Ministry of Education and Science of Ukraine

# ODESA NATIONAL UNIVERSITY OF TECHNOLOGY

International Competition of Student Scientific Works

# BLACK SEA SCIENCE 2023

# PROCEEDINGS

ODESA, ONUT 2023

Ministry of Education and Science of Ukraine

Odesa National University of Technology

International Competition of Student Scientific Works

# BLACK SEA SCIENCE 2023

## Proceedings



Odesa, ONUT
2023

# INTELLIGENT TECHNOLOGIES OF ASSESSMENT AND MODELING OF THE DEVELOPMENT OF AGRICULTURAL CROPS USING STREAMING DATA PROCESSING OF SATELLITE IMAGERY

**Author:** Dmytro Buts
**Advisors:** Oleksandr Fomin
Oleksii Diachenko
Odesa State Agrarian University (Ukraine)

***Abstract.*** *Agribusiness in Ukraine has reached a certain maturity, which is evidenced by the stabilization of the level of investments in agriculture and the growth of competition among producers of agricultural products until the beginning of 2022.*

*In the agro-industrial complex of Ukraine, the need for the volume and quality of the use of modern information technologies, in particular systems for collecting, storing and processing data from satellites, unmanned aerial vehicles, sensors of mobile weather stations, has significantly increased. At the same time, both the amount of data and the need for their quality processing and reliable conclusions that can be relied on when making decisions are increasing. As a result, there is a demand for industrial analytical systems and, in particular, advanced analytics.*

*Within the framework of this project, it is proposed to create a hardware and software information system for evaluating and modeling the development of agricultural crops using stream processing of satellite survey data in order to increase the efficiency and reliability of decision-making processes, diagnose the state of agronomic crops in large areas, optimize the management of agricultural resources, while increasing product quality, bringing significant economic benefits and saving agricultural resources.*

*The project corresponds to the theme of scientific research and scientific and technical (experimental) developments of the Ministry of Education and Science of Ukraine for 2022-2026 (item 23): perspective technologies of the agro-industrial complex and processing industry.*

***Keywords:*** *information system, satellite imaging, diagnostics, agronomic culture, artificial intelligence, precision agriculture.*

## I. INTRODUCTION

Land use monitoring is an important task in modern agriculture. Monitoring results are not limited to crop surveillance: they have a wide range of applications from monitoring the condition of agricultural resources, forecasting area and yield, crop assessment to planning harvesting activities and crop preservation [1, 2]. At the same time, the problem of identification of vegetative processes on large agricultural areas during growing cycles is one of the main ones [3].

Classification of the state of agricultural territories is one of the most important agricultural tasks in the world [4]. Due to the increase in food needs (for example, population growth) and the reduction of cultivated areas (for example, the expansion of cities) in countries importing agricultural products, growing plants requires more and more intensive methods, including triple sowing, the use of modern varieties of seeds, pesticides and fertilizers [5 ]. However, such intensive cultivation can lead to environmental problems such as water pollution, soil degradation, and microbial damage [6], threatening global food security and national economies.

As a result, those responsible for food safety are looking for solutions for sustainable agricultural production, starting with crop area monitoring and planning to obtain highly accurate information about the condition and location of crops at the level of individual fields in regions of interest [7, 8]. Indeed, in-depth knowledge such as the frequency of price increases and the estimation of production areas can help the government in formulating policies and justifying exports.

In the agro-industrial complex of Ukraine, the need for the volume and quality of the use of modern information technologies, in particular systems for collecting, storing and processing data from satellites, unmanned aerial vehicles, sensors of mobile weather stations, has significantly increased. The topicality of the topic is emphasized by current industry normative documents, in particular, the topic of scientific research and scientific and technical (experimental) developments of the Ministry of Education and Science of Ukraine for 2022-2026 (paragraph 23): perspective technologies of the agro-industrial complex and processing industry.

Under such conditions, the implementation of a software-hardware information system for evaluating and modeling the development of agricultural crops using streaming data from satellite imagery makes it possible to increase the efficiency and reliability of decision-making processes, diagnose the state of agronomic crops in large areas, and optimize the management of agricultural resources, which at the same time increases the quality products, provides significant economic benefits and savings of agricultural resources.

## II. LITERATURE ANALYSIS

Traditional systems for monitoring and planning vegetative processes are based on field observations [9]. This method of data collection requires an extremely high expenditure of time, money and human resources. In addition, agricultural areas are not static: they are subject to cycles of cultivation of different types of crops, weather conditions and climate change [10]. Therefore, methods of monitoring and planning vegetative processes based on field observations are too resource-intensive and slow.

Recently, monitoring systems in agriculture have been developed based on modern technologies of remote sensing using satellite data, which are able to meet the existing requirements for the cost and speed of mapping agricultural areas.

### 2.1. Satellite survey platforms

Satellite imagery is often inexpensive or even free has a wide spatial range covering a large geographic area, and has high temporal resolution (available throughout the year).

