Integrated Geospatial Analysis for Rural Development Metrics

Nataliia Kussul[1,2], Vlada Svirsh[1,3] and Bohdan Potuzhnyi[1,3]

[1] Institute of Physics and Technology, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiskyi Ave, 37, Kyiv, 03056, Ukraine
[2] Space Research Institute NASU-SSAU, Ukraine
[3] Engineering and Informational Technology Department, Bern University of Applied Science, Quellgasse, 21, Biel/Bienne, 2502, Switzerland

Abstract

This paper presents an innovative geospatial analysis framework to advance rural development policymaking in Ukraine. The core objectives are constructing enhanced infrastructure accessibility descriptors for villages, ensuring statistical integrity compared to preliminary data, and establishing detailed linkage maps to locate gaps. OpenStreetMap resources are integrated with humanitarian datasets to analyze over 10,000 rural settlements. Specialized algorithms transform distance metrics into graphed connectivity between villages and surrounding healthcare, education, transit, and other point-of-interest amenities proximal to them. Rigorous statistical testing proves consistency between initial rural accessibility distributions versus the graph-enhanced representations. Histograms, boxplots, and correlation analysis verify retained descriptive integrity. The outputs uniquely quantify village-tier granular infrastructure linkages to inform targeted revitalization. Gradient visualizations locate Ukraine's severest underserviced rural areas based on healthcare, schooling, poverty, and war displacement factors. This highlights the dire need and opportunity for strategic connectivity improvements centered on village accessibility requirements. Transitioning to rural graph neural networks assimilating real-time data streams promises responsive development policy recalibration as living conditions evolve.

Keywords

Geospatial Infrastructure Analysis, Statistical Modeling, Rural Development Metrics, Open Data Utilization

1. Introduction

Effective rural development policy hinges on addressing critical questions concerning the infrastructure gaps and accessibility barriers that villages encounter. This research is dedicated to informing such evidence-based decision-making in Ukraine through an innovative geospatial analysis framework focused on rural settlements.

Our core research questions include:

1. How can we construct enhanced descriptors of village accessibility to critical infrastructures like healthcare, transit, and public services based on real-world proximity linkages?
2. Can graph-based representations of rural infrastructure connectivity retain the statistical properties of original accessibility distributions while providing greater insights?
3. Which localized infrastructure linkage maps for each village can identify developmental gaps and priorities?

To tackle these questions, we integrate diverse geospatial data layers, drawing predominantly from collaborative OpenStreetMap resources. Our methodologies encompass specialized algorithms for rural data transformations as well as robust statistical testing. The outputs go

COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13, 2024, Lviv, Ukraine
nataliia.kussul@lll.kpi.ua (N. Kussul); vlada.svirsh25@gmail.com (V. Svirsh); bohdan.potuzhnyi@gmail.com (B. Potuzhnyi)
0000-0002-9704-9702 (N. Kussul); 0009-0004-4006-2705 (V. Svirsh); 0009-0001-2282-7012 (B. Potuzhnyi)

beyond previous efforts that were limited to macro-scale overviews or sparse point accessibility estimates.

Instead, our infrastructure graphs connect tens of thousands of Ukrainian rural localities to surrounding amenities such as medical, educational, and commercial points of interest proximal to them. The result is a detailed perspective on how village accessibility aligns with development needs, made actionable through the identification of localized gaps and opportunities.

With broader project support from the Ministry of Education and Science of Ukraine, these analytics will directly inform rural revitalization investments by providing insights into technology infrastructure integration and quantifying access inequality. Our approach and the novel descriptiveness it brings to rural areas also hold the promise of transferability to similar development contexts worldwide.

2. Related works

The application of geospatial analysis in rural development has seen significant evolution over the past few years, driven by advancements in data collection, processing techniques, and the growing availability of open-source data. This section reviews recent contributions that have shaped our current understanding and methodologies, specifically focusing on works that utilize geospatial analysis, open data sources like OpenStreetMap (OSM), and innovative analytical frameworks to address rural infrastructure development.

2.1. Geospatial characterization of rural settlements

In this section, we explore the evolution of geospatial analysis in rural development, highlighting significant advancements in the field as exemplified by the study "Geospatial Characterization of Rural Settlements and Potential Targets for Revitalization by Geoinformation Technology" by Yixuan Liu and colleagues [1]. Their research stands as a pivotal contribution, employing advanced spatial analysis techniques, such as kernel density, spatial autocorrelation, and regression analyses, to dissect the rural fabric of Jiangxi Province, China. Integrating remote sensing, topographic, and socioeconomic data, Liu et al. reveal a distinctive spatial distribution pattern of rural settlements, characterized by denser regions in the north and sparser areas in the south, shaped by physical and socioeconomic drivers.

