ABSTRACT

of the dissertation titled:

« DEVELOPMENT AND RESEARCH OF SPATIAL DATA ANALYSIS METHODS IN TERRITORIAL PLANNING SYSTEMS USING MACHINE LEARNING ALGORITHMS»,

submitted for the degree of Doctor of Philosophy (PhD)

in the educational program 8D06102 – Machine Learning & Data Science

AKYLBEKOV OLZHAS NAURYZBAYEVICH

Relevance of the Study. Contemporary trends in urbanization, demographic expansion, and territorial development exert escalating pressures on territorial planning systems. To plan the growth of cities and regions, you need to look at a wide range of geographic data, including climate, environment, and socio-economic factors. Traditional GIS-based methods frequently don't give very accurate results since they can't find hidden patterns in data with several dimensions. Using machine learning (ML) algorithms opens new ways to understand data, build adaptive models, and make spatial development forecasts more accurate. This is especially useful in complex landscapes and cities that change quickly.

The aim of the dissertation is to investigate and experimentally validate methods for spatial data analysis based on machine learning algorithms for territorial planning tasks, particularly for the classification and segmentation of areas according to their suitability for development.

The core idea of the study is to develop a reproducible computational pipeline that integrates heterogeneous spatial data (elevation, spectral features, climate, infrastructure) into a unified feature set, trains machine learning models, and generates final land suitability maps. The proposed approach is aimed at enhancing the objectivity and automation of urban planning decisions.

The object of the study is the spatial data of the target area (Alatau city, Almaty Region, Kazakhstan), including a digital elevation model, multispectral satellite imagery, climatic indicators, and data on transportation infrastructure.

The subject of the study is the set of machine learning methods and algorithms applied for pixel-based and area-based land classification under conditions of heterogeneous terrain and complex geospatial information.

Research objectives:

- To analyze existing approaches to the application of machine learning methods in spatial (territorial) planning;
- To develop a feature structure from multi-channel spatial data;
- To implement data preprocessing steps: normalization, patching, and augmentation;
- To train and compare various machine learning models — with a particular focus on the CNN–MLP architecture — and perform a comparative evaluation against U-Net, SegFormer, XGBoost, and LightGBM using Accuracy, F1-score, and Cohen’s Kappa metrics;
- To conduct cross-validation and compare model predictions with a reference scheme/master plan;

- To construct a final suitability map and assess its practical relevance for territorial planning tasks.

Research Methods. This study employs modern methods for spatial data analysis, including standardization and aggregation of multi-channel features, training and validation of machine learning models (including neural networks and boosting algorithms), spatial semantic segmentation techniques, suitability map generation, and statistical classification quality metrics such as Accuracy, Precision, Recall, F1-score, and Cohen's Kappa. The experiments were conducted using Python libraries and frameworks (TensorFlow/Keras, Scikit-learn, OpenCV, GDAL) as well as GIS software including QGIS and ArcGIS.

Key Provisions to Be Defended:

- A reproducible computational pipeline for geospatial data analysis is proposed, encompassing multi-channel feature extraction, model training, and suitability map generation;
- The high effectiveness of the CNN–MLP architecture is substantiated, offering a balanced trade-off between accuracy, robustness, and computational efficiency, particularly when working with limited training datasets;
- A comparative evaluation with alternative models (U-Net, SegFormer, XGBoost, LightGBM) highlights the advantages of hybrid approaches in complex geospatial environments;
- A methodology is developed for assessing the consistency of model predictions with reference maps from urban planning documentation;
- The practical applicability of machine learning approaches to the automation of territorial planning is experimentally confirmed, considering climatic, topographic, and infrastructural factors.

Description of the Main Research Results. A software prototype for intelligent spatial data analysis has been developed, implementing a full processing cycle — from feature preparation to the generation of final land suitability maps. Experiments were conducted using data from the Alatau area (Almaty Region), including multispectral imagery, digital elevation models, and infrastructure information.

- the validation results demonstrated high values across key performance metrics:
- classification accuracy reached up to 93%, depending on the model;
- coefficient of determination (R^2) reached up to 0.92 in regression tasks;
- F1-score reached up to 0.91 for the most significant class ("high suitability").

The CNN–MLP architecture delivered the best performance in terms of accuracy and robustness to variations in spatial data. Comparative analysis confirmed that while hybrid and transformer-based models (such as SegFormer and Vision Transformer) demonstrated strong generalization capabilities, CNN–MLP produced more stable and interpretable results, especially under conditions of limited labeled data and high spatial complexity.

Scientific novelty. A comprehensive approach is proposed for multi-criteria land suitability assessment, combining spatial analysis methods, machine learning,

and validation based on urban planning documentation. Model architectures were systematically investigated, along with data structure sensitivity and cross-validation stability. A methodology was developed for generating final suitability maps that take into account spatial patterns, building density, and anthropogenic pressure.

Practical significance. The results can be integrated into territorial planning information systems for preliminary development assessment, zoning, and urban decision support. The proposed models are applicable in the practices of smart urbanism, regional analytics, and environmental monitoring.

Alignment with scientific priorities and national programs. The dissertation results align with the priority areas of scientific and technological development of the Republic of Kazakhstan in the fields of digitalization, sustainable spatial development, and rational use of natural resources. The developed methods may be applied in smart city initiatives, digital cadastre systems, and automated monitoring, as well as in regional and national spatial planning tasks.

Personal contribution of the PhD candidate. The doctoral candidate independently defined the research goals and objectives, developed the conceptual and technical architecture of the solution, implemented the software prototype for spatial data analysis, trained machine learning models, conducted experimental validation, and formulated scientific conclusions. The author also prepared publications and conference abstracts based on the dissertation results.

Publications and dissemination. The main findings and results were presented at international and national scientific conferences, including those indexed in Scopus and Web of Science. The dissertation resulted in:

- 3 articles in journals indexed by Scopus/Web of Science, including Advances in Engineering Software (Q1, Software, 75%), IJACSA (Q3, General CS, 43%), and International Journal of Artificial Intelligence (Q3, Artificial Intelligence, 26%);
- 1 article in the journal «BULLETIN of D. Serikbayev East Kazakhstan Technical University»;

1 article in international conference proceedings.