

ABSTRACT

of the dissertation by Gailya Faritovna Aubakirova
on the topic «Development of a neural network system for forecasting grain crop
yield in the North Kazakhstan region»,
submitted for the degree of Doctor of Philosophy (PhD)
in the specialty 8D06101 – «Computer Science, computer engineering and control»

Forecasting the yield of grain crops is one of the relevant tasks in the digitalization of the agro-industrial complex, since the accuracy of yield forecasts directly affects production planning, rational use of resources, timely agrotechnological decision-making and the stability of the regional food system. This task is of particular importance for the North Kazakhstan Region, which is one of the main grain-producing territories of the Republic of Kazakhstan. In this region, the yield of spring wheat significantly depends on weather conditions, productive soil moisture reserves, the agrochemical state of the soil and the phytosanitary condition of crops.

The formation of grain crop yield is influenced by a complex of interrelated factors. Air temperature, precipitation, soil moisture availability, nutrient content, the spread of diseases, pests and weeds affect yield not separately, but in interaction with one another. Therefore, traditional statistical methods are not always able to describe complex nonlinear relationships between input parameters and yield with sufficient accuracy. This determines the need to apply machine learning methods, including artificial neural networks, which are capable of identifying hidden patterns in multidimensional data.

The systematization of domestic and international studies shows that the relevance of developing a neural network system for forecasting grain crop yield is determined by the following contradictions:

1. between the high spatial and temporal variability of grain crop yield and the limited capabilities of traditional linear models in describing nonlinear agroclimatic dependencies;
2. between the availability of regional statistical and meteorological data and the insufficient detail of information on the actual condition of a specific crop field;
3. between the development of digital technologies, sensor monitoring and artificial intelligence and the insufficient integration of field data into neural network models for yield forecasting;
4. between the need of agricultural producers for an operational and scientifically grounded yield forecast and the absence of a neural network system adapted to the conditions of the North Kazakhstan Region that combines data collection, processing and intelligent analysis.

Thus, the relevance of the dissertation research is determined by the practical need to improve the accuracy of grain crop yield forecasting and by the scientific task of developing a neural network system that provides comprehensive consideration of agroclimatic, agrochemical, soil, phytosanitary and local field data.

The object of the dissertation research is the process of forecasting grain crop yield in the North Kazakhstan Region.

The subject of the research is methods, algorithms, and software-hardware tools for collecting, processing and neural network analysis of data for forecasting grain crop yield, taking into account agroclimatic, soil-agrochemical, phytosanitary and local field factors.

The purpose of the dissertation research is to develop a neural network system for forecasting grain crop yield in the North Kazakhstan Region, providing comprehensive consideration of yield-forming factors and the possibility of integrating local field data into the forecasting process.

To achieve this purpose, the following tasks were solved in the dissertation:

1. To analyze the theoretical foundations of grain crop yield forecasting and determine the factors that significantly affect the formation of spring wheat yield under the conditions of the North Kazakhstan Region.
2. To form an initial dataset for neural network modeling of wheat yield by districts of the North Kazakhstan Region and to substantiate a system of input features including agrometeorological, soil-agrochemical and phytosanitary parameters.
3. To perform exploratory data analysis, preliminary processing, normalization and preparation of the training sample for constructing a neural network model for yield forecasting.
4. To develop and train a neural network model for forecasting wheat yield, assess its quality and compare it with basic statistical approaches.
5. To develop a device for collecting and transmitting local field data, ensuring the registration of soil and microclimatic parameters on a specific crop field.
6. To develop an algorithm for transforming, structuring and aggregating data obtained from the data collection and transmission device for subsequent use in a neural network model.
7. To substantiate an approach to integrating local field data into a neural network yield forecasting system and to transition from a regional forecast to a farm-level forecast.

The main provisions submitted for defense are:

1. A neural network system for forecasting grain crop yield based on the comprehensive consideration of agrometeorological, soil-agrochemical and phytosanitary factors, adapted to the conditions of the North Kazakhstan Region.
2. An artificial neural network model for forecasting spring wheat yield using a system of input features that characterize temperature regime, precipitation, productive moisture reserves, nitrogen and phosphorus content, and the phytosanitary condition of crops.
3. A device for collecting, processing and transmitting local field data intended for registering soil and microclimatic parameters directly on a crop field.

4. An algorithm for transforming and structuring field data, ensuring the transition from primary sensor signals to aggregated features suitable for use in a neural network yield forecasting model.
5. An approach to integrating local data from a specific farm into a neural network yield forecasting system, allowing the regional forecast to be refined with regard to actual field conditions.

The scientific novelty of the dissertation research consists in the development of a neural network system for forecasting grain crop yield that combines methods of intelligent data analysis, neural network modeling and local field monitoring. A distinctive feature of the proposed approach is the integration of a regional yield forecasting model with data obtained directly from the crop field.

