






[https://doi.org/10.51885/3134-8025\\_IICS\\_2026\\_1\\_1](https://doi.org/10.51885/3134-8025_IICS_2026_1_1)  
MPNТИ 28.23.15

## МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В АНАЛИЗЕ ВИЗУАЛЬНЫХ ДАННЫХ ДЛЯ 3D-МОДЕЛИРОВАНИЯ ГОРОДСКОЙ ИНФРАСТРУКТУРЫ

## МАШИНАЛЫҚ ОҚЫТУ АРҚЫЛЫ ВИЗУАЛДЫ МӘЛІМЕТТЕРДІ ӨҢДЕУ ЖӘНЕ 3D-ҚАЛА ҮЛГІЛЕРІН ЖАСАУ

## MACHINE LEARNING METHODS FOR 3D RECONSTRUCTION OF URBAN ENVIRONMENTS

О.Н. Акылбеков <sup>1\*</sup>, А.Н. Молдагулова <sup>1</sup> Г.С. Закария <sup>1</sup>  
С.Т. Байдильдинова <sup>2</sup>, А. Бекарыстанқызы <sup>3</sup>

<sup>1</sup>Казахский национальный исследовательский технический университет имени К.И. Сатпаева,  
г. Алматы, Казахстан

<sup>2</sup>Университет международного бизнеса имени К. Сагадиева, г. Алматы, Казахстан

<sup>3</sup>Университет Нархоз, г. Алматы, Казахстан

\*Автор-корреспондент: Акылбеков Олжас Наурызбаевич, e-mail: o.akylbekov@satbayev.university

### Ключевые слова:

машинное обучение,  
фото- и видеоданные,  
камера наблюдения,  
трёхмерное  
моделирование,  
городская среда,  
территориальное  
планирование,  
интеллектуальные  
системы.

### АННОТАЦИЯ

В статье рассматриваются методы машинного обучения, применяемые для анализа визуальных данных (фото- и видеоматериалов), полученных с камер наблюдения, с целью построения трёхмерных моделей городской инфраструктуры. Обоснована актуальность автоматизации пространственного анализа в условиях роста объёмов визуальной информации. В исследовании применены современные алгоритмы машинного обучения, включая логистическую регрессию, деревья решений, случайный лес, методы кластеризации и глубокие нейронные сети (U-Net, ResNet, ViT). Проведено количественное сравнение всех моделей по метрикам MAE, IoU и F1-score, что позволило объективно оценить их применимость для задач 3D-моделирования городской инфраструктуры. Научная новизна заключается в комплексной интеграции методов машинного обучения и компьютерного зрения для создания цифровых двойников городской среды. Практическая значимость подтверждается возможностью использования разработанных решений в системах территориального планирования и мониторинга городской инфраструктуры.

### Түйінді сөздер:

машиналық оқыту, фото  
және бейне деректер,  
бақылау камерасы,  
үшөлшемді модельдеу,  
қалалық орта, аумақтық

### ТҮЙІНДЕМЕ

Бұл мақалада қалалық инфрақұрылымның үшөлшемді үлгілерін құру үшін бақылау камераларынан алынған визуалды деректерді (фото және бейне) талдау барысында машиналық оқыту әдістерін қолдану қарастырылады. Визуалды ақпарат көлемінің үнемі артуы кеңістіктік талдауды автоматтандыру қажеттілігін туындатады.



жоспарлау,  
интеллектуалды  
жүйелер.

Зерттеуде логистикалық регрессия, шешім ағаштары, кездейсоқ орман (random forest), кластерлеу әдістері және терең нейрондық желілер (U-Net, ResNet, ViT) сияқты заманауи машиналық оқыту алгоритмдері қолданылды. Барлық модельдердің тиімділігі MAE, IoU және F1-score метрикалары арқылы салыстырмалы түрде бағаланды. Ғылыми жаңалық — машиналық оқыту мен компьютерлік көру әдістерін біріктіре отырып, қалалық ортаның цифрлық үлгілерін жасау. Жұмыстың практикалық маңыздылығы – алынған шешімдердің аумақтық жоспарлау мен қалалық инфрақұрылымды бақылау жүйелерінде қолдану мүмкіндігі.

---

**Keywords:**

machine learning, photo and video data, surveillance camera, three-dimensional modeling, urban environment, territorial planning, intelligent systems.

---

**ABSTRACT**

This article examines machine learning methods used to analyze visual data (photos and videos) from surveillance cameras for the purpose of constructing 3D models of urban infrastructure. The study highlights the relevance of automating spatial analysis in response to the growing volume of visual information. The research applied modern machine learning algorithms, including logistic regression, decision trees, random forest, clustering methods, and deep neural networks (U-Net, ResNet, ViT). A quantitative comparison of all models was performed using MAE, IoU, and F1-score metrics, allowing for an objective evaluation of their applicability in 3D urban modeling. The scientific novelty lies in the integrated application of machine learning and computer vision techniques to create digital twins of urban environments. The practical significance is demonstrated by the potential use of the developed solutions in territorial planning and urban infrastructure monitoring systems.