Remote sensing is a profitable solution for mapping land resources worldwide due to cheaper data collection compared to field work [5-7, 11]. At the same time, optical satellite sensors such as SPOT (Spot Image, France), Sentinel (European satellite system), and Landsat (National Oceanic and Atmospheric Administration, USA) are widely used. These sensors distinguish between land and crop areas, observing the Earth's surface mainly in the spectral range of 0.4-2.5 μm.

It is known that the use of several data sources, including radar time series, significantly improves the accuracy of mapping [11]. This approach is widely used around the world as a composition of satellite images in several spectral bands (for example, red, blue, and infrared).

In addition, existing methods of mapping agricultural areas based on satellite images often use artificially created spectral characteristics, namely vegetation indices.

### 2.2. Spectral characteristics

Due to the multispectral nature of satellite images, scientists have proposed different image indices for differentiating cultivated areas from non-cultivated areas [6, 7], vegetation states of plants, and conditions of their development. However, these indexes require a great deal of expert knowledge in manual processing. Existing technologies of automatic image processing, such as VGG, often work poorly on spectral images [9] but give significant errors due to the lack of statistical information and expert assessments for the territory of Ukraine.

The problems of remote sensing in general and satellite images, in particular, are different. First, although satellite images cover a large geographic area, they often have a relatively low spatial resolution, especially for older-generation sensors, resulting in inaccurate estimates of rice plantation areas.

Second, satellite images often suffer from adverse conditions such as cloud shadow or solar radiation [12].

Third, images are often produced by satellites in a polar orbit with a low sampling rate, which prevents 24-hour use.

Last but not least, existing spectral indices for vegetation identification in satellite imagery are empirical in nature and thus require further calibration and validation for different geographic regions and plant species.

### 2.3. Methods of machine learning

Overcoming these problems is possible only by creating a powerful information system for the automated processing of large volumes of data received from the satellite, data that is stored for a long period of time and forms a "history" of observations and expert data.

Solving the problem of data processing using traditional methods of mathematics and statistics is very difficult, therefore, this work uses an approach for evaluating and modeling the development of agricultural crops using streaming data processing of satellite imagery based on machine learning methods, primarily using convolutional neural networks.

Deep learning, in particular deep neural networks [4, 5, 12], has been successfully applied to image data analysis, such as image classification, object detection, and semantic segmentation [13].

Due to these successes, deep learning methods on satellite images for land use classification have recently been applied in various remote mapping studies [13-15].

Therefore, the goal of the work is to increase the efficiency and reliability of the processes of diagnosing the state of agronomic crops using machine learning methods based on streaming data from satellite surveys of large areas.

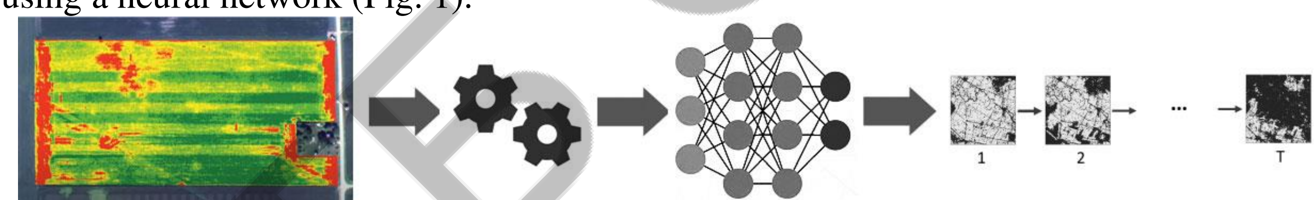
## III. OBJECT, SUBJECT, AND METHODS OF RESEARCH

***The object of research*** is the process of diagnosing the condition of agronomic crops.

***The subject of research*** is the means of the information system for diagnosing the condition of agronomic crops and methods of machine learning based on streaming data from satellite surveys of large areas.

Convolutional neural networks, which implement the popular idea of "deep learning" when building intelligent data processing systems, were chosen as a machine learning method to increase the efficiency and reliability of the processes of diagnosing the state of agronomic crops based on the streaming data processing of satellite surveys of large areas.

The intelligent data processing system itself is a streaming data processing pipeline using a neural network (Fig. 1).




I II III IV

Fig. 1. The structure of the intelligent data processing system in the form of a streaming pipeline of data processing using a neural network.

The multi-level nature of the deep learning neural network architecture will enable the collection of multi-spectral information from satellite imagery in both spatial and temporal dimensions (I). The input data goes to the data processing pipeline (II). The proposed deep neural network (III) is trained on the basis of input data in the form of satellite images in several spectral ranges (Fig. 2). At the output of the system, classification of plant conditions on each fragment of the land plot (IV) is performed according to the existing patterns (training sample).

