TASHKENT STATE UNIVERSITY OF ECONOMICS

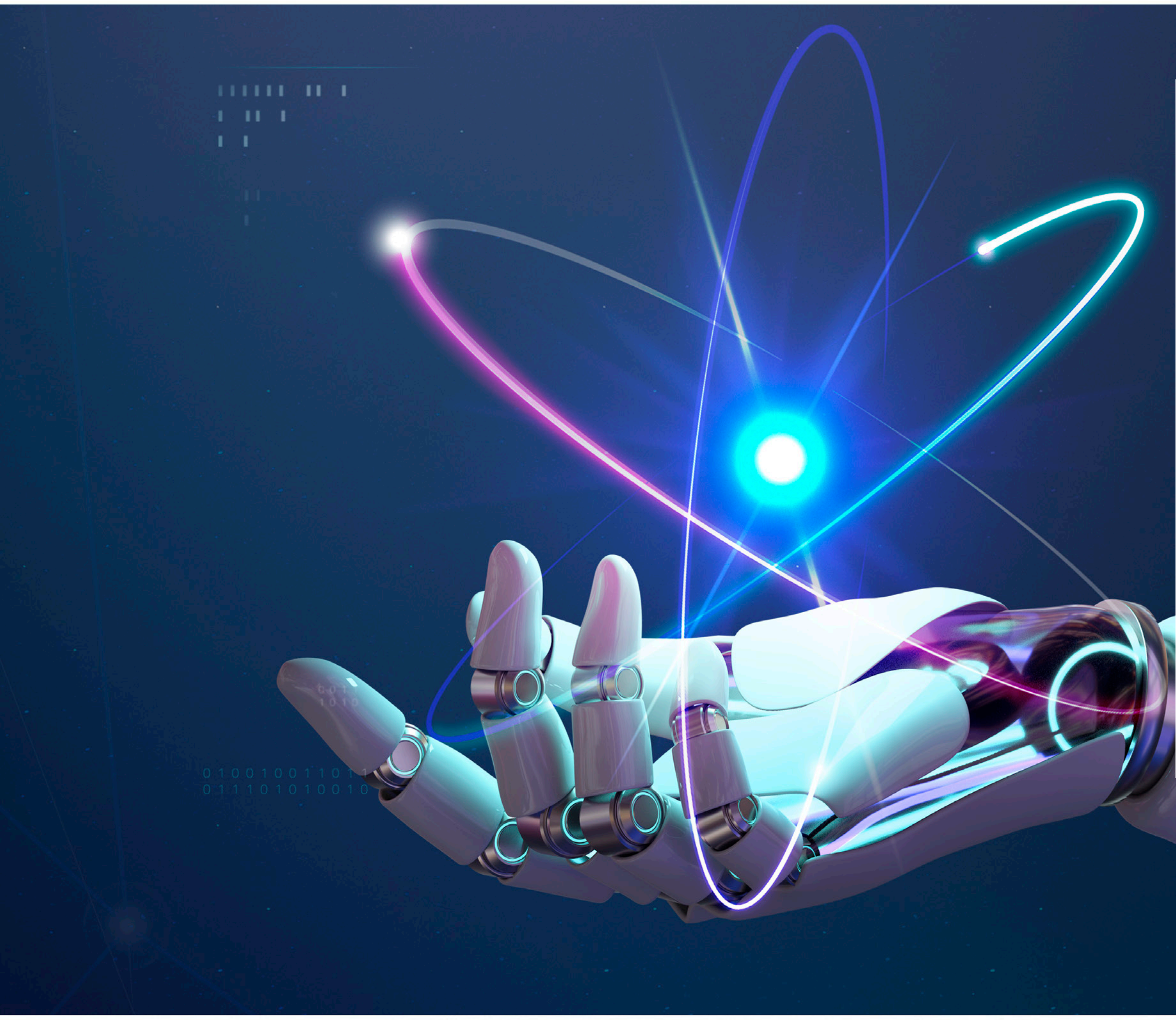
International journal of topics

INNOVATION SCIENCE AND TECHNOLOGY

Scopus || Electronic journal specializing in Scopus

ISSUE 3

Acceptance of papers **March, 2026**



Acceptance of papers

Published monthly

Topics

economics, technology, social sciences

ISSN 3060-5229

OpenAIRE Google Scholar Academic Resource Index ResearchBib BASE doi Digital Object Identifier OPEN ACCESS

