UDC 519.67

ENHANCING RURAL INFRASTRUCTURE IN UKRAINE: AN INTEGRATED GEOSPATIAL APPROACH

B. V. Potuzhnyi[1,2,a], V. R. Svirsh[1,2,b], N. M. Kussul[1,3,c]

[1] Institute of Physics and Technology

[2] Engineering and Informational Technology Department, Bern University of Applied Science

[3] Space Research Institute NASU-SSAU

Abstract

This study introduces a unified framework for improving rural infrastructure in Ukraine, combining geospatial data analysis and clustering techniques. By meticulously preparing and validating OpenStreetMap data for over 10 000 villages, we create a reliable foundation for our analysis. Utilizing this validated dataset, we apply advanced clustering to categorize villages by infrastructure quality, highlighting developmental gaps. Our approach merges data validation with an evaluative system to model village infrastructure, offering targeted insights for policy and development strategies. This integrated method provides a way to address infrastructure disparities, enabling data-driven decision-making for rural enhancement. Our findings aim to guide policymakers and development agencies towards strategic interventions, showcasing the potential of combining geospatial analysis with clustering for rural development.

Keywords: Geospatial Analysis, Rural Infrastructure, Data Validation, Clustering, Policy Recommendations

Introduction

In the pursuit of sustainable rural development, leveraging advanced geospatial analysis has emerged as a pivotal strategy. This research builds upon previous studies that have laid the groundwork in employing geospatial data, particularly from OpenStreetMap, to analyze and enhance village infrastructure within Ukraine. Recognizing the critical need for a comprehensive approach to address the disparities in rural infrastructure access, our study integrates meticulous data analysis with innovative clustering techniques. This fusion not only streamlines the identification of infrastructural gaps but also facilitates the formulation of targeted development strategies. By enhancing data validation processes and employing robust statistical models, we ensure the integrity of our approach, setting a precedent for data-driven rural planning. Our work aims to bridge the gap between theoretical models and practical applications, offering a scalable framework for policymakers and development agencies to catalyze rural revitalization efforts efficiently.

[a]bohdan.potuzhnyi@gmail.com
[b]vlada.svirsh25@gmail.com
[c]nataliia.kussul@lll.kpi.ua

1. Related works

1.1. Spatial Patterns and Socio-Environmental Evaluation in Rural Settlements

Recent advancements in geospatial analysis have significantly influenced rural development research. A pivotal contribution is made by the work of Liu et al. [4], which utilized advanced spatial analysis techniques, including kernel density, spatial auto-correlation, and regression analysis, to dissect the complex spatial distribution of rural settlements in Jiangxi Province, China. This study not only uncovers the distinct spatial patterns of rural settlements but also introduces the Socio-Environmental Evaluation Index (SEI) as a novel metric for assessing rural development inequality and guiding revitalization efforts. This research underscores the importance of integrating environmental and socioeconomic data to inform comprehensive rural development strategies.

1.2. Quality of Life Assessment in Ukrainian Rural Areas

In a related vein, Yailymova et al. [3] have conducted an insightful analysis into the quality of life in Ukraine's rural areas. Their methodology, which incorporates a comprehensive assessment of village remoteness from essential infrastructure and natural ecosystems, and consideration of proximity to conflict zones, addresses both the physical and the socio-political landscape affecting rural life quality. Highlighting the disparities faced by villages, especially in conflict-affected regions of Ukraine, their findings provide a data-driven foundation for targeted policy interventions. Together, these studies illustrate the critical role of geospatial analysis in addressing the nuanced challenges of rural development, emphasizing the need for targeted, data-informed approaches to rural revitalization.

Also a lot more other works researching this topic. For example, the work of [1], [8] and [6].

2. Materials

In our research, we utilized a diverse array of geospatial data sources to conduct a comprehensive analysis of rural infrastructure accessibility in Ukraine. The primary source of our spatial data was OpenStreetMap (OSM), an expansive, volunteer-generated map of the world offering detailed information on roads, buildings, and natural features. OSM's open-source nature and global coverage provide a rich foundation for geospatial analysis, enabling the identification and classification of infrastructure across varied landscapes [5].