Crucially, their work introduces the Socio-Environmental Evaluation Index (SEI), a novel metric for assessing rural development inequality and guiding targeted revitalization efforts. This approach not only enriches our understanding of rural settlement dynamics but also proposes a methodological framework for identifying revitalization priorities based on a comprehensive evaluation of socio-environmental factors. The study’s insights into the "dense north and sparse south" distribution and the development of the SEI represent a methodological leap in rural geospatial analysis, offering a nuanced perspective on rural development challenges and opportunities.

Liu et al.'s research aligns with broader trends in geospatial analysis, emphasizing the critical role of integrating environmental and socioeconomic data to inform rural development strategies. Their findings resonate with contemporary studies that examine the spatial heterogeneity of urban-rural integration, the conceptual expansion of city studies, and socio-spatial inequalities within various geographic contexts. By situating their work within this evolving landscape, Liu et al. contribute to a more informed and nuanced understanding of rural settlement patterns, underscoring the importance of geospatial analysis in crafting targeted and effective rural revitalization policies.

The advancements in geospatial technology and analytical methods showcased in Liu et al.'s study, along with related works, mark a significant step forward in rural development research. These contributions not only enhance our spatial understanding of rural areas but also offer practical tools for addressing the complex challenges of rural revitalization, emphasizing the value of geospatial analysis in navigating the intricate socio-environmental systems that define rural landscapes.

2.2. Spatio-temporal analysis of global urban building data in OSM

Building on the geospatial characterization of rural settlements discussed in Section 2.1, this section delves into the analysis of urban building data completeness within the OpenStreetMap (OSM) framework. It presents a spatio-temporal investigation to evaluate the extent and distribution of urban building footprints globally, examining disparities in data availability and quality that may impact comprehensive urban analysis and policymaking.

The study conducted by Herfort et al. (2023) [2] employs a machine-learning model to assess the completeness of the OSM building stock across 13,189 urban agglomerations. The findings reveal that while OSM's building footprint data exhibits over 80% completeness for a subset of urban centers, representing 16% of the urban population, the majority of cities — encompassing 48% of the urban populace — exhibit less than 20% completeness. This discrepancy highlights the importance of addressing data inequalities within the OSM platform to ensure unbiased insights into urban development.

Assessing OSM data inequalities is crucial as it directly affects the use of geospatial information in urban planning and the achievement of Sustainable Development Goals. The authors introduce a comprehensive framework for evaluating the completeness of OSM building data, considering factors such as the Human Development Index, population size, and geographic location to elucidate complex patterns of spatial bias in data coverage.

This section contributes to the overarching goal of the paper by emphasizing the necessity of integrating diverse data sources and analytical methods for a holistic understanding of both rural and urban infrastructure development. The insights from this analysis not only augment the rural focus of the earlier sections but also broaden the perspective on the applicability of open-source geospatial data for infrastructure analysis across multiple scales.

Herfort et al.'s investigation into urban OSM building data provides a critical reflection on the current state of geospatial data completeness, advocating for a more equitable distribution of data collection efforts. This approach is in line with the innovative framework presented in this paper, which underscores statistical accuracy, impartiality, and data diversity in geospatial analysis for rural development metrics.

2.3. Geospatial analysis of life quality in Ukrainian rural areas

Expanding upon the theme of rural development through geospatial analysis, Yailymova et al.'s work [3] introduces an algorithm to assess the quality of life in Ukraine's rural areas. Their methodology incorporates a comprehensive assessment of village remoteness from essential infrastructure and natural ecosystems, while also considering proximity to conflict zones. This innovative approach addresses not only the physical but also the socio-political landscape, affecting rural life quality.

Yailymova et al.'s study indicates a significant disparity in life quality, with many villages, particularly in eastern and southern Ukraine, facing challenges exacerbated by ongoing conflict. The study’s algorithmic assessment aligns with efforts to direct revitalization efforts where they are needed most, offering a data-driven foundation for policy decisions.