Visit the website
t.me/scopus_IST2100

Electronic publication, Issue 3. 265 pages.
Approved for publication on March, 2026.

THE SCIENTIFIC-POPULAR ELECTRONIC JOURNAL **"INNOVATION SCIENCE AND TECHNOLOGY"** HAS BEEN REGISTERED UNDER THE NUMBER **C-5669633** BY THE AGENCY FOR INFORMATION AND MASS COMMUNICATIONS (AOKA) OF THE REPUBLIC OF UZBEKISTAN, EFFECTIVE FROM OCTOBER 9, 2024.

CONTACTS

Phone: **+998 50 737 87 88**

Website: https://ist-journal.uz

Email: innovationist2025@gmail.com

The scientific electronic journal “Innovation Science and Technology” has been included in the list of scientific publications recommended for the publication of main scientific results of dissertations for the award of PhD and DSc degrees in economics and technical sciences, in accordance with the Resolution No. 370 of the Presidium of the Higher Attestation Commission of the Republic of Uzbekistan, dated May 8, 2025.

Editorial board:

CONTENTS

CONTENTS

CONTENTS

A MODEL FOR DETECTING URBAN INFRASTRUCTURE PROBLEMS IN CITIZENS' APPEALS BASED ON GEOLOCATION FEATURES

Mallayev Oybek Usmankulovich
Professor of the Department of Digital Technologies
Alfraganus University
E-mail: o.mallayev@afu.uz

Gazatov Jamoliddin Abduvoidovich
Independent Researcher
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
E-mail: jamoliddingazatov@mail.com

Aliyev Jaloliddin Kokand oglu
Independent Researcher
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
E-mail: jaloliddinaliyev0@gmail.com

Abstract: This article proposes a model for identifying and classifying urban infrastructure problems based on geolocation data derived from citizens' appeals. The study extracts features from geolocation points submitted by citizens, including timestamps and movement parameters.

Based on these features, methods have been developed for the automatic detection of issues such as traffic congestion, road damage, waste accumulation, and traffic signal malfunctions. The proposed model enables the classification of urban problems using machine learning algorithms.

The research results demonstrate that the use of geolocation-based citizen appeal data significantly enhances the efficiency of identifying urban issues and supports faster and more informed decision-making processes in urban management systems.

Key words: geo-map, machine learning, interactive services, GPS, NLP.

INTRODUCTION

In recent years, within the framework of the Smart City concept, issues related to the efficient and sustainable management of urban infrastructure have gained increasing importance. Population growth, the expansion of transportation flows, and the acceleration of urbanization processes are contributing to the growing need for advanced and adaptive urban management solutions [1].

One of the key sources for identifying urban infrastructure challenges is citizens' appeals. Through these appeals, residents provide valuable information about situations such as traffic congestion, road surface deterioration, waste accumulation, or traffic signal disruptions. In traditional systems, such appeals are typically recorded manually, and their analysis may require significant time and effort.

However, the rapid development of mobile devices and geolocation technologies has created new opportunities for the automated identification of urban infrastructure issues. By utilizing geolocation data, it is possible to obtain detailed information about a user's position, movement speed, time, and direction parameters, which can be effectively used for analytical purposes [2].

The purpose of this study is to develop a model for identifying and classifying urban infrastructure issues based on geolocation data obtained from citizens' appeals.