Complementing OSM data, we also incorporated settlement location information from the Humanitarian Data Exchange (HDX). Managed by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA), HDX is a pivotal resource for accessing up-to-date data on human settlements, particularly in areas affected by crises or in need of humanitarian aid. The HDX platform facilitated our access to the latest positional data on villages and cities within Ukraine as of mid-2021, ensuring our analysis reflected current geographic realities [7].

Another vital component of our dataset was the comprehensive listing of active elevators across Ukraine, sourced from the «Elevators in Ukraine» database. This dataset provided essential details on the distribution and accessibility of vertical transportation facilities, a critical aspect of urban infrastructure that impacts daily mobility and access in densely populated areas [2].

To enrich our analysis further, we employed proximity measures to various critical infrastructure elements, utilizing descriptors such as «distance to the closest objects». This approach, developed and utilized in the work of Yalimovas et al., enabled a nuanced examination of accessibility and infrastructure gaps across rural communities in Ukraine [3]. The integration of these proximity descriptors with our primary geospatial data layers offered a detailed and dynamic view of infrastructure accessibility, laying the groundwork for targeted development initiatives.

The aggregation and synthesis of this diverse data are meticulously detailed in Table 1, providing a structured overview of the geospatial data layers and sources that underpin our analysis. This comprehensive dataset forms the backbone of our study, allowing for an in-depth exploration of rural infrastructure development opportunities within Ukraine.

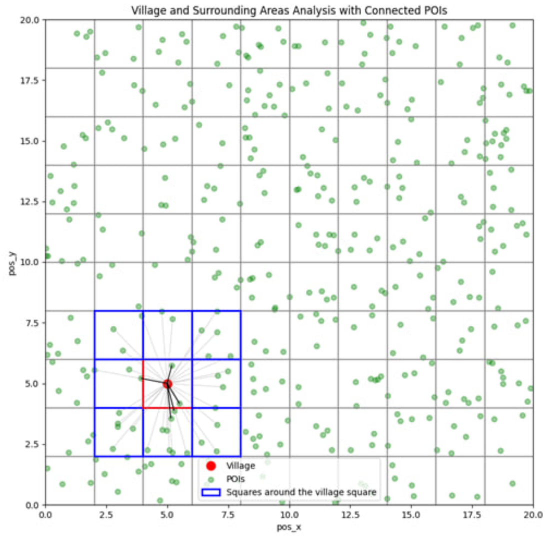

Figure 1. Algorithm visualization for distance calculation and POI selection

3. Creation of the new Graph-based descriptors

3.1. Algorithm for Distance Calculation and POI Selection

To enhance our dataset with detailed descriptors for rural areas in Ukraine, we utilized a specialized algorithm to systematically incorporate Points of Interest (POIs) relative to each village into a graph-structured format. This process is outlined by carefully defining buffer zones around each POI, based on maximum distances delineated in Table 2. Each zone is set to encompass POIs within a reachable perimeter from the villages, ensuring comprehensive data coverage.

Our algorithm initiates by identifying the specific buffer zone in which a village is situated and further considers the adjacent zones to encompass a wider range of POIs. This ensures a thorough inclusion of relevant POIs in our analysis. Within these areas, we prioritize proximity by selecting the closest POIs to each village, adhering to the «maximum quantity inside» criterion specified in Table 2. This criterion is crucial for maintaining a focused dataset that highlights the most accessible amenities to each rural community.

This meticulous process of distance calculation and POI selection is visually represented in Figure 1. The illustration demonstrates how distances from each village to the surrounding POIs are computed, highlighting the methodology's emphasis on spatial relationships. Key to this visualization are the connections (depicted as grey lines) that map the direct paths from villages (red dots) to POIs (green dots), with specific emphasis on the shortest routes to the five closest POIs (shown by black lines). This methodical approach ensures that our dataset not only captures the spatial proximity of essential services to rural areas but also provides an extensive framework for analyzing infrastructure accessibility across Ukraine.