In concert with the geospatial evaluations presented in previous sections, this research further underscores the dichotomy between rural and urban infrastructural development. It also highlights the acute challenges faced in war-torn regions, presenting a pressing case for targeted infrastructural and social intervention. Yailymova et al.'s contribution is thus a poignant reminder of the complex interplay between geography, infrastructure, and socio-political factors in shaping rural livelihoods.

2.4. Utilization of geospatial technology for village-level socio-infrastructural mapping

Maryada and Thatiparthi [4] present a case study in Chinnapendyala village, illustrating how geospatial technology can effectively map the social and infrastructural facilities at the micro-level. By combining spatial and non-spatial data, their study creates a detailed geodatabase, serving as a crucial tool for planners and policymakers in understanding and addressing the needs at the grassroots level.

Their methodology demonstrates the potential of GIS in visualizing and managing village-level development plans, emphasizing the integration of various parameters such as amenities, income, and social indicators. This approach aligns with the broader objective of sustainable and equitable rural development, showcasing the practical application of geospatial technology in enhancing the living conditions in rural areas.

2.5. Synthesis and future directions

Collectively, these studies highlight the transformative power of geospatial analysis in rural development. They present a spectrum of methodologies, from regional assessments to village-specific analyses, each contributing unique insights into the multifaceted nature of rural life and its enhancement through targeted development strategies.

Looking forward, the integration of additional data layers — reflecting agricultural activities, demographic changes, and environmental conditions — can enrich these analyses. Furthermore, the incorporation of real-time data and machine learning algorithms could provide even more nuanced, predictive insights into rural development needs and outcomes.

As this body of work continues to grow, the fusion of geospatial technology with other emerging data sciences holds the promise of driving informed, sustainable, and inclusive rural development policies across diverse global contexts.

3. Methods and materials

3.1. Materials

For this analysis, various geospatial data layers were extracted from OpenStreetMap (OSM) into a GeoDataFrame (GDF).

OpenStreetMap (OSM) is a collaborative project to create an editable map of the world, built by volunteers using aerial imagery, GPS devices, and low-tech field maps [5]. OSM data is open-source and includes comprehensive global coverage of roads, buildings, natural features, and provides a rich data foundation for geospatial analysis. In our study data layers extracted from OSM include major, secondary, and rural roads; land cover classification; and locations of schools, colleges, universities, hotels, hospitals, clinics, pharmacies, supermarkets, malls, banks, churches, libraries, kindergartens, and local, national and regional parks.

Additional data on village and city locations came from the Humanitarian Data Exchange (HDX), an open platform for sharing data across crises and countries [6]. The HDX is run by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA). For our analysis, HDX data on settlement locations in Ukraine as of mid-2021 was utilized.

Full list of the geospatial data across Ukraine used for our analysis is presented in Table 1.

Table 1
Geospatial data

Data	Layers	Source
Villages	Villages	The Humanitarian Data Exchange (as of 17.07.2021)[6]
Cities	City	The Humanitarian Data Exchange (as of 17.07.2021) [6]
Elevators	Elevators	Elevators in Ukraine (as of 23.02.2022) [7]
Roads	Major roads, Secondary roads, Rural roads	OSM [5]
Education	School, College, University	OSM[5]
Hotels	Hotel, guesthouse, shelter	OSM [5]
Medicine	Hospital, clinic, pharmacy	OSM [5]
Shops	Supermarket, mall, clothes, marketplace	OSM [5]
Bank	Bank	OSM [5]
Church	Church	OSM [5]
Library	Library	OSM [5]
Kindergarten	Kindergarten	OSM [5]
Parks	Local Park, National Park, Regional Park	OSM [5]

3.2. Methods

3.2.1. Methods for creation of extensive dataset with new descriptors

To develop a more comprehensive dataset with additional descriptors for rural areas in Ukraine, we implement a systematic process for identifying and integrating points of interest (POIs) from source geospatial data layers into groups for each village, presented in graph format. For each type of POI, we first segment all objects into buffer boxes of the same predefined size, corresponding to the maximum distance defined for each type in Table 2. Upon creating these square buffers, we proceed to identify the closest objects for each village. This involves determining which buffer a village is located in and defining the set of eight neighboring boxes. For each village, we limit the number of objects to a predefined number of closest objects, as specified in Table 2 under the column "maximum quantity inside." The process of POI buffering and defining the lookup buffers is illustrated in Figure 1, while the process of calculating distances to POIs and identifying the closest ones is described in Figure 2.