Geolocation refers to a set of data representing the spatial position of an object, event, or user within a coordinate system. In the analysis of citizens' appeals, geolocation is not limited to coordinates alone but serves as a multi-dimensional descriptor enriched with temporal and movement-related attributes. Therefore, it is considered a significant digital feature for identifying and analyzing urban infrastructure conditions.

In the proposed model, latitude, longitude, timestamp, GPS accuracy, speed, and direction are selected as input parameters. These parameters form the foundation for spatio-temporal modeling and enable a more comprehensive analysis of urban infrastructure dynamics.

LITERATURE REVIEW

Numerous scientific studies have examined the use of geolocation data and spatial analytics in urban infrastructure management. With the rapid development of digital technologies, urban systems increasingly rely on large-scale spatial data to enhance monitoring and support informed decision-making processes. Batty emphasizes that big data and geospatial information play a critical role in modern urban governance by enabling city authorities to analyze infrastructure conditions and improve urban services through data-driven approaches [3].

One of the important research directions in this field is urban computing, which integrates geospatial data, mobile sensing, and artificial intelligence to analyze urban dynamics. Zheng and colleagues introduced the concept of urban computing and proposed analytical methods for detecting traffic congestion using GPS trajectory data [1]. Their research demonstrates that movement speed, stop duration, and trajectory patterns can serve as key indicators for assessing traffic conditions in urban transport systems.

Clustering algorithms are widely applied for identifying spatial patterns and problem concentration areas. In particular, the DBSCAN algorithm has become one of the most commonly used density-based clustering methods for analyzing geolocation datasets [4]. DBSCAN enables the identification of clusters of spatial points within a defined radius and supports the detection of areas where specific urban issues are more frequently observed. This approach is especially useful for identifying congestion zones, waste accumulation points, and infrastructure-related irregularities.

Machine learning techniques have also demonstrated strong effectiveness in analyzing geolocation and urban data. The XGBoost algorithm, developed by Chen and Guestrin, is widely recognized for its ability to process large datasets with high computational efficiency and predictive accuracy [5]. Due to its gradient boosting mechanism, XGBoost has been successfully applied in various domains, including transportation analysis, infrastructure monitoring, and predictive analytics.

In addition to traditional machine learning methods, deep learning approaches have been increasingly applied in spatial data analysis. Goodfellow and Bengio highlight that deep learning models are capable of extracting complex patterns from large datasets and learning hierarchical representations of spatial and temporal data [7]. These capabilities enable deep neural networks to enhance the performance of intelligent urban monitoring systems.

Aggarwal notes that machine learning techniques provide effective tools for discovering hidden patterns in large-scale datasets and support advanced data analytics in smart city environments [8]. Similarly, Bishop emphasizes that probabilistic machine learning models play an important role in pattern recognition and classification tasks, enabling robust and reliable decision-making in complex data-driven systems [9].

Furthermore, Townsend discusses the concept of smart cities, where digital technologies, big data analytics, and intelligent information systems are integrated to enhance urban governance and infrastructure management [10]. According to this concept, real-time data obtained from mobile devices, sensors, and digital platforms significantly improves the responsiveness and efficiency of urban services.

Despite the growing number of studies in urban computing and geospatial analytics, most existing research primarily focuses on transportation monitoring and traffic analysis. At the same time, citizen-generated data—particularly geolocation information obtained from citizen appeals—offers additional opportunities and remains comparatively underutilized as a valuable data source for identifying urban infrastructure conditions.

Therefore, there is a clear research gap in developing integrated models that combine geolocation features, contextual information, and machine learning techniques for analyzing citizen appeals. Addressing this gap requires the development of a comprehensive analytical model capable of identifying infrastructure-related situations such as traffic congestion, road surface degradation, waste accumulation, and traffic signal irregularities based on geolocation data.

The model proposed in this research aims to address this need by integrating geolocation feature extraction, spatial clustering techniques, and machine learning classification methods to automatically identify urban infrastructure conditions using citizen-generated geolocation data.