---

**ВВЕДЕНИЕ**

Современное развитие цифровых технологий приводит к резкому увеличению объёмов неструктурированных данных, поступающих из надёжных источников – фото- и видеосъёмки, аэрофотосъёмки, спутниковых систем, сенсоров и датчиков. Эти потоки визуальной информации позволяют фиксировать изменения городской среды с высокой регулярностью и пространственной детализацией. Они отражают не только текущее состояние городской инфраструктуры, но и процессы её трансформации: уплотнение застройки, изменение транспортных потоков, использование общественных пространств и т.д. Однако из-за больших объёмов данных их ручная интерпретация становится невозможной, что требует применения современных аналитических методов, включая машинное обучение.

Методы машинного обучения позволяют автоматически выявлять закономерности в больших массивах визуальных данных и использовать их для пространственного анализа и прогнозирования. При этом возможен переход от низкоуровневого описания изображений к определению устойчивых признаков городской морфологии. Особенно актуально это в условиях урбанизированной среды, насыщенной сенсорами и камерами видеонаблюдения, формирующими непрерывный поток информации. Благодаря достижениям в области компьютерного зрения и глубокого обучения стало возможным извлечение пространственных признаков из изображений и реконструкция трёхмерных моделей городской среды (Hossain et al., 2020; Potok et al., 2018).

Несмотря на активное развитие методов анализа изображений, их прикладное применение в задачах пространственного планирования пока ограничено. Разработка устойчивых решений требует учёта вариативности данных, калибровки камер, различий

в городской застройке. Алгоритмы глубокого обучения демонстрируют высокую эффективность в обработке аннотированных фото- и видеоданных (Baran, 2020; Dietrich et al., 2020; Alzubaidi et al., 2021), позволяя проводить автоматическую сегментацию, классификацию объектов и генерацию трёхмерных представлений городской инфраструктуры.

Цель данного исследования — разработка и апробация подхода к построению 3D-моделей городской среды на основе визуальных данных с камер наблюдения с использованием и сравнением различных алгоритмов машинного обучения. Задачи включают: выбор и реализацию алгоритмов обработки визуальной информации; выделение пространственных признаков; анализ точности реконструкции; оценку применимости моделей для территориального планирования.

Методологическая основа работы сочетает технические и сравнительные подходы. Исходные данные представляют собой фото- и видеоматериалы городской среды, собранные в г. Алматы. Для их анализа применялись как классические алгоритмы (логистическая регрессия, деревья решений, случайный лес), так и нейросетевые архитектуры (U-Net, ResNet-50, ViT). В качестве метрик использовались IoU, F1-score, MAE и время обработки.

### **МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ**

Методологическая основа настоящего исследования опирается на технические, индуктивные и сравнительные подходы, адаптированные для пространственного анализа городской среды с применением современных методов машинного обучения.

Объект исследования включает визуальные данные (фото и видео), полученные с уличных камер наблюдения, размещённых в городской инфраструктуре. В качестве экспериментального материала использовались видеопотоки в форматах H.264 и MJPEG, а также фотоснимки в формате JPEG, полученные с устройств с разрешением не менее 1920×1080 пикселей.

Обработка данных осуществлялась на вычислительном комплексе с графическим процессором NVIDIA RTX 3080, с использованием Python 3.10, платформ TensorFlow 2.13 и PyTorch 2.0, а также библиотек OpenCV, NumPy и scikit-learn (Weng et al., 2024).

Предобработка изображений включала нормализацию, коррекцию перспективы и стабилизацию видео. Для извлечения ключевых признаков применялись алгоритмы SIFT и ORB, после чего осуществлялась корреляция между последовательными кадрами для восстановления информации о сцене. Полученные данные использовались для построения плотных карт и последующего формирования трёхмерных моделей методом Structure from Motion (Rodríguez-Lira et al., 2025; Moradi, 2025; Akyzbekov et al., 2025).

Технический подход предусматривал количественное описание параметров визуальных данных, включая пространственное распределение объектов, изменения освещённости, геометрию сцены и плотность застройки. Для анализа использовались следующие методы:

- гистограммы частот – для оценки распределения интенсивностей пикселей на изображениях;
- стандартное отклонение – как мера разброса значений яркости;
- коэффициент ранговой корреляции Спирмена – для выявления зависимостей между параметрами изображений и метками данных;
- PCA (Principal Component Analysis) – применялся как метод линейного снижения размерности, позволяющий выделить наиболее значимые компоненты признакового пространства без потери ключевой информации.

Использование PCA обеспечивало уменьшение избыточности в данных и ускорение обучения моделей за счёт исключения неинформативных признаков. Этот подход

особенно эффективен при обработке изображений с высокой вариативностью по освещению и ракурсам.

Индуктивный подход был реализован через аннотирование и кластеризацию видеоданных с применением алгоритмов K-средних и DBSCAN. Модели машинного обучения, обученные на датасетах Cityscapes и KITTI, использовались для выявления пространственных закономерностей городской застройки. Адаптированные архитектуры ResNet и U-Net применялись в рамках контролируемого и трансферного обучения.

Сравнительный анализ охватывал алгоритмы CNN, деревья решений, случайный лес и градиентный бустинг. Оценка производилась по метрикам точности классификации, времени обработки, устойчивости к шуму и способности к обобщению. Верификация осуществлялась с использованием 5-кратной кросс-валидации и метрик MAE, IoU и F1-score.

