

АННОТАЦИЯ

диссертационной работы на тему:

«РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДОВ АНАЛИЗА ПРОСТРАНСТВЕННЫХ ДАННЫХ В СИСТЕМАХ ТЕРРИТОРИАЛЬНОГО ПЛАНИРОВАНИЯ С ПРИМЕНЕНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ»,

представленной на соискание степени доктора философии (PhD)
по образовательной программе 8D06102 — «Machine Learning & Data Science»
АКЫЛБЕКОВ ОЛЖАС НАУРЫЗБАЕВИЧ

Актуальность темы. Современные процессы урбанизации, роста населения и освоения новых территорий предъявляют повышенные требования к системам территориального планирования. Управление пространственным развитием городов и регионов требует анализа разнородных геоданных — от климатических и экологических до социально-экономических. Традиционные ГИС-подходы не всегда справляются с задачами высокой точности из-за неспособности выявлять скрытые закономерности в многомерных данных. Применение алгоритмов машинного обучения (ML) открывает возможности интеллектуальной интерпретации данных, построения адаптивных моделей и повышения точности прогнозов пространственного развития, что особенно актуально в условиях сложного ландшафта и быстро меняющейся городской среды.

Целью диссертационной работы является исследование и экспериментальная проверка методов анализа пространственных данных на основе алгоритмов машинного обучения для задач территориального планирования, в частности, классификации и сегментации территорий по уровням пригодности для застройки.

Основная идея работы заключается в создании воспроизводимого вычислительного конвейера, включающего объединение разнородных пространственных данных (рельеф, спектральные признаки, климат, инфраструктура) в единый набор признаков, обучение моделей машинного обучения и формирование итоговых карт пригодности. Разработанный подход ориентирован на повышение объективности и автоматизацию градостроительных решений.

Объектом исследования являются пространственные данные исследуемой территории (территория Алатау, Алматинская область, Республика Казахстан), включая цифровую модель рельефа, мультиспектральные спутниковые снимки, климатические показатели и данные о транспортной инфраструктуре.

Предметом исследования выступают методы и алгоритмы машинного обучения, применяемые для пиксельной и площадной классификации территорий в условиях неоднородного рельефа и комплексной геопропространственной информации.

Задачи исследования:

- Провести анализ существующих подходов к использованию методов машинного обучения в территориальном планировании;
- Разработать структуру признаков из многоканальных пространственных данных;
- Реализовать этапы предобработки данных: нормализацию, разбиение на патчи, аугментацию;
- Обучить и сравнить различные модели машинного обучения — с фокусом на архитектуре CNN–MLP — и провести сравнительный анализ с моделями U-Net, SegFormer, XGBoost, LightGBM по метрикам Accuracy, F1-score, Cohen's Kappa;
- Провести кросс-валидацию и сопоставление результатов классификации с эталонной схемой/генеральным планом;
- Построить итоговую карту пригодности и оценить её практическую применимость для задач территориального планирования.

Методы исследования. В работе использованы современные методы анализа пространственных данных, включая стандартизацию и агрегацию многоканальных признаков, обучение и валидацию моделей машинного обучения (в том числе нейросетевых и бустинговых), методы пространственной семантической сегментации, построение карт пригодности, а также статистические метрики оценки качества классификации (Accuracy, Precision, Recall, F1-score, Cohen's Kappa). Эксперименты проведены с использованием библиотек и фреймворков Python (TensorFlow/Keras, Scikit-learn, OpenCV, GDAL) и геоинформационного ПО (QGIS, ArcGIS).