RESEARCH METHODOLOGY

The fundamental structure of geolocation data consists of the following components:
- Spatial component – latitude and longitude coordinates;
- Temporal component – the time when the appeal was submitted or when the point was recorded;
- Movement component – speed, direction, acceleration, and stop frequency;
- Accuracy component – GPS accuracy, signal quality, and device error;
- Context component – area type, street segment, proximity to objects, time of day, and day of the week.

These components transform geolocation into a comprehensive information source that answers not only the question "Where?", but also "When?", "Under what conditions?", "With what intensity?", and "In what environment?". In this regard, geolocation features have high diagnostic value for identifying situations reflected in citizens' appeals, such as traffic congestion, road surface deterioration, waste accumulation, or traffic signal disruptions.

The main properties of geolocation are as follows:
- spatial accuracy – indicates the precise location of a situation;
- dynamic nature – changes over time;
- noise – GPS inaccuracies and imprecise points may occur;
- clustering – similar situations tend to concentrate in specific areas;
- context dependence – the same coordinates may have different meanings depending on time and conditions.

For example, the concentration of coordinates in a certain area does not always indicate a specific issue. If the speed at those points is low, the stop ratio is high, and the time corresponds to morning peak hours, this may indicate traffic congestion. If the same location is near a traffic signal node and similar appeals are received from multiple users, this increases the likelihood of a traffic signal disruption.

Feature extraction from geolocation data is carried out in several stages. In the first stage, raw coordinates are cleaned by removing duplicate points, low-accuracy GPS records, unrealistic jumps, and records with inconsistent time sequences. Subsequently, the data are converted into a unified time format and spatial normalization is performed.

In the second stage, primary features are generated, including:
- latitude and longitude;
- timestamp;
- GPS accuracy;
- speed;
- direction;
- distance between two points;
- time interval.

In this study, it is recommended to calculate distance using the Haversine formula, while speed is derived from the distance and time difference. This approach ensures the accurate calculation of geographical distance between spatial points.

In the third stage, derived features are generated, which are particularly useful for modeling:
- average speed;
- maximum and minimum speed;
- stop duration;
- stop ratio;
- abrupt changes in direction;
- point density within an area;
- number of nearby appeals;
- frequency of repeated signals from the same location.

Among these, stop duration and stop ratio are identified as key indicators for detecting traffic congestion. A high stop ratio combined with low average speed increases the likelihood of congestion.

In the fourth stage, spatial-cluster features are extracted using density-based clustering algorithms such as DBSCAN. Through DBSCAN, points located within a specified radius are grouped into clusters and identified as hotspots [7]. This approach is particularly effective for detecting waste accumulation, road surface irregularities, or frequently recurring infrastructure-related situations.

In the fifth stage, semantic-contextual features are generated. These are derived not directly from geolocation data but from the contextual information associated with it:
- time of day (morning, midday, evening peak);
- day of the week (weekday or weekend);

- area type (highway, inner street, residential zone, intersection);
- proximity to objects (traffic lights, bus stops, waste containers, schools, markets);
- keywords extracted from the appeal text.

At this stage, geolocation features are enriched with textual features obtained through natural language processing (NLP). As a result, the model is capable of linking spatial signals with the semantic meaning of the associated text, thereby improving the accuracy and interpretability of urban infrastructure analysis.

In this study, a multi-stage model was developed to detect urban problems based on geolocation data. The data are collected through mobile devices and a Telegram bot. Each appeal contains the following parameters:

geographic coordinates (latitude, longitude)

timestamp

GPS accuracy

movement speed

direction

These data serve as the main input parameters for detecting urban problems. A citizen appeal is represented as a sequence of geolocation points:

$$T_i = \{p_{i1}, p_{i2}, ..., p_{in}\} \quad (1)$$

where:

$$p_i = (\varphi_k, \lambda_k, ..., t_k\} \quad (2)$$

φ - latitude, λ - longitude, and t – the timestamp.