This approach enables the creation of an encompassing set of POIs in proximity to each village, capturing both the breadth of nearby features (including cities, parks, etc.) and ensuring that they are located within an accessible distance. By systematically cataloging the closest POIs of diverse types to each rural settlement, we develop extensive descriptors capturing the accessibility and availability of key infrastructure and services. The end result is a graph-structured dataset linking villages to their nearest POIs across defined categories, along with distance metrics. This output provides granular insights into rural access to vital amenities at a national level. The graph-based structure, which connects rural villages to proximal POIs of different types through distance-based linkages, forms the foundation of our enhanced geospatial descriptors for rural infrastructure analysis.

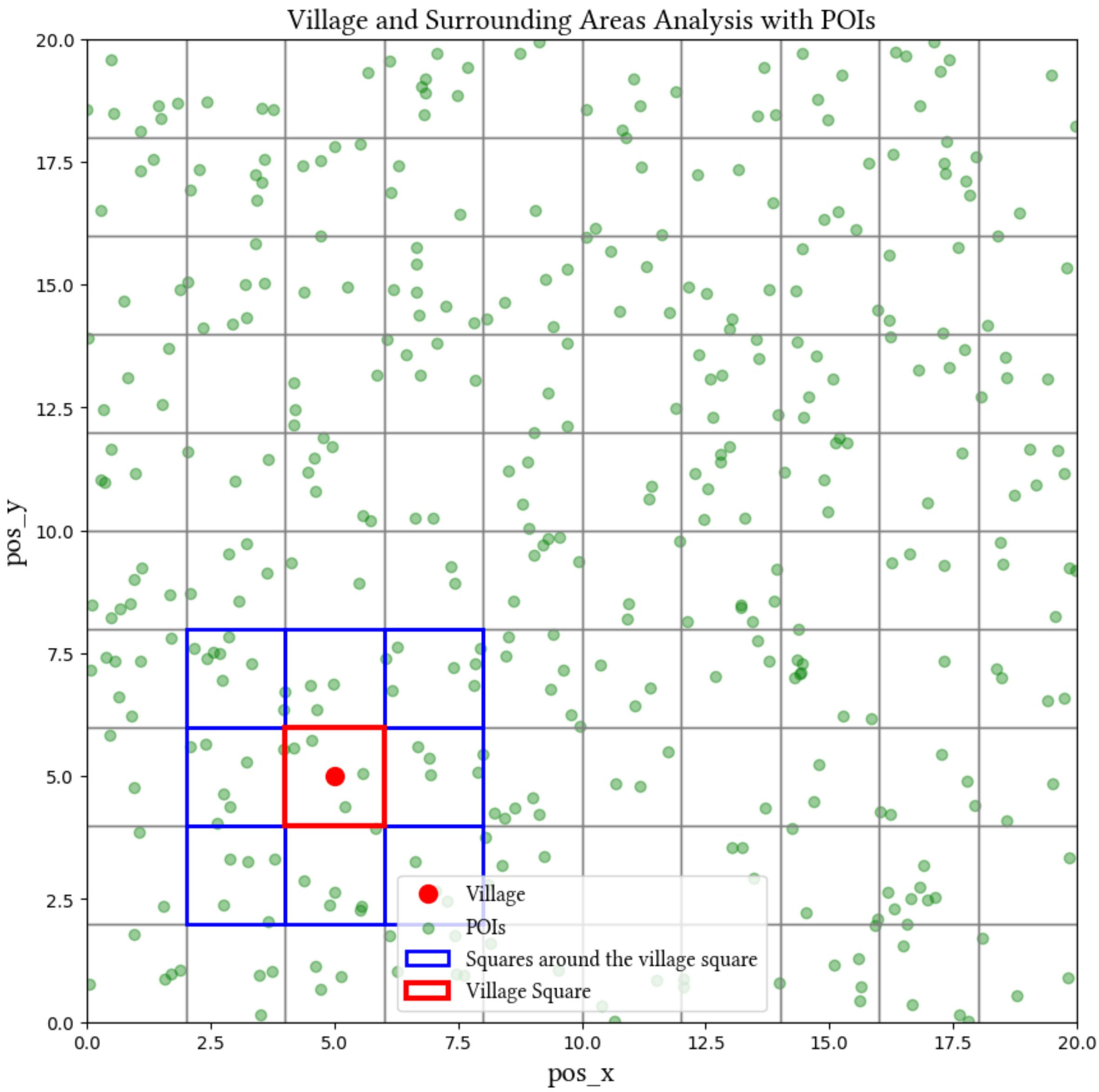

Figure 1: An illustration of the methodology used to define the regions for spatial analysis. The grey box represents all the created buffers. The blue boxes are those neighboring the village buffer, and the red box indicates the buffer in which the village is located. Green dots represent Points of Interest (POIs), and the red dots denote the village center.

