raster and vector models is considered. In exploring the natural world, geographic information systems (GIS) serve as tools to collect and interpret Earth's phenomena, transforming complex data into models that mimic reality for informed decision-making. GIS combines raster and vector data to represent geographical features, where raster models capture visual data through pixels and vector models provide detailed representations via points, lines, and polygons. This integration allows for a nuanced analysis of spatial relationships and patterns, essential for understanding geographic contexts. By balancing the strengths of raster and vector models, GIS facilitates a comprehensive approach to visualizing and analyzing spatial information, supporting various applications in environmental management and planning.

4. The research culminated in the formulation of a Conceptual Research Model for the development of a geographical information system (GIS) tailored to agricultural needs. This model is meticulously structured into four distinct phases, each delineating a crucial segment of the research trajectory. In the initial phase, a comprehensive identification of the objects and subjects operating within the ecosystem of the GIS is conducted, alongside the articulation of the specifications for the resultant product and the delineation of anticipated outcomes. This systematic approach ensures a robust foundation for the subsequent stages of GIS development, underpinning the research with a clear and methodical framework.

5. A combined mathematical model was developed to display the relationship between phenological indicators and the yield of agricultural crops, and biomonitoring. This model combines the threshold adaptive method of determining the region belonging to the image of crop sowing and the method of forecasting the values of the time series of phenological indicators based on the selection of its trend, seasonal and random components. The Otsu method was adapted to determine the adaptive threshold value.

6. The process of creating and training a hybrid neural network that integrates image data and soil information for yield prediction is described. It has been

established that the task requires the use of a combined type neural network for effectively finding a solution. A network architecture is proposed, which for receiving input data includes convolutional neural networks for image processing and fully connected layers for soil data analysis. 9 parameters of the soil, which are important for the task, have been determined. This integration allows the network to consider a variety of information, increasing its ability to accurately predict yield. The output layer of the first stage contains neurons that store the integrated seed values. In the second stage, the network uses recurrent neural networks to analyze data sequences, which adds the ability to take into account temporal dependencies and context. Proposed procedures of two-stage training with a teacher. The backpropagation method combined with stochastic gradient descent to update the weights was used to train the CNN. At the second stage, the Backpropagation method was applied Through Time in conjunction with the Gradient technique Clipping.

7. An information system for yield monitoring has been developed, which uses GIS data to improve the efficiency of the agricultural sector. The modular structure of the system is described, including modules for data collection, storage, processing, visualization and analysis, including the use of machine learning for forecasting and optimization of processes. The algorithm for implementing an integration model of artificial intelligence for yield monitoring based on a combination of multispectral images and geoinformation data, which includes seven stages: data collection and preparation, development of a neural network, its testing and validation, optimization, implementation in agrarian systems and further analysis of the results allows effective implement an artificial intelligence model for yield monitoring

8. The information system for monitoring the yield of agricultural crops was validated by methods of comparative analysis. A comparison of winter wheat, corn, and barley yield forecasts in Chernihiv region for 2019 from forecasting using the WOFOST simulation model and data from the State Statistics Service of Ukraine on

yield for 2019 shows that the monitoring model is able to provide fairly accurate yield forecasts. It was found that yield is largely determined by plant development in the first 3 months after emergence, emphasizing the importance of monitoring during this period. The obtained practical results emphasize the potential and limitations of the use of yield monitoring information technology using geoinformation systems.

9. The research underscores the pivotal role of information technology in yield monitoring as an essential foundational step for project management within the agrarian sector. By integrating advanced geoinformation systems in agriculture, stakeholders can significantly enhance decision-making processes, optimize resource allocation, and improve overall project outcomes. This synergy between information technology and agricultural management not only streamlines operations but also sets a new benchmark for efficiency and productivity in the field, establishing a vital link between data-driven insights and effective agricultural practices.