Основные положения, выносимые на защиту:

- Предложен воспроизводимый вычислительный конвейер анализа пространственных данных, включающий формирование многоканальных признаков, обучение моделей и генерацию карт пригодности;
- Обоснована высокая эффективность архитектуры CNN–MLP, обеспечивающей баланс между точностью, устойчивостью и вычислительной экономичностью при работе с ограниченными обучающими выборками;
- Выполнено сравнение с альтернативными моделями (U-Net, SegFormer, XGBoost, LightGBM), что позволило выявить преимущества гибридных подходов в условиях сложной геопространственной структуры;
- Разработана методика оценки согласованности прогнозов с эталонными схемами градостроительной документации;
- Экспериментально подтверждена применимость подхода к автоматизации территориального планирования с учетом климатических, рельефных и инфраструктурных факторов.

Описание основных результатов исследования. Разработан программный прототип интеллектуального анализа пространственных данных,

реализующий полный цикл: от подготовки признаков до генерации итоговых карт пригодности. Эксперименты проведены на данных города Алатау (Алматинская область) с использованием мультиспектральных изображений, цифровых моделей рельефа и инфраструктурной информации.

- Точность классификации (Accuracy): до 93 % в зависимости от модели и структуры входных данных;
- Коэффициент детерминации (R^2): до 0.92 в задачах регрессии;
- F1-мера: до 0.91 по ключевому классу "высокая пригодность".

Модели на основе архитектуры CNN–MLP продемонстрировали наилучшие показатели по точности и устойчивости к изменению входных данных. Сравнительный анализ подтвердил, что гибридные и трансформерные решения (в частности, SegFormer и Vision Transformer) обладают потенциалом обобщения, однако CNN–MLP показала более стабильные и интерпретируемые результаты, особенно при ограниченных обучающих данных и высокой пространственной сложности.

Научная новизна. Впервые предложен комплексный подход к многокритериальной оценке территориальной пригодности, сочетающий методы пространственного анализа, машинного обучения и валидации на основе градостроительной документации. Систематически исследованы архитектуры моделей, чувствительность к структуре данных и стабильность при перекрестной проверке. Разработана методика генерации итоговых карт пригодности, учитывающая пространственную структуру, плотность застройки и антропогенную нагрузку.

Практическая значимость. Результаты могут быть интегрированы в состав информационных систем территориального планирования для предварительной оценки застройки, зонирования территории и поддержки принятия градостроительных решений. Полученные модели применимы в практике "умного" градостроительства, регионального анализа и экологического мониторинга.

Соответствие направлениям развития науки и государственных программ. Результаты диссертации соответствуют приоритетным направлениям научно-технического развития Республики Казахстан в области цифровизации, устойчивого территориального развития и рационального использования природных ресурсов. Полученные методы могут быть применены в проектах «умных городов», цифрового кадастра и автоматизированного мониторинга, а также в задачах пространственного планирования на региональном и национальном уровнях.

Личный вклад докторанта. Докторантом лично выполнены: постановка цели и задач исследования, разработка концептуальной и технической архитектуры решения, реализация программного прототипа анализа пространственных данных, обучение моделей машинного обучения, проведение

экспериментальной валидации и формулирование научных выводов. Также автором подготовлены публикации и тезисы по результатам диссертации.

Апробация и публикации. Основные положения и результаты исследования докладывались на международных и республиканских научных конференциях, в том числе на мероприятиях, индексируемых в базах Scopus и Web of Science. По теме диссертации опубликованы:

- 3 статьи в изданиях, индексируемых Scopus/Web of Science, включая Advances in Engineering Software (Q1, Software, 75%), IJACSA (Q3, General CS, 43%), International Journal of Artificial Intelligence (Q3, Artificial Intelligence, 26%);
- 1 статья в журнале «ШҚТУ хабаршысы»;
- 1 статья в сборниках международных конференций.

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
AKYLBKOV 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.