The distance between two points is determined using the Haversine formula [6]:

$$d = 2R \arcsin \sqrt{\sin^2 \frac{\Delta\varphi}{2} + \cos\varphi_1 \cos\varphi_2 \sin^2 \frac{\Delta\lambda}{2}} \quad (3)$$

where R is the Earth's radius (6371 km6371 \, km6371km). Movement speed is calculated by the following formula:

$$v = \frac{d}{\Delta t} \quad (4)$$

where d is distance and, Δt - is the time difference.

Stop duration is calculated as follows:

$$StopTime = \sum \Delta t \cdot I(v \in \ ) \quad (5)$$

where I is the indicator function. The stop ratio is determined as:

$$StopRatio = \frac{StopTime}{T} \quad (6)$$

If StopRatio is high and the average speed is low, this indicates traffic congestion [1]. The DBSCAN algorithm is used to cluster geolocation points [4]. The core point condition is defined by Formula (7):

$$|N\varepsilon(p)| \geq MinPts \quad (7)$$

where ε - is the radius and MinPts is the minimum number of points. To identify problems, the XGBoost algorithm was used [5]. The Softmax probability function is expressed as:

$$P(y = k \mid x) = \frac{e^{z_k}}{\sum e^{z_j}} \quad (8)$$

The model identifies the following classes:

- traffic congestion
- road pothole
- waste accumulation
- traffic light malfunction (Table 1).

Table 1. Main features corresponding to each problem type

Problem Type	Most Important Geofeatures	Additional Indicators
Traffic congestion	avg_speed, stop_ratio, cluster_density	Peak time, proximity to intersection
Road pothole	sharp decrease in speed, repeated hotspot, direction change	Keywords “deep”, “hole”
Waste accumulation	static hotspot, cluster_density	Proximity to waste container, repeated complaint
Traffic light malfunction	stop_duration, intersection proximity, dist_traffic_light	Keywords “traffic light”, “not working”

This table clearly demonstrates the semantic relationship between the problem type and the extracted features.

ANALYSIS AND RESULTS

In the experimental study, 1,200 geolocation-based citizen appeals were analyzed (Table 2).

Table 2. Classification Performance of the Proposed Model

Problem Type	Precision	Recall	F1
Traffic jam	0.93	0.91	0.92
Pothole	0.86	0.84	0.85
Waste pile	0.90	0.88	0.89
Traffic light issue	0.91	0.90	0.90

The experimental results demonstrate that the XGBoost model provides high classification accuracy across all problem categories.

The highest performance was achieved in detecting traffic congestion, with an F1-score of 0.92. This result can be explained by the strong correlation between congestion patterns and geolocation-derived features such as movement speed, stop duration, and cluster density.

The detection accuracy for potholes was slightly lower (F1 = 0.85). This can be attributed to the fact that pothole detection often depends on additional contextual signals such as textual descriptions or visual confirmation.

Waste accumulation detection achieved an F1-score of 0.89, demonstrating that spatial clustering combined with repeated citizen reports can effectively identify environmental issues in urban areas.

Similarly, traffic light malfunction detection achieved an F1-score of 0.90. This problem type is strongly associated with stop duration and intersection proximity, which makes it easier to detect using geolocation data.

Comparative analysis. Compared to traditional manual complaint analysis methods, the proposed model significantly improves the efficiency of identifying urban infrastructure problems. The integration of geolocation analytics, clustering algorithms, and machine learning classification allows the system to process large volumes of citizen-generated data automatically.

Error analysis. Despite the promising results, several limitations were identified:

- GPS noise and inaccurate coordinates may affect spatial clustering accuracy.
- Some infrastructure problems require visual confirmation.
- Text descriptions provided by citizens may contain ambiguous or incomplete information.

These factors may introduce classification errors, particularly in distinguishing between similar infrastructure problems.