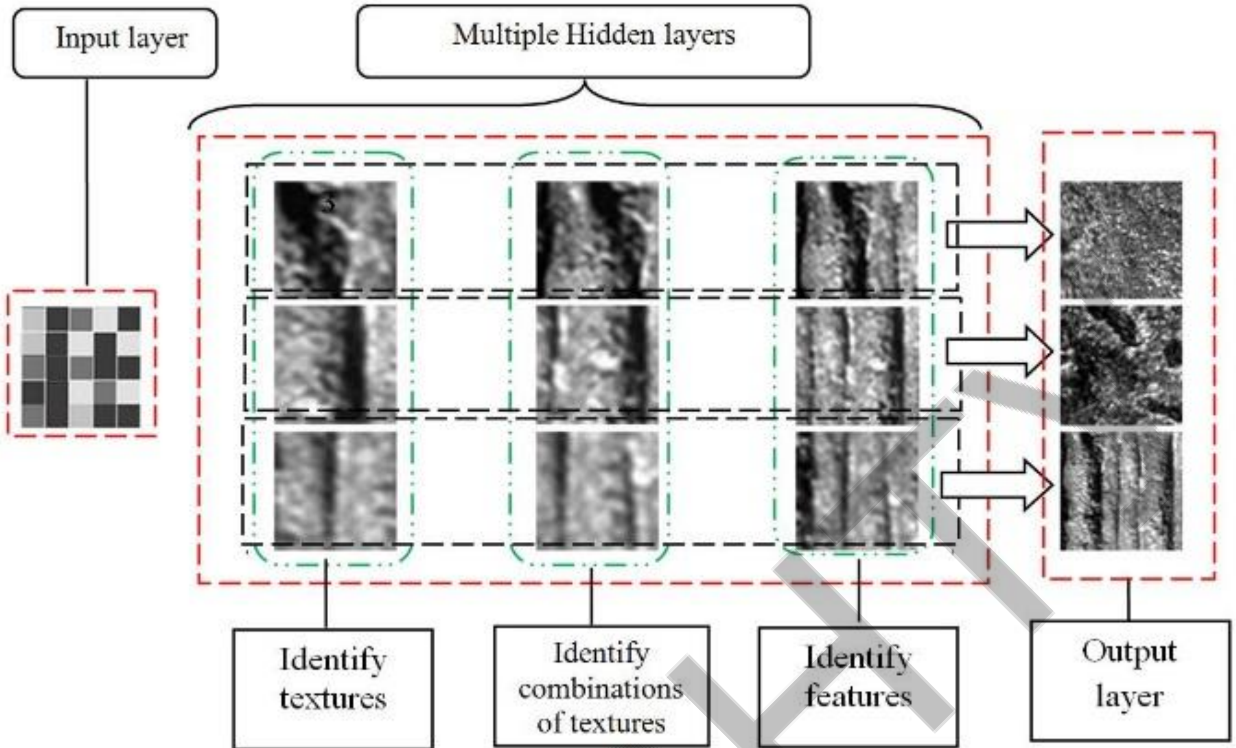



Fig. 2. The structure of a deep neural network that learns on the basis of input data in the form of satellite images in several spectral bands.

The paper proposes the use of a spatio-temporal-spectral deep neural network, which captures time dependencies at several time steps in the past and future directions using hidden layers, captures spatial patterns using convolutional layers, and captures spectral patterns to determine the location of abnormal areas at the pixel level using upsampled layers.

The given structure successfully performs the tasks of collecting data streams from satellite image sources and cleaning from obstacles such as cloud shadows, different angles of the sun's zenith position, and spatial discrepancies and does not require data preprocessing. Such a network is able to generalize spectral, spatial and temporal dependencies without relying on any predefined indices.

To create a powerful classifier with a relatively short training time, the convolutional neural network architecture was chosen and GoogLeNet was adopted. This architecture was proposed by Google specialists in 2014 in a research paper [16, 17].

The GoogLeNet architecture is very different from previous state-of-the-art architectures such as AlexNet and ZF-Net. It uses many different types of techniques, such as 1×1 convolution (to reduce the number of parameters (weights and biases) of the architecture) and global average pooling (reduces the number of parameters to be trained and improves the accuracy of the solution), which allows for a deeper architecture .

The general architecture of GoogLeNet consists of 22 levels (Fig. 3). The architecture is designed with computing efficiency in mind. The idea is that the architecture can be run on individual devices, even with small computing resources.