By adopting this algorithm, detailed in Figure 1 and

Table 1. Summary of Geospatial Data Layers and Sources

Data	Layers	Source	Objects	Description
Villages	Villages	The Humanitarian Data Exchange [7]		
Cities	City	The Humanitarian Data Exchange [7]	CITY2_NEAR, Kyiv_NEAR_	These objects describe the distance to the nearest city and capital of Ukraine
Elevators	Elevators	Elevators in Ukraine [2]	Elevators_	Distance to the closest elevator
Roads	Major roads, Secondary roads, Rural roads	OSM [5]	RD_m1_NEAR, RD_m2_NEAR, RD_m3_NEAR	These objects describe the distance to major, regional, and rural roads.
Education	School, College, University	OSM [5]	Education_	Distance to the education
Hotels	Hotel, guesthouse, shelter	OSM [5]	Hotels_NEA	Distance to the closest hotel
Medicine	Hospital, clinic, pharmacy	OSM [5]	Likarni_NE	Distance to the closest hospital
Shops	Supermarket, mall, clothes, marketplace	OSM [5]	Magaz_NEAR	Distance to the closest shop
Bank	Bank	OSM [5]	Bank_NEAR_	Distance to the closest bank
Church	Church	OSM [5]	Cerkva_NEA	Distance to the closest church
Library	Library	OSM [5]	Library_NE	Distance to the closest library
Kindergarten	Kindergarten	OSM [5]	Kinder_NEAR	Distance to the closest kindergarten
Parks	Local Park, National Park, Regional Park	OSM [5]	LokPark_NE, NatPark_NE, regPark_NE	These objects describe the distance to the nearest local park, national park, and regional park

structured according to the specifications in Table 2, we have developed an expansive dataset. This dataset crucially maps out the accessibility of vital amenities to rural communities, providing a granular perspective that is instrumental for our broader analysis of rural infrastructure accessibility. The resulting graph-based dataset, linking villages to their nearest POIs across defined categories, serves as the foundation for our enhanced geospatial descriptors for rural infrastructure analysis, offering actionable insights into the accessibility landscape at a national level.

3.2. Result of the algorithm application

The application of graph-based descriptors revolutionizes the analysis of rural infrastructure in Ukraine, offering precise insights into the accessibility of essential services for rural communities. These innovative descriptors facilitate a detailed understanding of connectivity between villages and critical Points of Interest (POIs), highlighting infrastructure gaps and underserved areas with unprecedented clarity. This approach not only supports data-driven decision-making for targeted development initiatives but also aids in prioritizing investments to address the most critical needs. By quantifying village accessibility to amenities like healthcare, education, and transportation, the graph-based analysis provides a foundational tool for both current and future rural development planning. It enables policymakers and development agencies to craft effective strategies, ensuring resources are allocated where they can have the most significant impact, thereby improving the overall quality of life in rural Ukraine.

Listing 1. JSON Example


```
[
  {
    "id_type": "admin4Pcod",
    "id": "UA2111000000",
    "distance": 3111.931012554032,
    "pos_x": 52060.12938924221,
    "pos_y": 5420926.907282681
  }
]
```

Table 2. Summary of Data and Parameters

Data	Description in the output graph	Maximum Distance	Maximum Quantity Inside
City	Id, pos x, pos y, distance	50km	5
Local parks	Id, type, area, pos x, pos y, distance	30km	5
National parks	Id, area, pos x, pos y, distance	50km	5
Regional parks	Id, area, pos x, pos y, distance	50km	5
Bank	Id, pos x, pos y, distance	30km	5
Church	Id, type, pos x, pos y, distance	30km	5
Education	Id, type, pos x, pos y, distance	30km	5
Elevator	Id, pos x, pos y, distance	30km	5
Hotel	Id, type, pos x, pos y, distance	30km	5
Kindergarten	Id, pos x, pos y, distance	30km	5
Library	Id, pos x, pos y, distance	30km	5
Medicine	Id, type, pos x, pos y, distance	30km	5
Shops	Id, type, pos x, pos y, distance	30km	5

3.3. Statistical Validation of Graph-Based Infrastructure Descriptors

Our analysis revealed that essential services traditionally cluster within specific geographic proximities, which our graph-based model adeptly captured. By examining the traditional distance metrics depicted in Figure 2 and juxtaposing them with the innovative graph-based descriptors in Figure 3, we were able to ascertain the model's fidelity in mapping real-world spatial relationships. The consistency of spatial correlations observed in the model confirms the preservation of inherent spatial dynamics, ensuring the descriptors' validity.