Future improvements. Future research may improve the proposed system by integrating:

- computer vision techniques for image-based problem verification
- deep learning models for more advanced spatial-temporal analysis
- real-time urban monitoring systems based on IoT sensors

Such improvements would enable the development of a more robust and intelligent urban infrastructure monitoring system.

CONCLUSION AND RECOMMENDATIONS

In this study, a model for identifying urban infrastructure conditions based on geolocation data obtained from citizens’ appeals was developed. The proposed model incorporates stages such as feature extraction from geolocation data, detection of spatial hotspots, and classification of urban conditions using machine learning algorithms.

In the research process, citizens' appeals were modeled as sequences of geolocation points. From these sequences, features such as distance, average speed, stop duration, stop ratio, direction dispersion, cluster density, and hotspot score were extracted. These features were integrated with textual descriptions of appeals, timestamps, and external GIS-based contextual data to form a unified and structured feature space.

As a result, a dataset architecture was developed to support the automatic classification of urban conditions such as traffic congestion, road surface irregularities, waste accumulation, and traffic signal disruptions. This architecture enables the effective structuring of geolocation data for machine learning applications and supports their integration into real-world urban management systems.

The obtained results confirm the high predictive performance and reliability of the XGBoost-based model in identifying urban infrastructure conditions.

LIST OF REFERENCES

1. Zheng, Y. (2014). Urban computing. ACM Transactions.
2. Han, J., & Kamber, M. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
3. Batty, M. (2016). Big data and the city. Cambridge, MA: MIT Press.
4. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise (DBSCAN). In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (pp. 226–231).
5. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).
6. Sinnott, R. W. (1984). Virtues of the Haversine. Sky and Telescope, 68(2), 159.
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. Cambridge, MA: MIT Press.
8. Aggarwal, C. C. (2015). Machine learning for data analytics. New York: Springer.
9. Bishop, C. M. (2006). Pattern recognition and machine learning. New York: Springer.
10. Townsend, A. M. (2013). Smart cities: Big data, civic hackers, and the quest for a new utopia. New York: W.W. Norton & Company.

Proofreader: Zokir ALIBEKOV

Layout and Designer: Oloviddin Sobir ugli

2026. № 3

To publish articles in this journal, you may submit articles, advertisements, stories, and other creative materials through the following links. Materials and advertisements are published on a paid basis.

You may subscribe to the journal at any time using the following details. Once subscribed, please send a screenshot or photo of your payment confirmation to our Telegram page @iqtisodiyot_77. Based on this, we will send the latest issue of the journal to your address each month.

“The journal “INNOVATION SCIENCE AND TECHNOLOGY” has been registered by the Agency for Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan from 09.10.2024 under the registration number №390637. License number: C-5669633. PNFL: 30407832680027

Our address: Tashkent city, Yunusobod district, 19th block, House 17.

TASHKENT STATE UNIVERSITY OF ECONOMICS

Google Scholar · OPEN ACCESS · ULRICHSWEB GLOBAL SERIALS DIRECTORY · Academic Resource Index ResearchBib · ISSN INTERNATIONAL STANDARD SERIAL NUMBER INTERNATIONAL CENTRE · CYBERLENINKA · OpenAIRE · ROAD · НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА LIBRARY.RU · BASE · INDEX COPERNICUS INTERNATIONAL · Crossref · doi Digital Object Identifier · CiteFactor Academic Scientific Journals · zenodo

Acceptance of articles

Published every monthly

Directions

Social, economic, political, technological, scientific

Scopus || Scientific electronic journal specializing in Scopus

CERTIFICATE NUMBER: №390637

ORDER NUMBER ACCORDING TO THE LICENSE REGISTER: C-5669633

CONTACT:

Contact us
+998 50 737 87 88

Telegram channel
t.me/scopus_IST2100

Journal official website
https://ist-journal.uz/index.php/IST