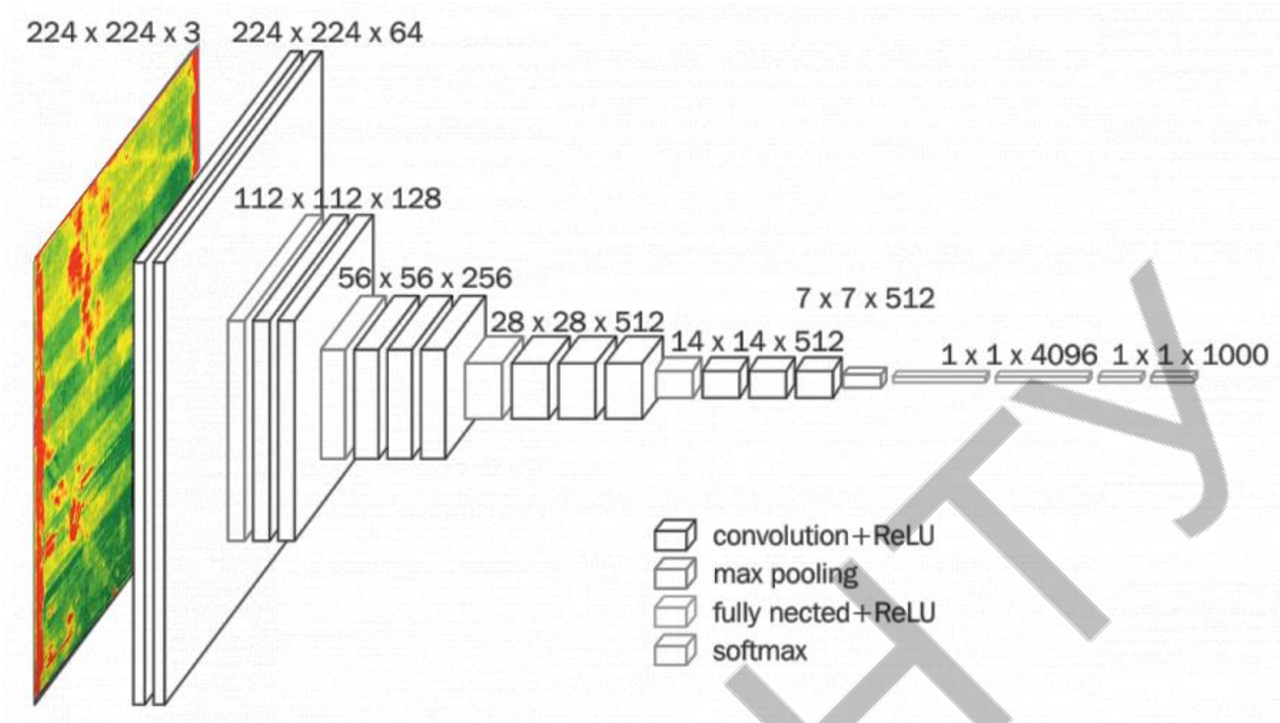



Fig. 3. General Architecture of GoogLeNet Convolutional Neural Network.

The initial architecture can be used in computer vision tasks that involve the use of convolutional filters.

In convolutional neural networks, a large part of the work is choosing the right layer to apply among the most common choices (1x1 filter, 3x3 filter, 5x5 filter, or max pooling). All we need is to find the optimal local design and repeat it spatially.

As these "initial modules" overlap, their output correlation statistics will inevitably change: as features of higher abstraction are captured by higher levels, their spatial concentration is expected to decrease, assuming that 3×3 and 5×5 Convolutions should increase when we pass to higher layers.

In our case, the GoogLeNet architecture accepts a 224 x 224 image with RGB color channels. All convolutions inside this architecture use Rectified Linear Units as their activation functions. So, the network operates with tensors (224, 224, 3) as input data. The model processes the input image and outputs a vector of 1000 values y=(y0, y1, …, y999). The vector y represents the classification probability for the corresponding class.

GoogleNet is trained using distributed machine learning systems with little parallelism of models and data. Asynchronous stochastic gradient descent with a momentum of 0.9 and a fixed learning rate schedule was used during training.

The results of the work of the proposed neural network are given in section IV when solving the problem of classifying the condition of plants according to vegetation indices, which are determined using satellite images in several spectral ranges.

## IV. RESULTS

In this work, the Sentinel platform was chosen as the base due to its high spatial resolution. In addition, its database contains archival data since 1999. Another advantage of this platform is its low time resolution (16 days), which is much smaller compared to reclamation cycles [3].

Access to the data received by the Sentinel satellite platform for research is provided by the Center for Satellite Land Monitoring of the Odesa State Agrarian Academy.

In fig. 4. shows the appearance of satellite images in two spectra: visible and infrared radiation.

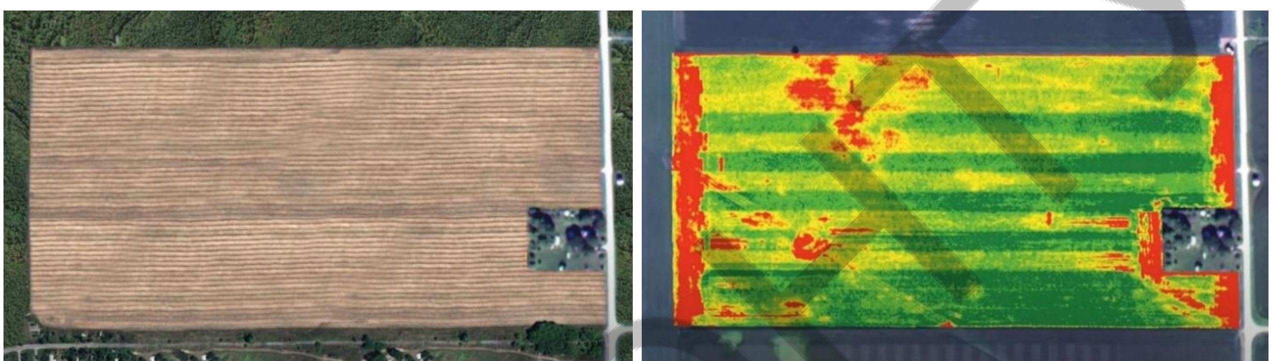


Fig. 4. View of satellite images in two spectra: visible and infrared radiation.