The scientific novelty of the dissertation is represented by the following results:

1. The structure of input features of a neural network model for forecasting wheat yield under the conditions of the North Kazakhstan Region has been substantiated, including agrometeorological, soil-agrochemical and phytosanitary parameters.
2. A neural network model for forecasting spring wheat yield has been developed, making it possible to take into account the nonlinear nature of relationships between yield-forming factors and the target variable.
3. The effectiveness of the neural network approach compared with a linear regression model in forecasting wheat yield based on multidimensional input data has been shown.
4. An autonomous device for collecting and transmitting agrometeorological and soil data from the studied field has been developed for the further formation of a local yield forecasting database.
5. An algorithm for processing local field data has been developed, including the conversion of sensor signals, verification of measurement correctness, assignment of timestamps, structuring of records and aggregation of data into daily, monthly and seasonal features.
6. An approach to integrating local data from a specific farm into a neural network yield forecasting system has been proposed, ensuring the transition from a generalized regional model to a more detailed local forecast.

The theoretical significance of the dissertation research lies in the development of methodological foundations for neural network forecasting of grain crop yield based on the comprehensive consideration of agroclimatic, soil-agrochemical, phytosanitary and local field factors. The work substantiates the need to move from isolated analysis of individual factors to a multidimensional model capable of taking into account nonlinear relationships between environmental parameters and yield. The proposed approach expands the application of artificial neural networks in the tasks of agricultural forecasting and digital farming.

The practical significance of the research lies in the possibility of applying the developed neural network system to support decision-making in agricultural production

in the North Kazakhstan Region. The practical significance of the work is reflected in the following areas:

1. For agricultural producers – the possibility of obtaining a yield forecast taking into account a complex of factors influencing crop formation.
2. For crop monitoring – the use of a data collection and transmission device for recording local soil and microclimate parameters directly in the field.
3. For the digitalization of agriculture – the formation of a unified database combining regional, agrochemical, phytosanitary and local field information.
4. For scientific research – the possibility of further developing neural network yield forecasting models with regard to new data sources and adapting the model to other crops or regions.
5. For the educational process – the use of the developed methods, algorithms and software-hardware solutions in training specialists in the fields of computer science, computer engineering, control and digital agriculture.

Approbation of results. The main results of the dissertation research have been tested through scientific publications, development of a software-hardware solution and implementation of the research results under agricultural production conditions. The developed system for collecting, processing and transmitting agrometeorological data was tested on the crop fields of LLP “Agrofirma Mamlutskaya”. The research results were published in 6 scientific papers, including articles in international peer-reviewed scientific journals and journals recommended by the authorized body in the field of science and higher education. In addition, the results are confirmed by an innovation patent of the Republic of Kazakhstan.

The personal contribution of the author consists in the formulation of the research problem, formation of the initial database, selection and substantiation of the input features of the neural network model, exploratory data analysis, construction and evaluation of the neural network model, development of the structure of the data collection and transmission device, analysis of the results of field testing, as well as generalization and interpretation of the obtained scientific results. The main results of theoretical and experimental studies were obtained by the author independently. In scientific works prepared in co-authorship, the applicant played the leading role in the analysis, processing and generalization of the results.

Structure and volume of the dissertation. The dissertation has a classical structure: introduction, main part consisting of four chapters, conclusion, list of references and appendices. The work includes theoretical analysis of the subject area, construction of a neural network yield forecasting model, development of a field data collection device and integration of local data into the neural network forecasting system. The number of pages, figures, tables and references is to be specified after completion of Chapter 4 and final technical editing of the dissertation.

The introduction substantiates the relevance of the research topic, formulates the purpose and objectives of the work, defines the object and subject of the research, and

presents the scientific novelty, practical significance, provisions submitted for defense, information on approbation of results and the author's personal contribution.

The first chapter considers the theoretical foundations of grain crop yield forecasting. It analyzes agrometeorological factors affecting yield formation under the conditions of the North Kazakhstan Region, traditional statistical forecasting methods, and modern machine learning methods. It is established that wheat yield is formed under the influence of a complex of factors, the relationship between which is complex and nonlinear. The feasibility of using artificial neural networks to solve the problem of yield forecasting is substantiated.

The second chapter forms the initial dataset for constructing a neural network model for forecasting spring wheat yield. Sixteen input features are defined, including agrochemical, meteorological and phytosanitary parameters. Exploratory data analysis is performed, including outlier checking, analysis of descriptive statistics, yield distribution, correlation relationships and relationships between key features and the target variable. Data preprocessing, normalization and preparation of the training sample are carried out. A neural network yield forecasting model is developed and evaluated, demonstrating higher forecast quality compared with multiple linear regression.

The third chapter develops and tests a device for collecting and transmitting local field data for neural network yield forecasting. The structure of the device is presented, including a sensor unit, microcontroller, memory module, real-time clock, GSM/GPRS communication module, autonomous power supply system and stationary data receiving module. The hardware and structural implementation of the device is considered, and field testing is carried out on a crop field. It is shown that the device provides registration of local soil and microclimatic parameters, and that the obtained data can be transformed, verified, aggregated and used to form a local feature database for a neural network model.

The fourth chapter considers the integration of local field data into the neural network yield forecasting system. The transition from regional forecasting to local model adaptation at the level of a specific farm is substantiated. The formation, preliminary processing and exploratory analysis of a local database are described, including data obtained from the data collection device, archival farm data, agrochemical and phytosanitary indicators. The construction of a local neural network model is envisaged, allowing the yield forecast to be refined with regard to actual crop field conditions. The final results of this section will be specified after completion of chapter 4.

The conclusion presents the main results of the dissertation research and formulates conclusions on the development of the neural network system for forecasting grain crop yield, the assessment of model quality, the creation of a field data collection device and the integration of local information into the forecasting system.