Фотограмметрия и методы компьютерного зрения использовались для построения 3D-моделей на основе изображений, аэрофотоснимков и спутниковых данных, что обеспечило многомасштабное покрытие территории. Применялись нейросетевые алгоритмы глубокой свёртки и методы многоракурсной стереорекострукции. Выделенные опорные точки кластеризовались и интегрировались в единую цифровую сцену, обеспечивая высокую точность геопривязки.

Программные эксперименты проводились на урбанистических ландшафтах восточной части города Алматы с последующим сопоставлением 3D-моделей с геоинформационными данными (форматы GeoTIFF и Shapefile), что обеспечило проверку достоверности пространственной реконструкции.

Для оценки качества применённых архитектур были выбраны модели U-Net, ResNet и Vision Transformer (ViT), обладающие высокой точностью сегментации и извлечения признаков. Валидация моделей осуществлялась с разбиением выборки в пропорции 70/30 и использованием 5-fold кросс-валидации.

Критерии оценки включали:

- IoU (Intersection over Union) — для анализа перекрытия сегментированных объектов;
- F1-score – для оценки классификационной точности;
- MAE (Mean Absolute Error) — при реконструкции геометрии объектов;
- время обработки — как критерий применимости в реальном времени.

Аннотирование данных проводилось вручную и проверялось экспертами. Эксперименты охватывали изображения и видео, полученные в различных условиях освещённости и плотности застройки, с использованием данных из города Алматы, Казахстан.

В дополнение к архитектурному выбору моделей были заданы конкретные параметры обучения. Модель U-Net использовалась с глубиной в пять уровней и симметричной деконволюционной структурой, а входные изображения нормализовались до размера 256×256 пикселей. Обучение проводилось в течение 50 эпох с применением оптимизатора Adam и функции потерь binary cross-entropy, размер батча составлял 16. Архитектура ResNet-50 использовала предварительно обученные веса ImageNet, с дообучением последних слоёв и входным размером изображений 224×224 пикселя; обучение осуществлялось на протяжении 30 эпох с использованием стохастического градиентного спуска и коэффициентом обучения 0.01. Для Vision Transformer была реализована конфигурация ViT-Base с размером патча 16×16, глубиной сети 12 слоёв и размерностью скрытого пространства 768. Входные изображения приводились к размеру 224×224 пикселя, обучение длилось 30 эпох с использованием оптимизатора AdamW и шагом обучения 0.0001. Классические алгоритмы (логистическая регрессия, дерево решений, случайный лес) использовались в качестве базовой линии. Глубина дерева

ограничивалась 10 уровнями, в случайном лесе использовалось 100 деревьев. Перед обучением данные проходили нормализацию с применением метода MinMax. Все вычисления проводились на аппаратной платформе с графическим ускорителем NVIDIA RTX 3080, в среде Python 3.10 с использованием библиотек PyTorch 2.0, scikit-learn и OpenCV.

## РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Машинное обучение — это технология искусственного интеллекта, отличающаяся от традиционных программных решений тем, что она опирается на индуктивные рассуждения и адаптируется к поступающим данным (Pan et al. 2025). По мере увеличения объёма и качества обучающей выборки модели машинного обучения демонстрируют более высокую точность анализа. В рамках настоящего исследования, посвящённого методам машинного обучения для анализа видео- и фотоинформации с целью построения трёхмерных пространственных моделей городской среды в системах территориального планирования, были рассмотрены ключевые подходы, такие как классификация, регрессия, кластеризация, анализ текста, проверка аномалий, а также последовательные шаблоны. Особое внимание уделено контролируемому и неконтролируемому обучению, применяемому для выявления структурных особенностей городской инфраструктуры (Akhavi Zadegan et al. 2025; Yu Shang и et. al, 2025; Rodríguez-Lira et al. 2025).

Практически значимым стало разграничение между поверхностным и глубоким обучением, особенно в задачах извлечения признаков городской сцены из визуальных данных. Глубокие нейронные сети, в отличие от поверхностных алгоритмов, способны формировать сложные иерархические представления, что критически важно при анализе видео- и фотоданных для 3D-моделирования городского пространства (Li et al., 2024; Liu et al., 2023). На этапе обучения изображениям присваиваются категории, содержащие как детерминированную, так и вероятностную информацию, что способствует качественной реконструкции городской среды [8–9].

Для задач бинарной классификации визуальных объектов городской среды, таких как определение принадлежности пикселей или сегментов изображения к категории «здание»/«не здание», была применена модель логистической регрессии. Данный алгоритм позволяет оценить вероятность отнесения объекта к заданному классу на основе набора визуальных признаков.

Математически логистическая регрессия описывается следующим выражением:

$$P(Y = 1 | X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)'}}$$

где:  $P(Y = 1 | X)$  — вероятность принадлежности объекта к классу 1 (например, элемент городской застройки);  $x_1, x_2, \dots, x_n$  — признаки, извлечённые из изображения (цвет, градиент, текстурные характеристики, плотность пикселей и др.);  $\beta_0$  — свободный коэффициент (смещение);  $\beta_1, \beta_2, \dots, \beta_n$  — веса признаков, полученные в процессе обучения модели;  $e$  — экспоненциальная функция, применяемая в логистическом преобразовании.