### 4.1. Formation of the educational sample

As vegetation indices (artificially created informative spectral characteristics) are included in the training sample, 5 known values are used in the work [5, 6].

1. Normalized Vegetation Diversity Index (NDVI) (calculated from near-infrared light reflected by vegetation and visible light; intended only to detect living vegetation, as healthier and stronger plants absorb more visible light and reflect more near-infrared light).
2. Land Surface Water Index (LSWI).
3. Soil-adjusted vegetation index (SAVI).
4. Enhanced Vegetation Index (EVI) (uses additional wavelengths of light to reduce NDVI inaccuracies, including solar incidence angle, light distortion and refraction, and noisy ground signals).
5. Plant vegetation index (RGVI).

Another such index is the EVI, which. EVI also allows you to track changes over time.

### 4.2. Study of data distribution

In order to extract informative data from the stream of satellite images, a preliminary study of the spectral characteristics of the data must be conducted. In fig. 4 shows the distribution of indices included in the training sample formed on the basis of satellite images. From these distributions, the most informative indices can be established, on the basis of which a classifier of plant states based on a neural network is built.

It can be seen that each index has its own contribution to the identification of plant states. The most informative for the process of diagnosing plant conditions are the NDVI and RGVI indices. Therefore, based on the spectral characteristics of the indices in question, shown in fig. 5, the indices: NDVI, RGVI were selected for the further formation of the training data sample.

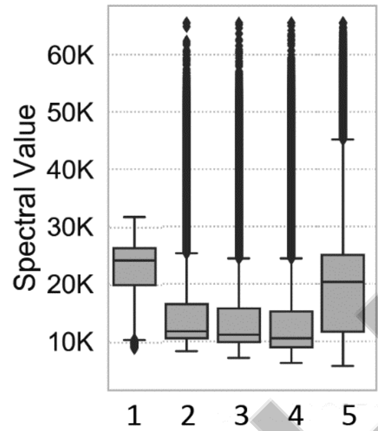



Fig. 5. Value distribution of spectral channels.

**4.3. Study of spectral data correlation**

To understand the relationship between any two spectral bands, as well as their own distribution, the work uses paired graphs. In fig. 6a shows the pairwise correlation of data for all indices considered in the work. In fig. 6b shows a correlation map in the form of a table of correlation coefficients between pairs of indices.

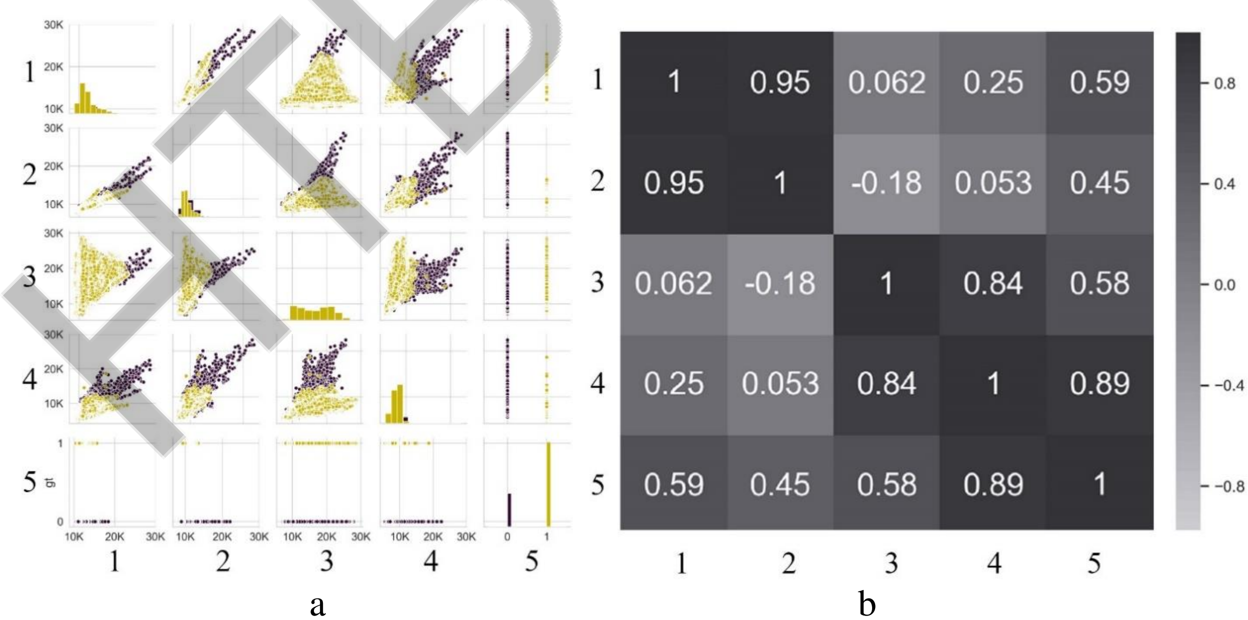



Fig. 6. Correlation of spectral bands: a – in spectral channels (on satisfactory – dark, unsatisfactory vegetation – light), b – spectral Correlation

### 4.2. Construction of a neural network classifier of plant vegetation states

The GoogleNet deep neural network architecture, discussed above, is used to build a plant vegetation state classifier that combines spectral, spatial, and temporal information simultaneously.