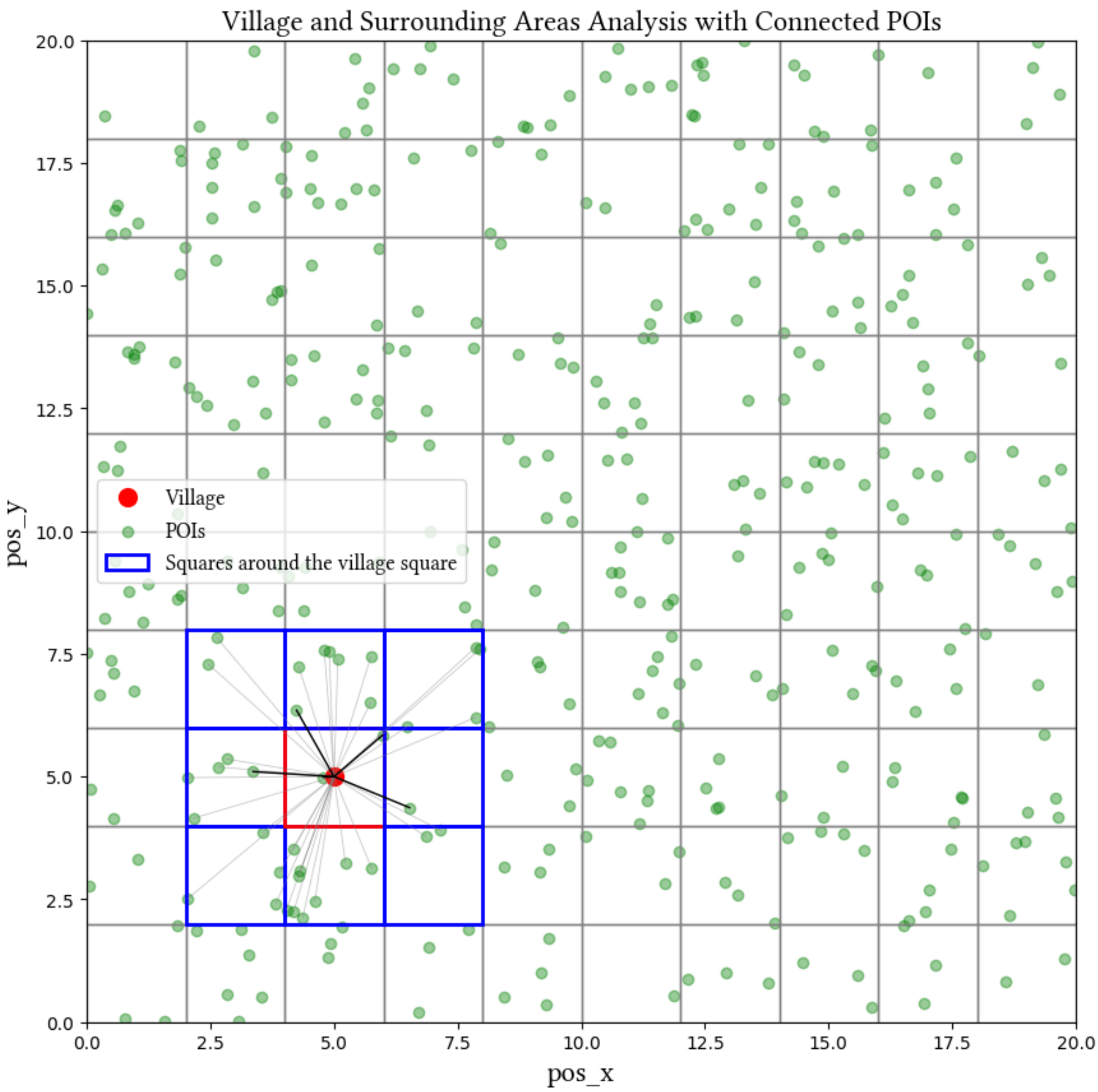

Figure 2: This figure demonstrates the calculation of distances within the designated regions, consistent with the methodology outlined in Figure 1. It includes the addition of grey lines, which represent the distances from the village center to all Points of Interest (POIs) within the observed buffers. The black lines specifically denote the distances to the five closest POIs. Red dots indicate the location of the village, while green dots mark the POIs.