REFERENCES

1. Mingxin, Huang, & Vatskel, V. (2019). Digital image analysis technologies for decision support systems in agricultural. Management of development of complex systems, 37, 164 – 167, https://doi.org/10.6084/m9.figshare.9783227
2. Mingxin, Huang. (2019). Review of monitoring and forecasting tools of the crop yield. Management of development of complex systems, 38, 161 – 167, https://doi.org/10.6084/m9.figshare.9788696
3. Huang, M., & Shabala, Y. Y. (2019). Conceptual model of geogerphic information system for agriculture. Scientific Bulletin of Uzhhorod University. Series of Mathematics and Informatics, 2(35), 149–155. https://doi.org/10.24144/2616-7700.2019.2(35).149-155
4. Mingxin, Huang. (2024). Development of a yield monitoring model based on analysis of surveys and images of the field. Management of development of complex systems, 57, 67 – 71. https://doi.org/10.32347/2412-9933.2024.57.67-71
5. Mingxin, Huang. (2020). Use of geoinformation systems for agricultural problems. Science Journal Innovation Technologies Transfer. 61-64. https://doi.org/10.36381/iamsti.4.2020.61-64
6. Mingxin, Huang. (2018). Use of geoinformation systems in agriculture. V International Scientific and Practical Conference "Information Technologies and Interactions", November 20-21, 2018, 63.
7. Mingxin, Huang. (2019). Information technology for efficient management of crop yields. XV International Scientific and Practical Conference "Project Management in the Development of Society", May 17-18, 2019, 40-41
8. Mingxin, Huang. (2019). Use of digital images if geographical areas for thr purposes of agricultural. I International Scientific and Practical Conference IMTCK2019, 65-66

9. Mingxin, Huang. (2019). Participatory sensing for monitoring and forecasting of environmental pollution. VI International Scientific and Practical Conference "Information Technologies and Interactions", December 20, 2019, 95-96.

10. Mingxin, Huang. (2020). Using Geoinformation Systems For Agriculture Tasks. Seventh international scientific-practical conference «Management of the development of technologies» Topic: "Information technology development of educational content» Kyiv, 25 – 26 March 2020, 127-128. [In Ukrainian]

11. Huang, M., Biloshchytskyi, A., Andrashko, Y. & Omirbayev, S. (2021). A Conceptual Model for Diversification Strategies Choice, 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 1-5, https://doi.org/10.1109/SIST50301.2021.9465973 [Scopus, Web of Science]

12. Food and Agriculture Organization of the United Nations. Retrieved from: http://www.fao.org/home/en/

13. Gueririni, F. (2015) The future of agriculture? Smart Farming. Retrieved from: https://www.forbes.com/sites/federicoguerrini/2015/02/18/the-future-of-agriculture-smart-farming/#84d0ef13c42c [in English].

14. About ArcGIS. The mapping and analytics platform. Retrieved from: http://www.esri.com/software/arcgis

15. DeJoia A., Duncan M. (2015). What is Precision Agriculture and why is it important Retrieved from: https://soilsmatter.wordpress.com/2015/02/27/what-is-precision-agriculture-and-why- is-it-important/.

16. Kuchansky A., Biloshchytskyi A., Andrashko Yu., Biloshchytska S., Shabala Ye., Myronov O. (2018). Development of adaptive combined models for predicting time series based on similarity identification. Eastern-European Journal of Enterprise Technologies. 2018. Vol. 1/4 (91). P. 32–42. DOI: 10.15587/1729-4061.2018.121620.

17. Petitjean F., Inglada J., Gançarski P. Satellite image time series analysis under time warping IEEE Trans. Geosci. Remote Sens. 2012. 50 (8). 3081-3095.

18. Eerens H., Haesen D., Rembold F., Urbano F., Tote C., Bydekerke L. Image time series processing for agriculture monitoring. Environmental Modelling and Software. 2014, 53, 154–162.

19. Drone data collection and analytics for agriculture. Quantify plant and soil health, improve productivity and maximize field output. Retrieved from: https://www.precisionhawk.com/agriculture.