The input data of the network consists of mxn pixel matrices for each spectral band (each pixel represents a real geolocation), where m and n are the sizes of the region of interest. In our case, the size mxn is equal to 1x1 meter.

The input consists of lxp pixel matrices for each spectral band, where l and p are the size of the region of interest. In our case, the lxp sizes is 10x10 meters.

The GoogleNet convolutional neural network architecture is implemented on the Keras platform using the TensorFlow library (as a backend for Keras).

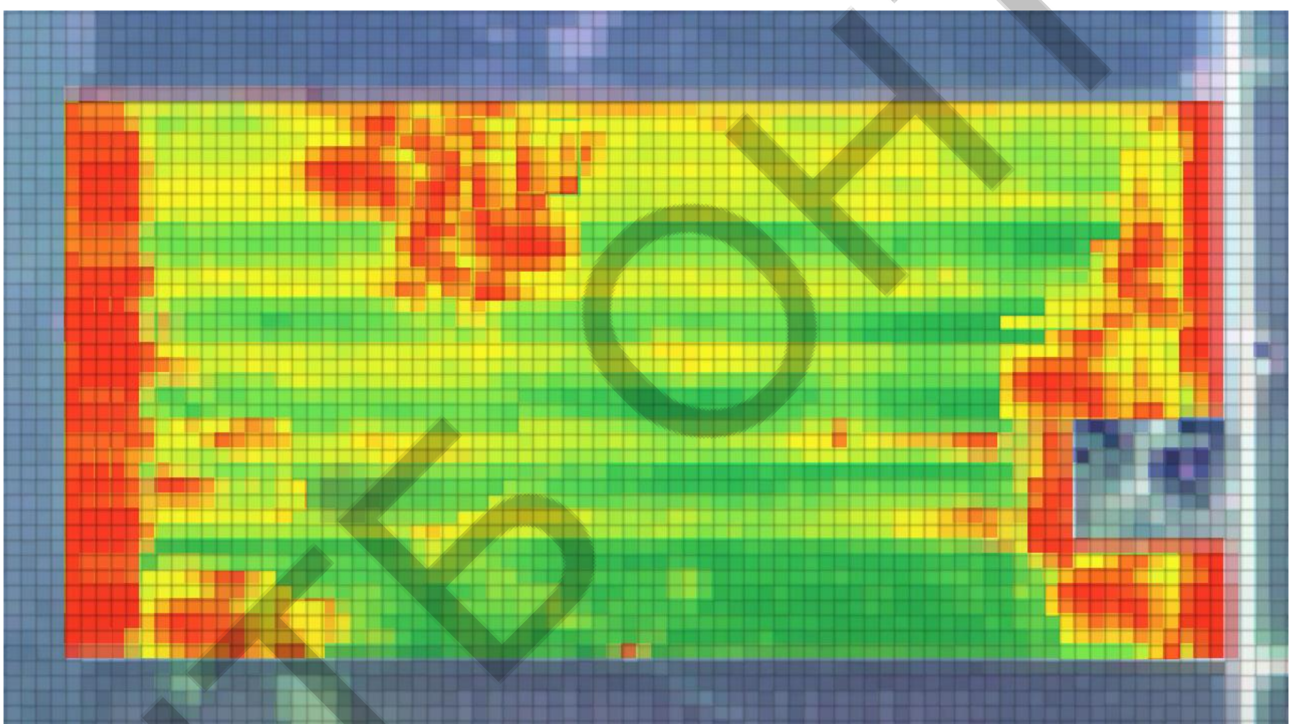


Fig. 7. Classification of vegetation states of plants on the test site (green ¬¬– satisfactory, red – unsatisfactory vegetation)

Graphs of neural network training on the training sample of plant vegetation states on the test site are shown in fig. 8. The classification accuracy was 93.39%.