3.2.2. Statistical analysis methodology

To support the thorough evaluation of our geospatial datasets related to rural infrastructure in Ukraine, from both the initial data and the newly constructed graph descriptors, we employ a range of numerical and graphical statistical techniques. Rather than relying on any one approach, combining multiple methods provides a rigorous, well-rounded perspective on the data distributions.

We extensively utilize histograms in our analytical workflow. Visualizing the frequency of infrastructure distance metrics across villages in histograms rapidly communicates the shape of their distributions. We can promptly identify normally distributed features versus those skewed in accessibility towards higher or lower range values. Outliers also emerge clearly in histograms. An example is a rural settlement situated anomalously far from the nearest medical clinic compared to most other villages.

Complementing the histograms' graphical distributional insights, box plots concisely represent the internal data spread unpinning those distributions — key quartiles, outliers, and extreme value range. Comparing box plots side-by-side enables quickly discerning median similarities and differences across infrastructure categories. For instance, we can discover which part of Ukrainian villages is within 10 km of a school and what is the median distance to a major city.

While the graphical approaches indicate distributional properties, correlation matrices directly spotlight the relationships between accessibility of hospitals, schools, roads, and other facilities. The correlation coefficients surface the dimensions of rural infrastructure with the tightest links, guiding deeper investigations into these aligned accessibility gaps.

In addition to learning from data visualizations, we leverage two fundamental statistical functions. Summary statistics like means, medians and standard deviations underscore central tendencies and variation for infrastructure distances and additional features. We utilize Python's Pandas library which offers convenient built-in descriptive statistic calculation functions through syntax like dataframe.describe(). More crucially, with our revamped graph dataset connecting villages to their nearest-accessibility POIs across categories, we construct distribution plots showing the number of POIs available within set distance radii of each rural settlement. By comparing the aggregate distributions to the preliminary data, we affirm retention of inherent geospatial relationships in the enhanced descriptive dataset.

This multi-pronged methodology combining histograms, box plots, correlation matrices, statistical summaries and distribution analysis supports robust evaluation of the intricacies in the geospatial rural accessibility datasets guiding infrastructure improvement initiatives for Ukrainian villages. The techniques provide cross-validating numerical and visual evidence.

4. Experiment

The key goal of our experimental effort is the development of an enhanced set of geospatial descriptors to represent rural accessibility to critical infrastructure based on a graph database model. To achieve this, we conduct a multi-phase analysis encompassing:

1. Statistical profiling of the baseline geospatial datasets on rural Ukraine to validate integrity and alignment with development priorities.
2. Implementation of specialized algorithms to construct graph-based linkages between rural settlements and surrounding multi-category points-of-interest (POIs) representing accessible infrastructure.
3. Quantitative and graphical statistical analysis assessing whether the graph-based enhancement retains the qualitative properties of the preliminary rural accessibility distributions.

The output is an advanced graph dataset quantifying village-level access to key amenities within critical distance thresholds. This powers upstream analytics to precisely locate gaps and bottlenecks in rural infrastructure limiting development.

4.1. Verification of baseline geospatial data

Initially, we verify the completeness and validity of the assembled geospatial data layers on rural infrastructure in Ukraine. We are working with the GeoDataFrames (GDFs), the points of which can be seen in Figure 3.