Вероятность  $P(Y = 1 | X)$  моделируется с помощью сигмоидной функции активации, которая ограничивает значение на выходе в диапазоне от 0 до 1. Оптимизация параметров модели проводилась методом градиентного спуска с минимизацией функции потерь binary cross-entropy, что обеспечивало устойчивую сходимость даже на небольших подмножествах обучающей выборки.

В контексте данного исследования логистическая регрессия использовалась как базовая модель для классификации пикселей на семантические категории городской сцены. Несмотря на простоту, алгоритм продемонстрировал приемлемую точность на ограниченных подмножествах данных и служил сравнительной базой для оценки производительности более сложных архитектур глубокого обучения.

Алгоритм AdaBoost, построенный на механизме ансамблирования слабых классификаторов (обычно деревьев решений глубины 1–3), продемонстрировал высокую эффективность при анализе видеоинформации, направленной на построение трёхмерных моделей городской среды. Принцип работы алгоритма заключается в последовательном обучении слабых моделей с акцентом на примеры, которые ранее были классифицированы с ошибкой, что позволяет улучшать итоговую точность классификации.

Математически итоговая модель AdaBoost представляется как взвешенная сумма решений отдельных слабых классификаторов:

$$H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right),$$

где  $h_t(x)$  – слабый классификатор на  $t$ -й итерации,  $\alpha_t$  – вес, зависящий от точности соответствующего классификатора.

В условиях эксперимента алгоритм показал устойчивость к шуму и артефактам в визуальных данных, а также способность точно выделять элементы городской инфраструктуры, такие как здания, дороги и уличные объекты. При сравнении с более тяжёлыми нейросетевыми архитектурами AdaBoost продемонстрировал высокую обобщающую способность и низкое время обработки, особенно в задачах бинарной и многоклассовой классификации при ограниченных объёмах данных. Эти качества делают его применимым в условиях реального времени, когда критично учитывать вычислительные ресурсы (Wang et al., 2024; Weng et al., 2024).

Идея состоит в том, чтобы поместить файлы в более широкий социально-экономический контекст и определить с теоретической точки зрения связи, которые они представляют с другими средствами наблюдения, а также с городской средой, в которой они находятся. Такой подход позволяет выявить зависимости между зафиксированными визуальными данными и пространственными характеристиками городской инфраструктуры, включая плотность застройки, транспортные потоки и активность населения (Li et al., 2023; Niu, 2024). Эти взаимосвязи наглядно представлены на рисунке 1, где визуализируется городская сцена с учётом слоёв социального, функционального и пространственного анализа, демонстрируя, как методы машинного обучения интегрируются в систему территориального планирования



**Рисунок 1.** Городская сцена Алматы в дневное время

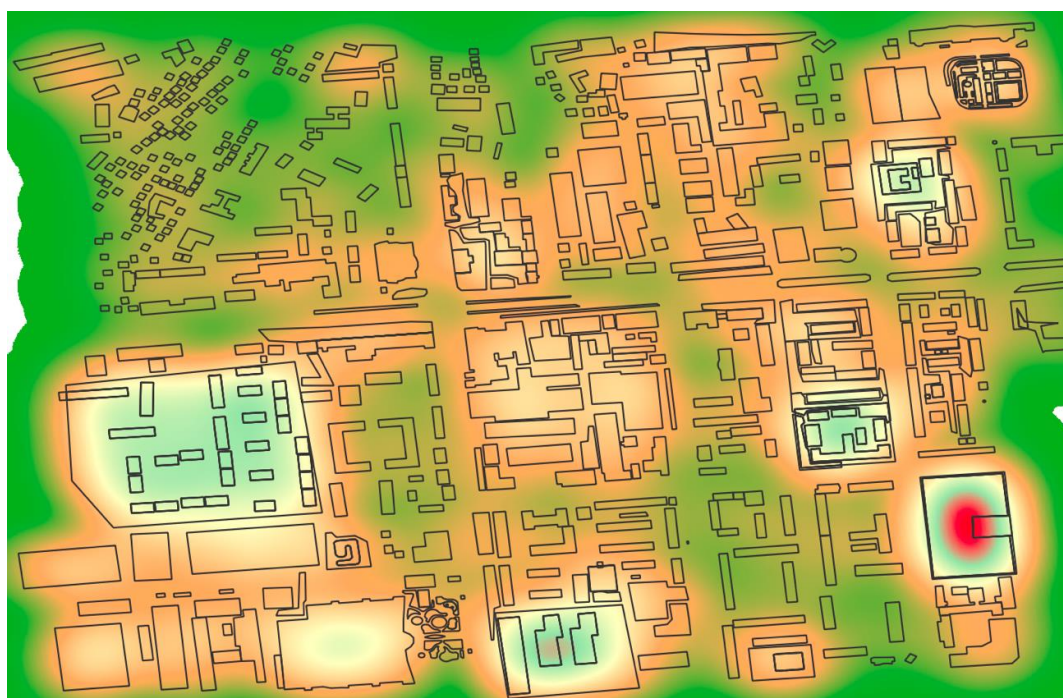
*Примечание – снято авторами в рамках исследования. Снимок сделан в центральной части города. Изображение используется для анализа транспортной нагрузки, плотности застройки и*

*визуального контекста городской среды. Данные могут быть использованы в системах территориального планирования и машинного обучения для извлечения признаков городской морфологии*

Особое внимание уделено процессу тестирования обученных моделей на независимых наборах данных. В условиях, приближённых к реальной съёмке городской сцены, уровень точности классификации напрямую зависит от стабильности точек интереса, ракурса и освещённости. Эти параметры критичны при создании пространственно точных 3D-моделей городской среды. Глубокие нейросети, способные к многослойной абстракции признаков, позволили адаптироваться к этим вызовам и улучшить пространственную реконструкцию (Buyukdemircioglu et al., 2022).