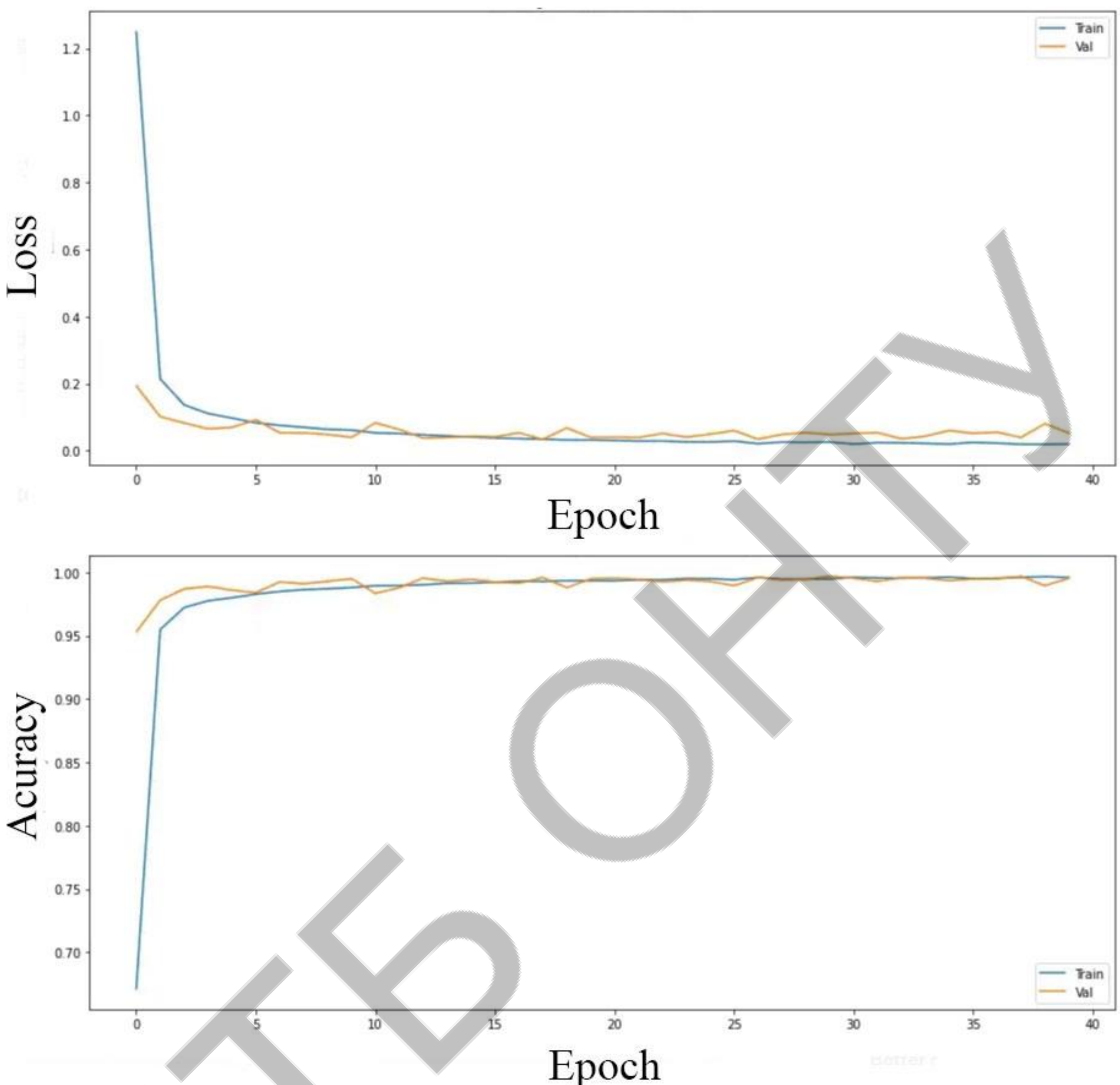



Fig. 8. Training of a neural network on a training sample of vegetation states of plants on the test site: a – training and validation loss; b – training and validation accuracy

Numerical characteristics of the process of building a neural network classifier (training time, testing time, classification accuracy) in comparison with the method of field observations are given in Table 1.

Table 1. The precision, training times, and testing times on the plants' vegetation dataset

| **Methods** | **Training time** | **Testing time** | **Precision** |
|---|---|---|---|
| **Neural network** | 9.8 hrs | 6.1 s | 0.9339 |
| **Field observations** | - | - | 0.7702 |

Table 1 summarizes the reliabilities of diagnosing the conditions of agricultural plants on the test plot, obtained by processing satellite images with a neural network and evaluated using field observations. At the same time, the table does not show the training and testing time of the classifier for the field observation method due to the lack of this data. But these values, according to expert assessments, clearly exceed the values obtained for the neural network up to ten times.

It can be seen from the table that the method of determining the state of plant vegetation using satellite photography has advantages in accuracy by 16% and speed by 5-10 times (according to expert estimates) compared to traditional field observations.

## V. CONCLUSIONS

The work proposes the creation of a software and hardware information system for evaluating and modeling the development of agricultural crops using streaming data from satellite surveys in order to increase the efficiency and reliability of decision-making processes, diagnose the state of agronomic crops in large areas, optimize the management of agricultural resources, while increasing the quality of products , bringing significant economic benefits and saving agricultural resources.

The information system consists of two main components: stream data processing for collecting and cleaning raw images from adverse conditions of satellite images and multi-temporal mapping with high spatial resolution using a classifier based on a deep learning neural network for automatic fixation of features (diagnosis) of vegetation of agricultural plants.