20. Oganov A.V., Gogunsky V.D. Use the Theory of Constrains in PMO implementation at the organization.GESJ. Computer Science and Telecommunications. 2013. 4(40).59-65

21. Morozov, V., Kalnichenko, O., & Liubyma, I. (2017). Managing projects configuration in development distributed information systems. 2nd IEEE International Conference on Advances Information and Communication. 154–157. 10.1109/aiact.2017.8020088

22. Kuchansky A., Biloshchytskyi A., Andrashko Yu., Vatskel V., Biloshchytska S., Danchenko O., Vatskel I. (2018). Combined models for forecasting the air pollution level in infocommunication systems for the environment state monitoring. 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS). 125–130. doi: 10.1109/IDAACS-SWS.2018.8525608

23. Babu A. Jagadeesh, Thirumalaivasan D., Venugopal K. (2006). STAO: a component architecture for raster and time series modeling. Environmental Modelling & Software. 21(5). 653-664.

24. Atzberger, C., (2013). Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens. 5 (2), 949-981.

25. Baruth, B., Royer, A., Klisch, A., & Genovese, G., (2008). The use of remote sensing within the MARS Crop Yield Monitoring System of the European Commission, pp. 935-941. ISPRS, Commission VIII, Stresa

26. Zhao, G., Bryan, B.A., King, D., Luo, Z., Wang, E., Bende-Michl, U., Song, X., Yu, Q., (2013). Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining grid computing and parallel processing. Environ. Model. Softw. 41 (3), 231-238.

27. Cracknell, A.P. (1997). The Advanced Very High Resolution Radiometer (AVHRR). Taylor & Francis, London, Bristol (PA).

28. Maisongrande, P., Duchemin, B., & Dedieu, G. (2004). VEGETATION/SPOT: an operational mission for the Earth monitoring; presentation of new standard products. Int. J. Remote Sens. 25, 9-14

29. Justice, C.O., Vermote, E., Townshend, J.R.G., Defries, R., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J.L., Riggs, G., Strahler, A., Lucht, W., Myneni, R.B., Knyazikhin, Y., Running, S.W., Nemani, R.R., Wan, Z.M., Huete, A.R., van Leeuwen, W., Wolfe, R.E., Giglio, L., Muller, J.P., Lewis, P., & Barnsley, M.J. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. IEEE Trans. Geosci. Remote Sens. 36 (4), 1228-1249.

30. Fensholt, R., Anyamba, A., Huber, S., Proud, S.R., Tucker, C.J., Small, J., Pak, E., Rasmussen, M.O., Sandholt, I., & Shisanya, C. (2011). Analysing the advantages of high temporal resolution geostationary MSG SEVIRI data compared to Polar Operational Environmental Satellite data for land surface monitoring in Africa. Int. J. Appl. Earth Obs. Geoinformation 13 (5), 721-729.

31. Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sens. 5 (4), 1704-1733

32. Data is Vital We provide the data you need to make sound decisions. Retrieved from: https://www.gminsights.com/

33. Rasinmäki, J. (2003). Modelling spatio-temporal environmental data. Environmental Modelling & Software, 18(10), 877-886.

34. Babu, A. J., Thirumalaivasan, D., & Venugopal, K. (2006). STAO: a component architecture for raster and time series modeling. Environmental Modelling & Software, 21(5), 653-664.

35. MATLAB for Artificial Intelligence. Retrieved from: http://www.mathworks.it/

36. The R Project for Statistical Computing. Retrieved from: https://www.r-project.org/

37. Bringing advanced geospatial technologies to the world. Retrieved from: http://grass.osgeo.org/

38. TerrSet 2020 Software Features. Retrieved from: http://clarklabs.org/

39. Fully Benefit From the Value of Remotely Sensed Data. Retrieved from: https://www.nv5geospatialsoftware.com/

40. Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., Harlan, J.C., (1974). Monitoring the Vernal Advancements and Retro Gradation of Natural Vegetation. Greenbelt, MD, USA, p. 371.

41. JRC Science Hub - European Commission. Retrieved from: http://www.marsop.info/marsop3 /

42. USGS FEWS NET Data Portal. Retrieved from: http://earlywarning.usgs.gov/fews/

43. Becker-Reshef, I., Justice, C., Sullivan, M., Vermote, E., Tucker, C., Anyamba, A., Small, J., Pak, E., Masuoka, E., Schmaltz, J., Hansen, M., Pittman, K., Birkett, C., Williams, D., Reynolds, C., Doorn, B., 2010. Monitoring global croplands with coarse resolution earth observations: the Global Agriculture Monitoring (GLAM) Project. Remote Sens. 2 (6), 1589-1609.