This methodological integrity is crucial, as it supports the reliability of our findings. The graph-based model provides a scaffold for development strategists and policymakers to visualize and prioritize improvements in rural infrastructure connectivity. Moreover, the alignment of our graph-based descriptors with real-world accessibility patterns underscores their potential as a tool for strategic planning. By retaining the detailed nuances of rural accessibility and service distribution, our model establishes a reliable foundation for informed interventions aimed at enhancing the quality of life for rural inhabitants.

The implications of these findings are far-reaching. With a validated model that accurately reflects the spatial distribution of services, stakeholders can make targeted decisions to bridge the gaps in rural infrastructure. Such strategic enhancements are expected to bolster the socio-economic fabric of these communities, facilitating better access to essential services and, consequently, fostering regional development. This study thus not only contributes a validated analytical tool but also provides a roadmap for applying such tools in the pursuit of equitable rural development.

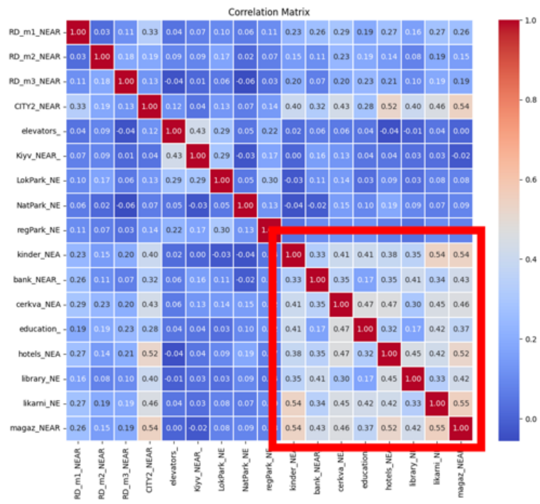

Figure 2. Correlation map of the traditional distance metrics

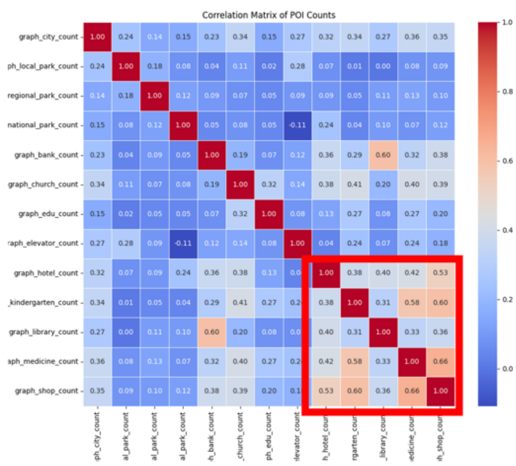

Figure 3. Correlation map of the graph-based infrastructure descriptors

4. Clustering Analysis

4.1. Custom vector creation

In this study, we introduce a pioneering approach titled «Custom Vector Creation» for the systematic analysis and enhancement of rural infrastructure through advanced clustering techniques and the integration of geospatial data. This methodology represents a significant advancement in the field of geoinformatics, particularly in the application of spatial data analysis for rural development. The crux of our approach lies in the fusion of proximity-based metrics and graph-based spatial representations to construct comprehensive infrastructure profiles for each village.

By synthesizing distance measurements to essential infrastructure components with intricate graph-based spatial details, we generate unique vectors that encapsulate the multifaceted characteristics of village infrastructure. This amalgamation enables a nuanced comparison and categorization of villages based on their infrastructure quality, offering a more refined perspective than traditional methods. The precision afforded by this technique allows for the identification of specific infrastructure needs and gaps, facilitating targeted interventions.