АНДАТПА

диссертациялық жұмысына тақырыбы:

**«МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІН ҚОЛДАНА
ОТЫРЫП, АУМАҚТЫҚ ЖОСПАРЛАУ ЖҮЙЕЛЕРІНДЕ
КЕҢІСТІКТІК ДЕРЕКТЕРДІ ТАЛДАУ ӘДІСТЕРІН ӘЗІРЛЕУ ЖӘНЕ
ЗЕРТТЕУ», 8D06102 — «Machine Learning & Data Science» білім беру
бағдарламасы бойынша философия докторы (PhD) ғылыми дәрежесін алу
үшін ұсынылған
АКЫЛБЕКОВ ОЛЖАС НАУРЫЗБАЕВИЧ**

Зерттеу жұмысының өзектілігі. Қазіргі урбанизация, халық санының өсуі және жаңа аумақтарды игеру үдерістері аумақтық жоспарлау жүйелерінің тиімділігіне жоғары талаптар қояды. Қалалар мен аймақтардың кеңістіктік дамуын басқару табиғи-климаттық, инфрақұрылымдық, әлеуметтік-экономикалық және экологиялық деректердің ірі көлемін талдауды қажет етеді. Дәстүрлі геоақпараттық әдістер көп өлшемді деректердегі жасырын заңдылықтарды ескере алмауына байланысты болжамның дәлдігін әрдайым қамтамасыз ете алмайды. Ал машиналық оқыту алгоритмдерін қолдану кеңістіктік модельдеудің, жерді қашықтықтан зондтау (ДЗЗ) деректерін интеллектуалды интерпретациялаудың және талдау үдерістерін автоматтандырудың жаңа мүмкіндіктерін ашады.

Зерттеу жұмысының мақсаты — құрылысқа жарамдылық деңгейлері бойынша аумақтарды жіктеу және сегменттеу сияқты аумақтық жоспарлау міндеттері үшін машиналық оқыту алгоритмдері негізінде кеңістіктік деректерді талдау әдістерін әзірлеу және олардың тәжірибелік тиімділігін тексеру.

Зерттеу нысаны — Алатау қаласы (Алматы облысы) бойынша цифрлық бедер үлгілері, мультиспектралды спутниктік суреттер, климаттық көрсеткіштер және көлік инфрақұрылымы жөніндегі кеңістіктік деректер.

Зерттеу пәні — әртекті рельеф пен күрделі геоақпараттық жағдайларда аумақтарды пиксельдік және аумақтық жіктеу үшін қолданылатын машиналық оқыту әдістері мен алгоритмдері.

Зерттеу міндеттері:

- Машиналық оқыту әдістерін аумақтық жоспарлауда қолданудың қолданыстағы тәсілдерін талдау;
- Көпарналы кеңістіктік деректер негізінде белгілік құрылымды әзірлеу;
- Деректерді алдын ала өңдеу: нормализация, патчтарға бөлу және аугментация жүргізу;
- CNN–MLP архитектурасына баса назар аудара отырып, машиналық оқыту модельдерін (U-Net, SegFormer, XGBoost, LightGBM) оқыту және оларды негізгі метрикалар (Accuracy, F1-score, Cohen's Kappa) бойынша салыстыру;

- Модельдердің болжамдарын эталонды схема/бас жоспармен салыстыра отырып, кросс-валидация өткізу;
- Қорытынды жарамдылық картасын құру және оның практикалық маңыздылығын бағалау.

Зерттеу әдістері. Кеңістіктік деректерді стандарттау, көпарналы белгілерді біріктіру, нейрондық және бустингтік модельдерді оқыту, кеңістіктік семантикалық сегментация, жарамдылық карталарын құру, сондай-ақ классификация сапасын бағалауға арналған статистикалық метрикалар (Accuracy, Precision, Recall, F1-score, Cohen's Kappa) қолданылды. Эксперименттер Python кітапханалары мен фреймворктары (TensorFlow/Keras, Scikit-learn, OpenCV, GDAL) және QGIS, ArcGIS ГАЖ құралдары арқылы жүзеге асырылды.