44. Crop Explorer. Retrieved from: https://ipad.fas.usda.gov/cropexplorer/

45. Blower, J.D., Gemmell, A.L., Griffiths, G.H., Haines, K., Santokhee, A., & Yang, X., (2013). A Web Map Service implementation for the visualization

of multidimensional gridded environmental data. Environ. Model. Softw. 47 (9), 218-224

46. Díaz, L., Bröring, A., McInerney, D., Libertá, G., & Foerster, T., (2013). Publishing sensor observations into Geospatial Information Infrastructures: a use case in fire danger assessment. Environ. Model. Softw. 48 (10), 65-80.

47. Dubois, G., Schulz, M., Skøien, J., Bastin, L., Peedell, S., 2013. eHabitat, a multipurpose Web Processing Service for ecological modeling. Environ. Model. Softw. 41 (3), 123-133.

48. Jönsson, P., Eklundh, L., 2004. TIMESAT e a program for analysing time-series of satellite sensor data. Comput. Geosci. 30, 833-845.

49. Welcome to the Climate Hazards Center (CHC). Retrieved from: https://www.chc.ucsb.edu/

50. AgrometShell (AMS) Software for crop yield forecasting initiated by the Food and Agriculture Organization of the United Nations. Retrieved from: http://www.hoefsloot.com/agrometshell.htm

51. Gavahi, K., Abbaszadeh, P., & Moradkhani, H. (2021). DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. Expert Systems with Applications, 184, 115511

52. Matsumura, K., Gaitan, C. F., Sugimoto, K., Cannon, A. J., & Hsieh, W. W. (2015). Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. The Journal of Agricultural Science, 153(3), 399-410

53. Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in plant science, 10, 452963

54. Guo, W. W., & Xue, H. (2014). Crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models. Mathematical Problems in Engineering, 2014

55. Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. Remote Sensing of Environment, 259, 112408

56. Tian, H., Wang, P., Tansey, K., Zhang, J., Zhang, S., & Li, H. (2021). An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. Agricultural and Forest Meteorology, 310, 108629.

57. Tian, H., Wang, P., Tansey, K., Zhang, S., Zhang, J., & Li, H. (2020). An IPSO-BP neural network for estimating wheat yield using two remotely sensed variables in the Guanzhong Plain, PR China. Computers and Electronics in Agriculture, 169, 105180.

58. Li, L., Wang, B., Feng, P., Wang, H., He, Q., Wang, Y., ... & Yu, Q. (2021). Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China. Agricultural and Forest Meteorology, 308, 108558.

59. Liang, B., Liu, H., Quine, T. A., Chen, X., Hallett, P. D., Cressey, E. L., ... & Hartley, I. P. (2021). Analysing and simulating spatial patterns of crop yield in Guizhou Province based on artificial neural networks. Progress in Physical Geography: Earth and Environment, 45(1), 33-52.

60. Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sensing, 12(2), 236.

61. White, M. A, & Nemani, R.R. (2006). Real-time monitoring and short-term forecasting of land surface phenology." Remote Sensing of Environment, 104(1), 43-49.

62. Reed, B.C., White, M., &Brown, J.F. (2003). Remote sensing phenology." Phenology: an Integrative Environmental Science, 39, 365-381.

63. Verbesselt J, Hyndman R, Newnham G, Culvenor D (2010). Detecting trend and seasonal changes in satellite image time series." Remote Sensing of Environment, 114(1), 106-115.

64. Verbesselt J, Hyndman R, Zeileis A, Culvenor D (2010). Phenological Change Detection while Accounting for Abrupt and Gradual Trends in Satellite Image Time Series Remote Sensing of Environment, 114(12), 2970-2980.

65. Charles A. Poynton (2003). Digital Video and HDTV: Algorithms and Interfaces. Morgan Kaufmann. ISBN 1558607927.

66. Loeffler, C., Ligtenberg, A., & Moschytz, G. S. (1989, May). Practical fast 1-D DCT algorithms with 11 multiplications. In International Conference on Acoustics, Speech, and Signal Processing, (pp. 988-991). IEEE.

67. Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83, 195–213.

68. Fulton, J.P.; C.J. Sobolik; S.A. Shearer; S.F. Higgins; T.F. Burks (2009). Grain yield monitor flow sensor for accuracy for simulated varying field slopes. 25 (1). Applied in Engineering in Agriculture: 15–21.

69. Atherton, B.C.; M.T. Morgan; S.A. Shearer; T.S. Stombaugh; A.D. Ward (1999). "Site-Specific Farming: A Perspective on Information Needs, Benefits and Limitations". Soil and Water Conservation Society. 54 (Second Quarter): 455–461.

70. P. Revesz, R.Chen, P. Kanjamala, Y. Li, Y. Liu, and Y. Wang, “The MLPQ/GIS constraint database system,” In Proceedings of the 2000 ACM SIGMOD international conference on Management of data, 601 p., 2000

71. Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66. http://doi.org/10.1109/TSMC.1979.4310076

72. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. Pattern recognition, 77, 354-377.

73. Sun, M., Song, Z., Jiang, X., Pan, J., & Pang, Y. (2017). Learning pooling for convolutional neural network. Neurocomputing, 224, 96-104.

74. Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015, April). Convolutional, long short-term memory, fully connected deep neural networks. In

2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4580-4584). Ieee.

75. Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388-427.

76. Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. Water, 11(7), 1387.

77. Wang, X., Xu, J., Shi, W., & Liu, J. (2019, October). OGRU: An optimized gated recurrent unit neural network. In Journal of Physics: Conference Series (Vol. 1325, No. 1, p. 012089). IOP Publishing.

78. He, J., Li, L., Xu, J., & Zheng, C. (2018). ReLU deep neural networks and linear finite elements. arXiv preprint arXiv:1807.03973.

79. Xu, J., Li, Z., Du, B., Zhang, M., & Liu, J. (2020, July). Reluplex made more practical: Leaky ReLU. In 2020 IEEE Symposium on Computers and communications (ISCC) (pp. 1-7). IEEE.

80. Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289.

81. Kalaiselvi, T., Padmapriya, S. T., Somasundaram, K., & Praveenkumar, S. (2022). E-Tanh: a novel activation function for image processing neural network models. Neural Computing and Applications, 34(19), 16563-16575.

82. Rojas, R., & Rojas, R. (1996). The backpropagation algorithm. Neural networks: a systematic introduction, 149-182.

83. Zhang, S., Choromanska, A. E., & LeCun, Y. (2015). Deep learning with elastic averaging SGD. Advances in neural information processing systems, 28.

84. Li, M., Zhang, T., Chen, Y., & Smola, A. J. (2014, August). Efficient mini-batch training for stochastic optimization. In Proceedings of the 20th ACM

SIGKDD international conference on Knowledge discovery and data mining (pp. 661-670).

85. Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 78(10), 1550-1560.

86. Liu, M., Zhuang, Z., Lei, Y., & Liao, C. (2022). A communication-efficient distributed gradient clipping algorithm for training deep neural networks. Advances in Neural Information Processing Systems, 35, 26204-26217.