Our custom vector creation process employs a dual-source data model, leveraging the strengths of both proximity measures and graph-based information to capture an in-depth view of rural infrastructure landscapes. This hybrid approach not only enhances the granularity of our analysis but also improves the accuracy of the subsequent clustering process, thereby ensuring a more effective categorization of villages for development planning.

4.2. K-Means Clustering for Infrastructure Categorization

In the pursuit of delineating rural infrastructure disparities, our research introduces an advanced K-Means clustering methodology tailored for infrastructure categorization. This technique is pivotal in discerning and categorizing the infrastructural nuances within rural village landscapes. The clustering operation commences with the aggregation of multi-dimensional data on village infrastructure into unique vectors, each reflective of a specific point of interest (POI). These vectors are then subjected to the K-Means algorithm, resulting in the preliminary grouping of villages based on the similarity of their infrastructure attributes.

Upon completion of the initial clustering phase, a post-sorting step is implemented to reorder the clusters. This reorganization is based on the infrastructure quality, ranked from those with the highest needs to the least, utilizing the centroids derived from the K-Means process. This post-sorting step is instrumental in elevating the clarity of the clusters, facilitating a more strategic approach to development planning by prioritizing areas with critical infrastructure needs.

Figure 4 exemplifies this process, showcasing clusters of road types through a 3D visualization where each

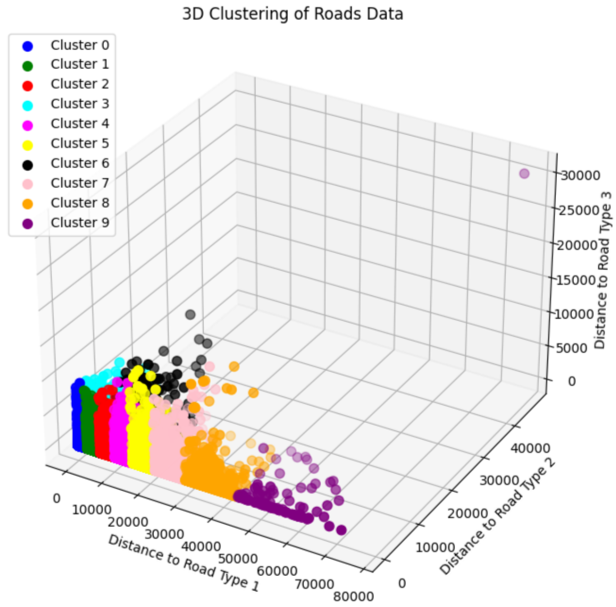

Figure 4. Road clustering visualization using improved algorithm

color represents a distinct cluster. It is discernible from the figure that the clusters are arranged in an order indicative of their proximity to crucial infrastructure, with Cluster 0 harboring villages closest to essential road types and Cluster 9 representing those furthest away. This visualization not only confirms the efficacy of our clustering method but also vividly highlights the geographic spread and infrastructure accessibility within the analyzed rural regions, thus providing a valuable tool for policymakers and developmental agencies in strategic planning and resource allocation.

4.3. Quality formula implementation and distribution analysis

In the domain of geospatial analysis for rural development, this research adopts a rigorous, data-driven approach to clustering techniques. The crux of our methodological innovation lies in the Custom Vector Creation and subsequent distribution analysis, aiming to systematically categorize villages based on infrastructure quality. The novel formula implemented for quality assessment is as follows:

$$QualityScore = 0.8 \times \left(\frac{1}{n} \sum_{i=1}^{n} category_i \right) + 0.2 \times clusterindex, \quad (1)$$

where:

- $category_i$ represents the infrastructure category being assessed;
- n is the number of infrastructure categories included;
- $clusterindex$ is the assigned number post-clustering, indicating the relative quality of infrastructure within the cluster.