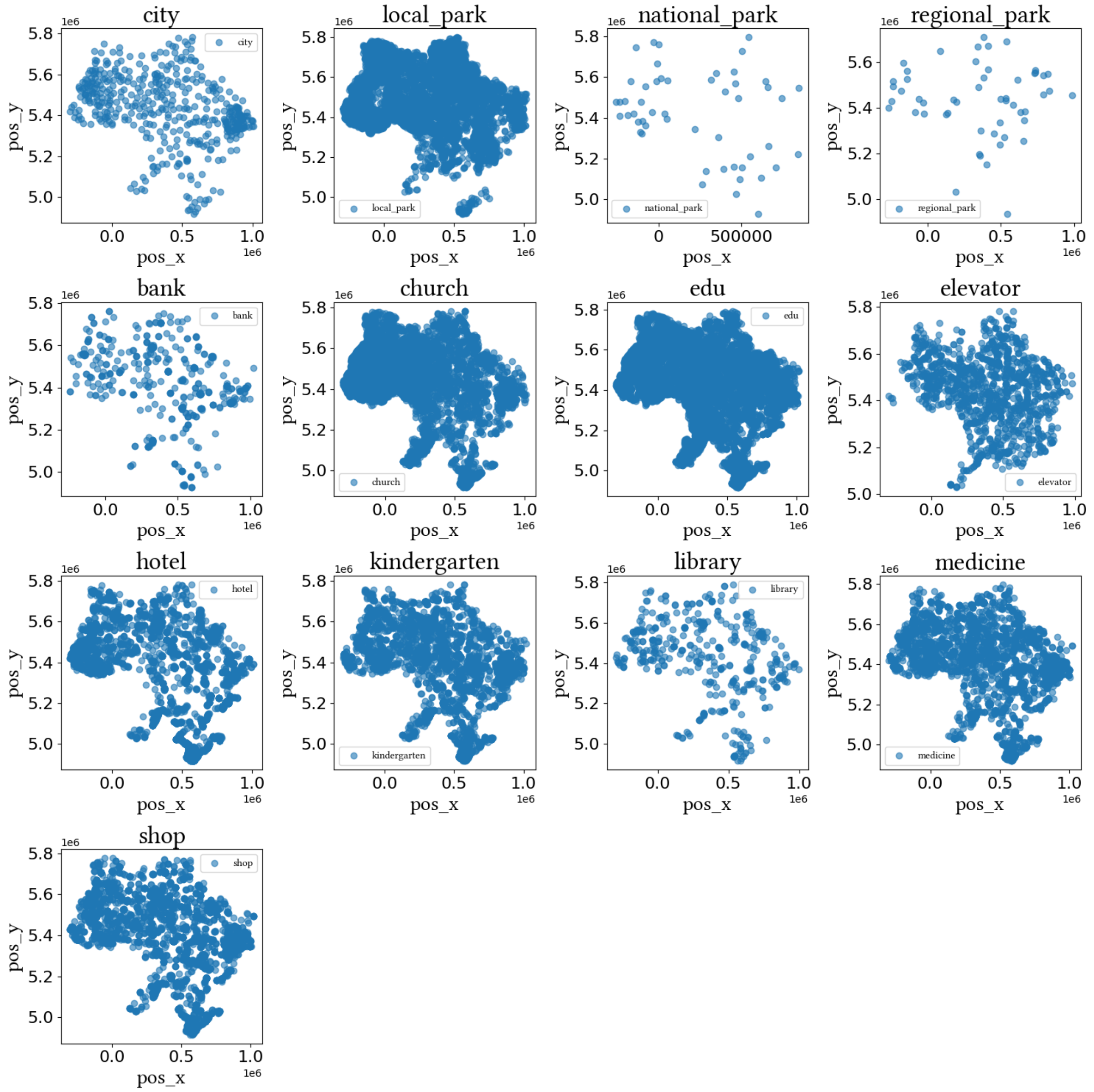

Figure 3: Graphical representation of the Points of Interest (POIs) that will be used for each of the GeoDataFrames (GDFs) analyzed. Each subplot depicts one type of POI, as categorized in Table 2. The blue dots represent individual POIs. The positions (pos_x and pos_y) are derived from longitude and latitude values, converted to meters.

Upon examination, it is evident that most of these plots exhibit a more or less normal distribution across Ukraine. "Normal" in this context implies that there are no significant gaps, stripes, or any other missing areas within the data. After analyzing this dataset and verifying its accuracy, we are proceeding to the subsequent steps.

4.2. Creation of rural accessibility graph descriptor

With clean baseline data secured, we implement our proposed methodology of rural catchment area segmentation, POI extraction and selection, distance calculation, and graph database construction to transform the preliminary datasets into enhanced accessibility descriptors.

Our experiment primarily involves the use of a table that describes which data was included in the newly created descriptors. For each GDF, we have identified specific values that determine which columns must be included in the output. These details can be found in Table 2.

Table 2
Creation scheme

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

The outcomes of this segment of the experiment will be discussed in the section Results.