87. PostGIS. Retrieved at: https://postgis.net/

88. Public cadastral map. (2019). Retrieved at: https://map.land.gov.ua

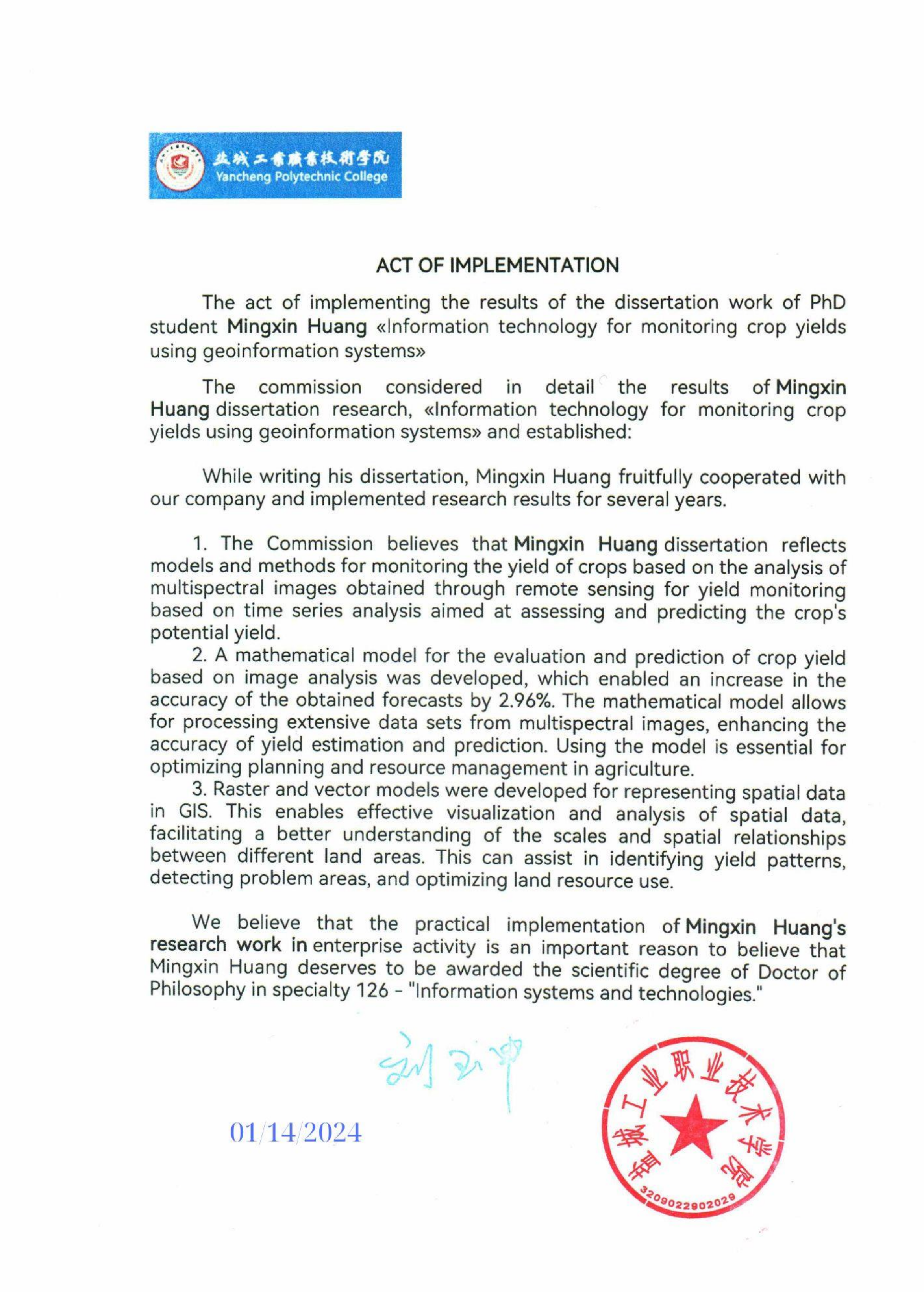
89. Copernicus: Sentinel-2 - The Optical Imaging Mission for Land Services. (2021). Retrieved at: https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-2

90. The volume of production, yield and the area of agricultural crops collected by their species. (2021). Retrieved at: http://www.ukrstat.gov.ua/operativ/operativ2021/sg/ovuzpsg/Arh_ovuzpsg_2021_e.html

91. WOFOST - WOrld FOod STudies. (2021). Retrieved at: https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental-Research/Facilities-Tools/Software-models-and-databases/WOFOST.htm

92. USGS science for a changing world. (2022). Retrieved at: https://earthexplorer.usgs.gov/

APPENDIX A. ACT OF IMPLEMENTATION



盐城工业职业技术学院
Yancheng Polytechnic College

ACT OF IMPLEMENTATION

The act of implementing the results of the dissertation work of PhD student **Mingxin Huang** «Information technology for monitoring crop yields using geoinformation systems»

The commission considered in detail the results of **Mingxin Huang** dissertation research, «Information technology for monitoring crop yields using geoinformation systems» and established:

While writing his dissertation, Mingxin Huang fruitfully cooperated with our company and implemented research results for several years.