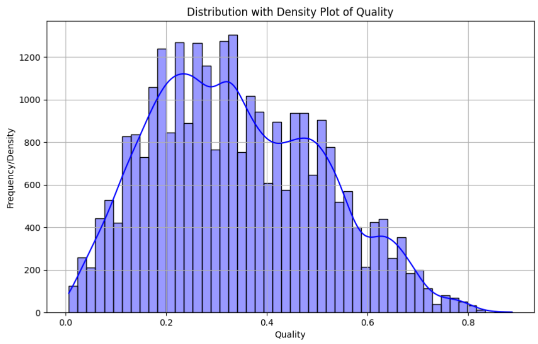

Figure 5. Road clustering visualization using improved algorithm

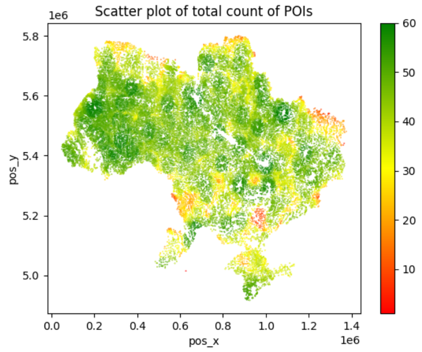

Figure 6. Road clustering visualization using improved algorithm

This formula (1) underscores the comprehensive assessment of infrastructure quality by accounting for both the individual infrastructure categories and the village's relative position within its cluster.

Figure 5 portrays the distribution of the overall infrastructure quality of villages, visualized through a histogram overlaid with a density plot. The quality scores, derived from the aforementioned formula, demonstrate a normal-like distribution, signifying a central tendency around the mean quality score. This distribution validates the clustering methodology and its effectiveness in the identification and grouping of villages with similar infrastructure profiles. Such graphical representation enables stakeholders to decipher the infrastructure landscape across rural regions quickly, serving as a catalyst for informed decision-making and resource allocation in development planning.

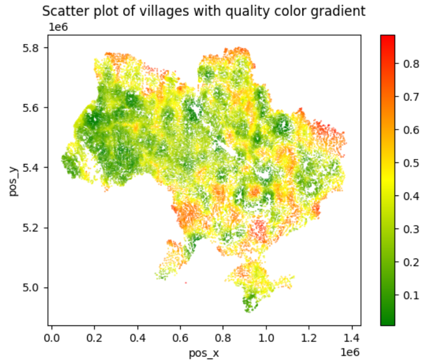

Figure 7. Road clustering visualization using improved algorithm

5. Discussion

This research presents a sophisticated model for analyzing and enhancing the infrastructure in Ukrainian villages through advanced clustering techniques and comprehensive geospatial analysis. The study identifies a distinct pattern in the distribution of rural infrastructure, which is critical for strategic development and planning.

The maps of Ukraine (Figure 6 and Figure 7) generated by our geospatial analysis illustrate a gradient of infrastructure quality. The western regions of Ukraine consistently display higher levels of development, while the southern and eastern regions reveal infrastructural deficits. These patterns highlight the potential disparities in resource allocation and accessibility, suggesting areas where development efforts could be maximally effective.

The study reaffirms the utility of clustering techniques in categorizing villages by infrastructure quality. This approach elucidates the disparities and facilitates the identification of specific areas that require urgent developmental strategies. The visualization of infrastructure quality across rural Ukraine provides an empirical basis for policymakers to prioritize interventions, aiming to foster equitable development throughout the nation.

Moreover, the research underscores the potential of integrating geospatial data with clustering to inform policy decisions. This methodology allows for a more nuanced understanding of infrastructural needs and the socio-economic factors that influence them, which is essential for crafting long-term sustainable development policies.

Our findings also suggest a broader application of the methodology, revealing underlying socio-economic and geographic determinants that could be addressed by future policy interventions. This perspective highlights the importance of going beyond infrastructure to consider the complex socio-political forces that shape rural development.

While the study provides comprehensive insights into the current state of rural infrastructure, it also recognizes the dynamic nature of development. Continuous updates and enhancements to geospatial data quality are necessary to ensure the accuracy and reliability of future analyses.