В рамках задач территориального планирования была апробирована методика самообучения и глубокого представления визуальных данных. Это позволило перейти от пиксельного уровня к концептуальным пространственным меткам, критически важным для 3D-моделирования городской среды. Регуляризация и оптимизация моделей минимизировали переобучение и улучшили переносимость решений между различными типами визуальных данных (Chen et al., 2025).

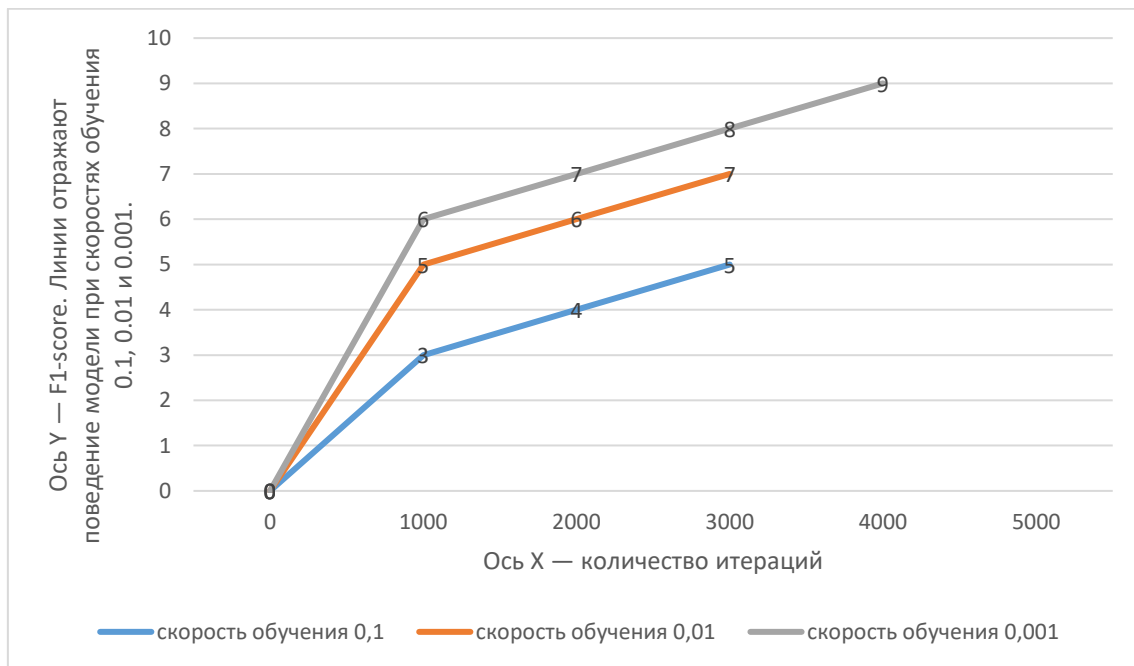
Такой подход позволяет выявить зависимости между зафиксированными визуальными данными и пространственными характеристиками городской инфраструктуры, включая плотность застройки, транспортные потоки и активность населения. На рисунке 2 представлен пример участка центральной части Алматы, где в единой визуализации совмещены контуры зданий и тепловое распределение активности. Такая интеграция даёт возможность одновременно рассматривать социальные, функциональные и пространственные слои, что демонстрирует потенциал применения методов машинного обучения для построения трёхмерных моделей городской среды и их использования в системах территориального планирования.



**Рисунок 2.** Участок центральной части Алматы (Казахстан, Алматы)

*Примечание – данные были обработаны и визуализированы с использованием программного обеспечения QGIS; исходное изображение зафиксировано авторами в рамках исследования*

Методы деревьев решений и случайных лесов использовались для анализа городской инфраструктуры, выделяя функциональные зоны по визуальным признакам. Для оценки влияния гиперпараметров дополнительно исследовалась зависимость качества классификации от скорости обучения и числа итераций. Как видно из графика, уже на первых сотнях-тысячах итераций модели на деревьях достигают стабильных значений F1-score и IoU, обучаясь существенно быстрее и с меньшими вычислительными затратами, чем более тяжёлые нейросетевые архитектуры (U-Net, ResNet-50, ViT). При увеличении числа шагов обучения оптимальные результаты достигаются при меньшей скорости (0,001), что обеспечивает более плавную сходимость и более глубокий разбор признаков. Полученные зависимости важны для настройки алгоритмов в автоматизированных системах территориального моделирования (Chajaei & Bagheri, 2025).



**Рисунок 3.** Эволюция скорости обучения метода деревьев решений при анализе фото- и видеофайлов

*Примечание – составлено авторами на основе результатов. По оси X отложено количество итераций, по оси Y – нормированное значение F1-score/IoU; линии отражают поведение модели при скоростях обучения 0,1, 0,01 и 0,001. Методы деревьев решений демонстрируют более быстрое обучение по сравнению с другими моделями, при близких значениях IoU и F1-score. Составлено автором на основе экспериментальных данных.*

Важный результат – использование градиентного спуска в качестве метода оптимизации при построении моделей, анализирующих видео- и фотоинформацию с целью восстановления трёхмерной структуры городской сцены. Метод показал стабильную сходимость даже при наличии искажений в исходных изображениях.