The conducted simulation of the development of agricultural crops using the streaming processing of satellite survey data demonstrates a significant increase in the efficiency and reliability of the process of diagnosing the condition of agronomic crops on the test plot in comparison with the method of field observations. The implemented approach to determining the state of plant vegetation using satellite imagery has the advantages of 16% accuracy and 5-10 times faster (according to expert estimates) compared to traditional field observations.

Although this work is focused on increasing the efficiency and reliability of the processes of diagnosing the state of agronomic crops in Ukraine, the implemented approach is sufficiently general for other regions for full national planning.

The developed system has profound implications for government officials and agricultural managers and other decision-makers who strive for sustainable agricultural production.

The work can be continued in several directions. First, the diagnostic process can be improved by using additional vegetation indices to increase the level of reliability of mapping. Secondly, the proposed assessment of the condition of plants in large areas can be applied to other, non-food crops (for example, forest fire monitoring, water resources dynamics, etc.) for sectoral and national planning.

The author's contribution can be summarized as follows. The author implemented the software part of the information system for evaluating and modeling the development of agricultural crops using streaming data from satellite imagery. As a classifier of vegetation states of agricultural plants, the use of a deep learning neural network is proposed. A training sample of satellite multispectral images was formed for the test agricultural area (Odesa region, Ovidiopol district, geocoordinates of the area: 46.287396, 30.439197) and simulation of the development of agricultural crops in this area was carried out using streaming data processing of satellite images.

## VI. REFERENCES

1. Bastiaanssen, W.G.M., Molden, D.J. and Makin, I.W. (2000) "Remote sensing for irrigated agriculture: Examples from research and possible applications", Agricultural Water Management, Vol. 46, No. 2, pp.137-155. ISSN 0378-3774. DOI 10.1016/S0378-3774(00)00080-9.

2. Lobell, D.B.; Ortiz-Monasterio, J.I.; Asner, G.P.; Naylor, R.L.; Falcon, W.P. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. Agron. J. 2005, 97, 241–249.

3. Tan, G.; Shibasaki, R. Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration. Ecol. Model. 2003, 168, 357–370.

4. Jiménez-Jiménez, S.I.; Marcial-Pablo, M.d.J.; Ojeda-Bustamante, W.; Sifuentes-Ibarra, E.; Inzunza-Ibarra, M.A.; Sánchez-Cohen, I. VICAL: Global Calculator to Estimate Vegetation Indices for Agricultural Areas with Landsat and Sentinel-2 Data. Agronomy 2022, 12, 1518. https://doi.org/10.3390/agronomy12071518

5. Bannari, A., Morin, D., Bonn, F., and Huete, A. (1995) "A review of vegetation indices", Remote Sensing Reviews, Vol. 13, No. 1, pp. 95-120. DOI 10.1080/02757259509532298.

6. Meera Gandhi, G., Parthiban, S., Thummalu, N. and Christy, A. (2015) "NDVI: Vegetation change detection using remote sensing and GIS – A case study of Vellore District", Procedia Computer Science, Vol. 57, pp.1199 – 1210. ISSN 1877-0509. DOI 10.1016/j.procs.2015.07.415.

7. NASA and USGS (2017) "Landsat". [Online]. Available: http://eoedu.belspo.be/en/satellites/landsat.htm

8. Panda, S., Ames, D. and Panigrahi, S. (2010) "Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques", Remote Sensing, Vol. 2, pp. 673-696. ISSN 2072-4292. DOI 10.3390/rs2030673.

9. Potić, I., Bugarski, M. and Matić-Varenica, J. (2017) "Soil Moisture Determination using Remote Sensing data for the property protection and increase of agriculture production", Paper presented at the " Worldbank conference on land and poverty", The World Bank, Washington DC, March 20-24, 2017.

10. USGS (2017) "Landsat and Agriculture - Case Studies on the Uses and Benets of Landsat Imagery in Agricultural Monitoring and Production" [Online]. Available: https://pubs.usgs.gov/of/2017/1034/ofr20171034.pdf

11. Sources must be cited in the standard APA style. Available: https://apastyle.apa.org/style-grammar-guidelines/references/examples

12. C. Szegedy *et al*., "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

13. Szegedy, Christian & Liu, Wei & Jia, Yangqing & Sermanet, Pierre & Reed, Scott & Anguelov, Dragomir & Erhan, Dumitru & Vanhoucke, Vincent & Rabinovich, Andrew. (2014). "Going Deeper with Convolutions".

14. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E.; Hubbard, W., Jackel, L.D. (December 1989). "Backpropagation Applied to Handwritten Zip Code Recognition". Neural Computation. 1(4): 541–551.

15. Szegedy, Christian & Vanhoucke, Vincent & Ioffe, Sergey & Shlens, Jon & Wojna, ZB. (2016). "Rethinking the Inception Architecture for Computer Vision". 10.1109/CVPR.2016.308.

16. Ott, Jordan & Linstead, Erik & LaHaye, Nick & Baldi, Pierre. (2020). Learning in the machine: "To share or not to share?". Neural Networks. 126. 10.1016/j.neunet.2020.03.016.

17. Simonyan, Karen & Zisserman, Andrew. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". arXiv 1409.1556.