In summary, the findings of this research contribute to a detailed understanding of the infrastructural landscape in rural Ukraine. The methods and conclusions drawn from this study offer a blueprint for using geospatial analysis to facilitate informed decision-making and resource allocation, paving the way for targeted and effective rural development initiatives. The discussion sets forth the transformative potential of such analytical approaches in rural development policies.

Conclusions

This research has provided a comprehensive analysis of rural infrastructure in Ukraine by leveraging advanced geospatial data and clustering techniques. Through the meticulous preparation and validation of OpenStreetMap data and the novel application of clustering, the study has demonstrated the potential to categorize villages effectively by infrastructure quality, highlighting the developmental disparities across various regions.

The integration of graph-based spatial information has enabled a granular examination of the infrastructure landscape, offering actionable insights for policymakers and development agencies. By identifying specific areas with the most critical needs, the study facilitates the efficient allocation of resources and strategic intervention planning. Furthermore, the research methodology's robustness and precision in categorizing infrastructure quality serve as a benchmark for future rural development projects.

Looking ahead, the adaptability of this approach to different geographic contexts and the integration of additional socio-economic data can expand its applicability. This opens avenues for broader applications, including strategic rebuilding efforts in post-conflict scenarios and the potential to guide sustainable development initiatives worldwide.

The study's outcomes underscore the importance of a data-driven approach in rural revitalization efforts and set the stage for transformative development policies that can significantly enhance the quality of life for rural populations.

With the backing of Ukraine's Ministry of Education and Science by the project of «Information Technologies of Geospatial Analysis for the Development of Rural Areas and Communities», the project's findings are well-positioned to influence real-world infrastructure development programs and contribute to a more balanced and sustainable recovery process.

In conclusion, this research has not only contributed a novel framework to the field of rural infrastructure analysis but also established a foundation for future innovations that can advance the development of rural communities both in Ukraine and globally.

References

1. A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap / B. Herfort, S. Lautenbach, J. Porto de Albuquerque, J. Anderson, A. Zipf // Nature Communications. — 2023. — Vol. 14. — P. 3985. — DOI: 10.1038/s41467-023-33956-z.
2. Elevators in Ukraine. — 2022. — Accessed: 2022. https://elevatorist.com/karta-elevatorovukrainy.
3. Geospatial Analysis of Life Quality in Ukrainian Rural Areas / H. Yailymova, B. Yailymov, N. Kussul, A. Shelestov // Proceedings of the 13th International Conference on Dependable Systems, Services and Technologies (DESSERT). — Athens, Greece, 2023. — P. 1–5. — DOI: 10.1109/DESSERT61349.2023.10416517. — URL: https://doi.org/10.1109/DESSERT61349.2023.10416517.
4. Geospatial characterization of rural settlements and potential targets for revitalization by geoinformation technology / Y. Liu [et al.] // Sci. Rep. — 2022. — Vol. 12. — P. 8399. — DOI: 10.1038/s41598-022-12294-2. — URL: https://doi.org/10.1038/s41598-022-12294-2.
5. OpenStreetMap: Ukraine. — 2021. — Accessed: 2021. https://download.geofabrik.de/europe/ukraine.html.
6. *Roy S. M.*, *Prasad N. S. R.*, *Srinuvas B.* Achieving Sustainable Village Development Through Geoinformation Application // ICT Analysis and Applications. Vol. 314. — 2022. — P. 833–843. — (Lecture Notes in Networks and Systems). — DOI: 10.1007/978-981-16-3331-1_70. — URL: https://doi.org/10.1007/978-981-16-3331-1_70.
7. The Humanitarian Data Exchange. — 2021. — Accessed: 2021. https://data.humdata.org/.
8. *Ye X.*, *Wei Y. D.* Geospatial Analysis of Regional Development in China: The Case of Zhejiang Province and the Wenzhou Model // The Professional Geographer. — 2013. — Vol. 46, no. 6. — P. 445–464. — DOI: 10.2747/1538-7216.46.6.445.