Кластерный анализ городских объектов проводился с использованием формулы несходства Брея-Кертиса:

$$\frac{\sum T t = 1 |x_{it} - x_{jt}|}{\sum T t = 1 (x_{it} + x_{jt})}$$

что позволило определить пространственные группы объектов и участков городской инфраструктуры. Это облегчает задачи урбанистического планирования и распределения функциональных зон [18].

Для прогнозирования функционального назначения участков использовался байесовский классификатор:

$$P(A/B) = P(A \cap B) / P(B) = [P(A) * P(B/A)] / P(B),$$

который показал высокую эффективность при работе с неполными или шумными данными городской съёмки. В условиях реального городского пространства, где часто отсутствует полная информация, такой подход обеспечивает устойчивость выводов и масштабируемость в рамках систем территориального планирования (Lynda et al., 2025).

Особого внимания заслуживает применение архитектуры Vision Transformer (ViT), разработанной в Токуо Tech, при анализе визуальных данных без использования традиционных линз. Данный метод показал высокую точность в реконструкции изображений городской среды и открыл перспективы для компактных сенсорных систем, применимых в пространственном мониторинге (Pan et al., 2022).

**Таблица 1.** Сравнение эффективности моделей машинного обучения на тестовом наборе данных (городская сцена, Алматы)

Модель	IoU (%)	F1-score	MAE (м)	Время обработки (сек/кадр)
Logistic Regression	58,3	0,69	0,34	0,12
Decision Tree	61,2	0,71	0,31	0,15
Random Forest	66,5	0,75	0,27	0,28
U-Net	74,2	0,83	0,21	0,95
ResNet-50	69,5	0,78	0,26	0,82
ViT	77,1	0,85	0,19	1,25

*Примечание – составлено авторами*

Архитектура Vision Transformer (ViT) продемонстрировала наивысшие показатели по метрикам IoU (77,1 %) и F1-score (0,85), что свидетельствует о её высокой точности при семантической сегментации городской сцены. Вместе с тем ViT характеризуется наибольшим временем обработки (1,25 сек/кадр), что может быть ограничивающим фактором в приложениях реального времени. U-Net обеспечила сбалансированное соотношение между качеством и вычислительной эффективностью (IoU – 74,2 %, время – 0,95 сек/кадр), что делает её оптимальной для практического применения в системах мониторинга городской инфраструктуры. ResNet-50 показала достойный результат по F1-score (0,78), однако по точности и скорости уступает как ViT, так и U-Net. Это может быть связано с ограниченной способностью модели эффективно обрабатывать как глобальные, так и локальные признаки при сравнительно высокой глубине сети. Классические алгоритмы – логистическая регрессия, дерево решений и случайный лес – продемонстрировали более низкие значения метрик точности (IoU < 67 %), однако значительно превосходят нейросетевые подходы по скорости обработки. Это определяет их применимость в задачах предварительной фильтрации данных или в условиях ограниченных вычислительных ресурсов. Метрика MAE (средняя абсолютная ошибка) позволяет дополнительно оценить точность восстановления геометрических характеристик объектов. Наименьшее значение MAE (0,19 м) зафиксировано у модели ViT, за ней следует U-Net (0,21 м), что подтверждает их превосходство в реконструкции пространственных структур.

Таким образом, различия в производительности моделей объясняются их архитектурными особенностями, уровнем абстракции признаков и степенью устойчивости к шуму. Глубокие нейронные сети демонстрируют высокую точность при увеличенных вычислительных затратах, тогда как традиционные алгоритмы обеспечивают оперативную обработку, но уступают в детальности и способности к обобщению. Выбор модели в практическом применении должен определяться требованиями к точности, скорости и ресурсам целевой системы.

## ЗАКЛЮЧЕНИЕ

Настоящее исследование было направлено на оценку применимости алгоритмов машинного обучения к задачам пространственного анализа городской среды на основе визуальных данных. Проведённое сравнение классических и нейросетевых моделей позволило выявить существенные различия в их производительности, интерпретируемости и применимости к условиям реального времени. Нейросетевые архитектуры, в частности Vision Transformer (ViT) и U-Net, продемонстрировали наилучшие показатели точности в задачах сегментации и 3D-реконструкции, тогда как классические методы, включая логистическую регрессию и деревья решений, показали лучшие результаты по скорости обработки и устойчивости при ограниченных вычислительных ресурсах.

В ходе экспериментов была подтверждена эффективность использования глубоких моделей при наличии полноценных аннотированных датасетов и стабильных условий визуального захвата. Тем не менее высокая вычислительная сложность некоторых архитектур ограничивает их применение в системах реального времени без предварительной оптимизации. Особое значение имеет правильная настройка параметров моделей — глубины, размера входного изображения, количества эпох — что в случае нейросетей существенно влияет на результативность и сходимость. Несмотря на полученные метрики, в анализе результатов остаются нерешёнными вопросы интерпретации различий между архитектурами, объяснение которых требует дополнительного изучения чувствительности моделей к типам признаков и качеству входных данных.