Қорғалуға ұсынылатын негізгі қағидалар:

- Кеңістіктік деректерді талдауға арналған қайталанатын есептеуіш конвейер ұсынылды: көпарналы белгілерді дайындаудан бастап жарамдылық карталарын құруға дейін;
- CNN–MLP архитектурасының тиімділігі негізделді — бұл модель шектеулі оқыту үлгілерінде жоғары дәлдік, тұрақтылық және түсініктілік көрсетті;
- Альтернативті модельдермен (U-Net, SegFormer, XGBoost, LightGBM) салыстырмалы бағалау жүргізіліп, гибридті тәсілдердің күрделі геокеңістіктік жағдайларда басым екендігі дәлелденді;
- Модель болжамдарының бас жоспардың эталондық карталарымен сәйкестігін бағалау әдістемесі әзірленді;
- Климаттық, рельефтік және инфрақұрылымдық факторларды ескерумен аумақтық жоспарлау процестерін автоматтандыру үшін машиналық оқытуды қолданудың практикалық мүмкіндігі тәжірибелік түрде расталды.

Негізгі нәтижелер. Кеңістіктік деректерді интеллектуалды талдау бойынша бағдарламалық прототип әзірленіп, жарамдылық карталарын генерациялауды толық қамтитын есептік конвейер жүзеге асырылды. Алатау ауданы деректері негізінде мультиспектралды суреттер, цифрлық бедер үлгілері және инфрақұрылымдық ақпараттар қолданылып, жүргізілген валидация нәтижелері жоғары көрсеткіштерді көрсетті:

- дәлдік (Accuracy) – 93% дейін;
- детерминация коэффициенті (R^2) – 0.92 дейін;
- F1-мөлшері – 0.91 дейін («жоғары жарамдылық» класы).

CNN–MLP архитектурасы кеңістіктік күрделілігі жоғары деректермен жұмыс істеу барысында ең жақсы және тұрақты нәтижелер көрсетті. Салыстырмалы талдау SegFormer және Vision Transformer секілді трансформерлік модельдер жақсы жалпылау қабілетіне ие екенін көрсетті, алайда CNN–MLP практикалық тұрақтылық және интерпретация жағынан басым болды, әсіресе белгіленген деректер көлемі шектеулі жағдайда.

Ғылыми жаңалығы. Кеңістіктік жарамдылықты бағалау бойынша көпкритерийлі тәсіл алғаш рет ұсынылып, кеңістіктік талдау, машиналық оқыту және ресми жоспарлау құжаттарымен салыстыру әдістері біріктірілді. Модельдер архитектурасының әсері, деректер құрылымына сезімталдығы мен тұрақтылығы жүйелі түрде зерттелді.

Практикалық маңызы. Зерттеу нәтижелері аумақтық жоспарлау жүйелерінде, жерді зонирлеу мен құрылысқа алдын ала баға беру, экологиялық бақылау және «ақылды қала» технологияларында қолдануға жарамды.

Мемлекеттік және ғылыми даму басымдықтарына сәйкестігі. Жұмыс нәтижелері Қазақстан Республикасының цифрландыру, орнықты аумақтық даму және табиғи ресурстарды тиімді пайдалану бағыттары бойынша мемлекеттік стратегияларға толық сәйкес келеді. Әдістерді ұлттық және өңірлік деңгейдегі «ақылды қала», кадастр және мониторинг жобаларында пайдалануға болады.

Автордың жеке үлесі. Докторант зерттеу мақсаттары мен міндеттерін дербес айқындап, тұжырымдамалық және техникалық шешімдерді әзірледі, бағдарламалық прототипті жүзеге асырды, модельдерді оқытып, эксперименттік валидация жүргізді және ғылыми нәтижелерді тұжырымдады. Нәтижелер негізінде ғылыми мақалалар мен тезистер дайындалды.

Апробация және жарияланымдар.

- Scopus/Web of Science базаларында индекстелетін 3 мақала (Advances in Engineering Software – Q1, IJACSA – Q3, International Journal of Artificial Intelligence – Q3);
- «ШҚТУ хабаршысы» журналында 1 мақала;
- Халықаралық конференциялар жинақтарында 1 мақала жарияланды.