1. The Commission believes that **Mingxin Huang** dissertation reflects models and methods for monitoring the yield of crops based on the analysis of multispectral images obtained through remote sensing for yield monitoring based on time series analysis aimed at assessing and predicting the crop's potential yield.
2. A mathematical model for the evaluation and prediction of crop yield based on image analysis was developed, which enabled an increase in the accuracy of the obtained forecasts by 2.96%. The mathematical model allows for processing extensive data sets from multispectral images, enhancing the accuracy of yield estimation and prediction. Using the model is essential for optimizing planning and resource management in agriculture.
3. Raster and vector models were developed for representing spatial data in GIS. This enables effective visualization and analysis of spatial data, facilitating a better understanding of the scales and spatial relationships between different land areas. This can assist in identifying yield patterns, detecting problem areas, and optimizing land resource use.

We believe that the practical implementation of **Mingxin Huang's research work in** enterprise activity is an important reason to believe that Mingxin Huang deserves to be awarded the scientific degree of Doctor of Philosophy in specialty 126 - "Information systems and technologies."

01/14/2024

3209022902029

APPENDIX B. LIST OF THE APPLICANT'S PUBLICATIONS ON THE THEME OF THE DISSERTATION AND INFORMATION ON THE APPROVAL OF THE RESULTS OF THE DISSERTATION

Articles in professional publications of Ukraine (included in the list of the Ministry of Education and Science of Ukraine)

1. **Mingxin, Huang,** & Vatskel, V. (2019). Digital image analysis technologies for decision support systems in agricultural. Management of development of complex systems, 37, 164–167, https://doi.org/10.6084/m9.figshare.9783227 [Category B]
2. **Mingxin, Huang**. (2019). Review of monitoring and forecasting tools of the crop yield. Management of development of complex systems, 38, 161 – 167, https://doi.org/10.6084/m9.figshare.9788696 [Category B]
3. **Huang, M.,** & Shabala, Y. Y. (2019). Conceptual model of geographic information system for agriculture. Scientific Bulletin of Uzhhorod University. Series of Mathematics and Informatics, 2(35), 149–155. https://doi.org/10.24144/2616-7700.2019.2(35).149-155 [Category B]
4. **Mingxin, Huang.** (2024). Development of a yield monitoring model based on analysis of surveys and images of the field. Management of development of complex systems, 57, 67 – 71. https://doi.org/10.32347/2412-9933.2024.57.67-71 [Category B]

Articles in professional publications of Ukraine

(not included in the list of the Ministry of Education and Science of Ukraine)

1. **Mingxin, Huang.** (2020). Use of geoinformation systems for agricultural problems. *Science Journal Innovation Technologies Transfer*. 61-64. https://doi.org/10.36381/iamsti.4.2020.61-64

Approbation works

1. **Mingxin, Huang**. (2018). Use of geoinformation systems in agriculture. V International Scientific and Practical Conference "Information Technologies and Interactions", November 20-21, 2018, 63.

2. **Mingxin, Huang**. (2019). Information technology for efficient management of crop yields. . XV International Scientific and Practical Conference "Project Management in the Development of Society", May 17-18, 2019, 40–41

3. **Mingxin, Huang**. (2019). Use of digital images if geographical areas for thr purposes of agricultural. I International Scientific and Practical Conference IMTCK2019, 65–66.

4. **Mingxin, Huang**. (2019). Participatory sensing for monitoring and forecasting of environmental pollution. VI International Scientific and Practical Conference "Information Technologies and Interactions", December 20, 2019, 95-96.

5. **Mingxin, Huang.** (2020). Using Geoinformation Systems For Agriculture Tasks. *Seventh international scientific-practical conference «Management of the development of technologies» Topic: "Information technology development of educational content» Kyiv, 25 – 26 March 2020*, 127-128. [In Ukrainian]

6. **Huang, M.,** Biloshchytskyi, A., Andrashko, Y. & Omirbayev, S. (2021). A Conceptual Model for Diversification Strategies Choice, *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Nur-Sultan, Kazakhstan, 1–5, https://doi.org/10.1109/SIST50301.2021.9465973 [**Scopus, Web of Science**]