Исследование сопровождается рядом ограничений. Во-первых, отсутствует единый протокол стандартизации визуальных данных, полученных из различных источников. Во-вторых, модели не тестировались на мультисенсорных данных (LiDAR, тепловизоры), что ограничивает универсальность результатов. В-третьих, не была охвачена оценка устойчивости моделей к шуму, искажениям и неполноте данных, характерным для городских условий наблюдения.

В качестве перспективных направлений дальнейших исследований следует отметить интеграцию моделей машинного обучения с геоинформационными системами, применение генеративных методов для автоматического дополнения недостающих участков городской сцены, а также развитие архитектур с учётом требований explainable AI. Расширение экспериментальной базы и внедрение методов самообучения представляют собой ключевые векторы повышения масштабируемости и адаптивности разрабатываемых решений в рамках концепции «умного города».

Таким образом, полученные результаты подтверждают научную и практическую значимость комплексного подхода, сочетающего визуальную аналитику, методы компьютерного зрения и машинного обучения для пространственного моделирования городской среды. Разработанная методология может служить основой для построения автоматизированных систем мониторинга, анализа и планирования городской инфраструктуры в условиях быстро развивающихся урбанистических пространств.

**КОНФЛИКТ ИНТЕРЕСОВ:** Авторы заявляют об отсутствии конфликта интересов.

**ФИНАНСИРОВАНИЕ:** Данное исследование выполнено в рамках подготовки диссертации на соискание степени доктора философии (PhD) по государственному образовательному заказу без привлечения дополнительного грантового финансирования. Работа проведена в Satbayev University в рамках образовательной и исследовательской деятельности докторанта, а также совместно с Университетом Нархоз и Университетом международного бизнеса им. К. Сагадиева.

**БЛАГОДАРНОСТИ:** Авторы выражают искреннюю признательность коллегам из Satbayev University за методологическую поддержку, активное участие в обсуждении теоретических и прикладных аспектов применения методов машинного обучения, а также за помощь в интерпретации полученных результатов. Отдельная благодарность выражается анонимным рецензентам за ценные замечания, способствовавшие повышению научной обоснованности и качества статьи. Поддержка исследовательской среды Satbayev University сыграла важную роль в реализации данного проекта, направленного на развитие цифровых подходов в системах территориального планирования.

**УВЕДОМЛЕНИЕ ОБ ИСПОЛЬЗОВАНИИ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА:** В процессе подготовки настоящей научной статьи авторы использовали возможности генеративного искусственного интеллекта исключительно для улучшения формулировок, редактирования и структурирования текста. Все научные идеи, аналитика, интерпретация результатов и выводы являются исключительно авторскими. Генеративные модели не применялись для создания данных или подмены научного анализа.

#### СПИСОК ЛИТЕРАТУРЫ

- Akhavi Zadegan, A., Vivet, D., Hadachi, A. (2025). Challenges and advancements in image-based 3D reconstruction of large-scale urban environments: a review of deep learning and classical methods. *Frontiers in Computer Science*, 7. <https://doi.org/10.3389/fcomp.2025.1467103>
- Akylbekov, O. N., Kasenkhan, A. M., & Baidildinova, S. T. (2025). AI methods in education: A global review and examples of application in Kazakhstan. *Bulletin of D. Serikbayev EKTU*, pp. 103–113. <https://doi.org/10.51885/1561-421220252103>
- Akylbekov, O., Al Said, N., Martínez-García, R., & Gura, D. (2022). ML models and neural networks for analyzing 3D data spatial planning tasks: Example of Khasansky urban district of the Russian Federation. *Advances in Engineering Software*, 173, 103251. <https://doi.org/10.1016/j.advengsoft.2022.103251>
- Buyukdemircioglu, M., et al. (2022). Deep learning for 3D building reconstruction: A review. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2022, 359–366. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-359-2022>
- Čahojová, L., Jarolímek, I., Klímová, B., Kollár, M., Michalková, M., Mikula, K., Ožvat, A. A., Slabejová, D., & Šibíková, M. (2025). Integration of ground-based and remote sensing data with deep learning algorithms for mapping habitats in Natura 2000 protected oak forests. *Basic and Applied Ecology*. <https://doi.org/10.1016/j.baae.2025.01.006>
- Chajaei, F., & Bagheri, H. (2025). LOD1 3D City Model from LiDAR: The Impact of Segmentation Accuracy on Quality of Urban 3D Modeling and Morphology Extraction. *Remote Sensing Applications: Society and Environment*. <https://doi.org/10.1016/j.rsase.2025.101534>
- Chen, S., Wang, Z., Hong, C., Sun, Y., Jia, H., & Liu, W. (2025). Self-supervised learning of end-to-end 3D LiDAR odometry for urban scene modeling. *Remote Sensing*, 17(15), 2661. <https://doi.org/10.3390/rs17152661>

- Li, F., et al. (2023). Machine learning and remote sensing integration for urban studies. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2023.104653>
- Li, Y., et al. (2024). A method for extracting buildings from remote sensing imagery using a 3D Joint Attention module. *Scientific Reports*. <https://www.nature.com/articles/s41598-024-70019-z>
- Liu, J., et al. (2023). Deep learning-based multi-view stereo matching and 3D real-scene reconstruction framework (Deep3D). *Computers & Geosciences*. <https://doi.org/10.1016/j.isprsjprs.2023.08.015>
- Lynda, D., Logeswari, G., Tamilarasi, K., & Rakesh, S. (2025). Hybrid Bayesian deep learning model for predicting urban heat island intensity in African cities. *Scientific Reports*. <https://doi.org/10.1038/s41598-025-13492-4>
- Moradi, L. (2025). Implications of Artificial Intelligence for Assessing the Built Environment: Urban data collection methods and machine learning algorithms. *Journal of Urban Planning and Development*. <https://doi.org/10.1080/10630732.2025.2468142>
- Niu (2024). Overview of image-based 3D reconstruction technology. *JEOS: Journal of the European Optical Society*. <https://doi.org/10.1051/jeos/2024018>
- Pan, J., et al. (2025). Machine Learning-Enhanced 3D GIS Urban Noise Simulation. *Geoinformatics Journal*, 14(6), 223.
- Pan, X., Chen, X., Takeyama, S., & Yamaguchi, M. (2022). Image reconstruction with transformer for mask-based lensless imaging. *Optics Letters*, 47(7), 1843–1846. <https://doi.org/10.1364/OL.455378>
- Rodríguez-Lira, D.-C., Córdova-Esparza, D.-M., Terven, J., Romero-González, J.-A., Alvarez-Alvarado, J. M., González-Barbosa, J.-J., & Ramírez-Pedraza, A. (2025). Recent developments in image-based 3D reconstruction using deep learning: Methodologies and applications. *Electronics*, 14(15), 3032. <https://doi.org/10.3390/electronics14153032>
- Wang, C., et al. (2024). Deep learning-based 3D reconstruction from multiple images. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2024.128018>
- Weng, Q., et al. (2024). How will AI transform urban observing, sensing, imaging, and mapping? *Nature*. <https://doi.org/10.1038/s42949-024-00188-3>
- Yu Shang et al. (2024). UrbanWorld: An Urban World Model for 3D City Generation. arXiv preprint arXiv:2407.11965. <https://arxiv.org/abs/2407.11965>

**Авторлар туралы мәліметтер**  
**Информация об авторах**  
**Information about authors**



**Ақылбеков Олжас Наурызбаевич** – Ақпараттық технологиялар магистрі, Қ.И. Сәтбаев атындағы Қазақ ұлттық зерттеу техникалық университеті, Алматы, Қазақстан

**Ақылбеков Олжас Наурызбаевич** – Магистр информационных технологий, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан

**Akylbekov Olzhas Nauryzbayevich** – Master of Information Technology, Kazakh National Research Technical University named after K.I. Satpayev, Almaty, Kazakhstan

e-mail: o.akylbekov@satbayev.university

ORCID: <https://orcid.org/0000-0002-7188-5550> ,



**Молдагулова Айман Николаевна** – физика-математика ғылымдарының кандидаты, Қ.И. Сәтбаев атындағы Қазақ ұлттық зерттеу техникалық университеті, Алматы, Қазақстан

**Молдагулова Айман Николаевна** – Кандидат физико-математических наук, Казахский Национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Қазақстан

**Moldagulova Aiman Nikolaevna** – Candidate of Physical and Mathematical Sciences, Kazakh National Research Technical University named after K.I. Satpayev, Almaty, Kazakhstan,  
e-mail: a.moldagulova@satbayev.university,  
ORCID: <https://orcid.org/0000-0002-1596-561X>,



**Закария Гульназ Саянкызы** – Қолданбалы геоинформатика ғылымдарының магистрі, Қ.И. Сәтбаев атындағы Қазақ ұлттық зерттеу техникалық университеті, Алматы, Қазақстан

**Закария Гульназ Саянкызы**– Магистр наук в области прикладной геоинформатики, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан

**Zakariya Gulnaz Sayankyzy** – Master of Science in Applied Geoinformatics, Kazakh National Research Technical University named after K.I. Satpayev, Almaty, Kazakhstan,  
e-mail: g.zakariya@satbayev.university,  
ORCID: <https://orcid.org/0009-0001-7774-7634>



**Байдильдинова Сабина Талгатовна** – педагогика ғылымдарының магистрі, Қ. Сағадиев атындағы Халықаралық бизнес университеті, Алматы, Қазақстан

**Байдильдинова Сабина Талгатовна** – Магистр педагогических наук, Университет международного бизнеса им. К. Сағадиева, Алматы, Казахстан

**Baidildinova Sabina Talgatkyzy** – Master of Pedagogical Sciences, K. Sagadiyev University of International Business, Almaty, Kazakhstan,  
e-mail: baidildinova.s@gmail.com,  
ORCID: <https://orcid.org/0009-0002-5026-6504>



**Бекарыстанкызы Акбаян** – PhD, қауымдастырылған профессор, Цифрлық технологиялар мектебі, NARXOZ Университеті, Алматы қ., Қазақстан

**Бекарыстанкызы Акбаян** – PhD, доцент Школы цифровых технологий, Университет Нархоз, г. Алматы, Казахстан.

**Bekarystankyzy Akbayan** – PHD, Associate Professor, School of Digital Technologies, Narхоз University, Almaty, Kazakhstan,  
e-mail: akbayan.b@gmail.com,  
ORCID: <https://orcid.org/0000-0003-3984-2718>