4.3. Quantitative assessment of graph-based transformation

Finally, we employ a suite of statistical analysis techniques comparing key attributes between the baseline datasets and the new graph-based accessibility representation to ensure fundamental distributions remain consistent. We evaluate two types of data. The first type was pre-calculated by project members and encompasses various attributes, with a detailed description provided in Table 3.

Table 3
Distances to closest objects

Name	Description
RD_m1_NEAR	Distance to closest major road
RD_m2_NEAR	Distance to closest regional road
RD_m3_NEAR	Distance to closest rural road
CITY2_NEAR	Distance to closest city
elevators_	Distance to closest elevator
Kiyv_NEAR_	Distance to the capital of Ukraine, Kyiv
LokPark_NE	Distance to the closest local park
regPark_NE	Distance to the closest regional park
kinder_NEA	Distance to the closest kindergarten
bank_NEAR_	Distance to the closest bank
cerkva_NEA	Distance to the closest church
education_	Distance to the closest object of type education
hotels_NEA	Distance to the closest object of type hotel
library_NE	Distance to the closest library
likarni_NE	Distance to the closest hospital
magaz_NEAR	Distance to the closest shopping store

As discussed in Section 3.2.2, we have applied a comprehensive set of statistical tools to assess how the datasets behave, their correlations, and their distributions. Pertaining to the newly created data structure mentioned in Section 4.1, we have generated several graphs, including graph_city, graph_local_park, graph_regional_park, graph_bank, graph_church, graph_edu (educational institutions), graph_elevator, graph_hotel, graph_kindergarten, graph_library, graph_medicine (medical facilities), and graph_shop. Within the scope of this paper, we will focus exclusively on analyzing the statistics related to the 'distance' attribute for each of these entities.

The multi-stage experiment applies specialized algorithms and cloud-based geospatial analysis at scale to advance rural accessibility metrics from basic distance estimates to detailed infrastructure linkage descriptors mapped to individual villages.

5. Results

The key objectives of our rural infrastructure analysis are:

1. To develop enhanced descriptors of accessibility to critical amenities for each village based on a graph model linking settlements to real-world infrastructure and services distributed in their vicinity.

2. To ensure quantitative integrity such that the new descriptors retain the fundamental properties and distributions of the preliminary accessibility metrics.
3. To establish detailed infrastructure linkage maps for each rural community to inform development planning.

Our multi-faceted experiment achieves these goals through the key outputs described in the following subsections.

5.1. Graph-based accessibility descriptors

Transformation of the initial distance estimates into detailed graph connectivity between rural villages and surrounding multi-category POIs succeeds in constructing advanced accessibility descriptors. We have developed a structure as follows:


```
[
 {
  "id_type": "admin4Pcod",
  "id": "UA2111000000",
  "distance": 3111.931012554032,
  "pos_x": 52060.12938924221,
  "pos_y": 5420926.907282681
 },
 {
  "id_type": "admin4Pcod",
  "id": "UA2110100000",
  "distance": 23160.7396750182,
  "pos_x": 60818.384628206666,
  "pos_y": 5441084.188000026
 },
 {
  "id_type": "admin4Pcod",
  "id": "UA2123210100",
  "distance": 41187.70115749954,
  "pos_x": 76317.64176459657,
  "pos_y": 5451978.68918336
 },
 {
  "id_type": "admin4Pcod",
  "id": "UA2110400000",
  "distance": 41376.3746009097,
  "pos_x": 90058.87699085698,
  "pos_y": 5416411.645983889
 },
 {
  "id_type": "admin4Pcod",
  "id": "UA2110200000",
  "distance": 43577.79427222703,
  "pos_x": 80594.92931475205,
  "pos_y": 5391219.696301765
 }
]
```


When additional descriptors are needed, such as for regional parks, new key values are introduced into our graph structure. This enhancement ensures our graph's comprehensiveness, as outlined in Table 2.

5.2. Statistical analysis

One of the initial findings from this statistical analysis is illustrated in Figure 4. Examining RD_m2_NEAR and RD_m3_NEAR, it is apparent that nearly all villages have satisfactory access to roads of any type. However, there is significant room for improvement in RD_m1_NEAR, which exhibits a right-skewed distribution. Kyiv_NEAR shows a triangular distribution, indicating that the majority of villages are situated 150-500 km away, which is favorable as it suggests uniform access to the capital and, consequently, to business opportunities with companies based there. The distributions of other variables exemplify log-normal distribution, indicating that while some villages are in close proximity to our POIs, many others have room for improvement in terms of location and access to various facilities.

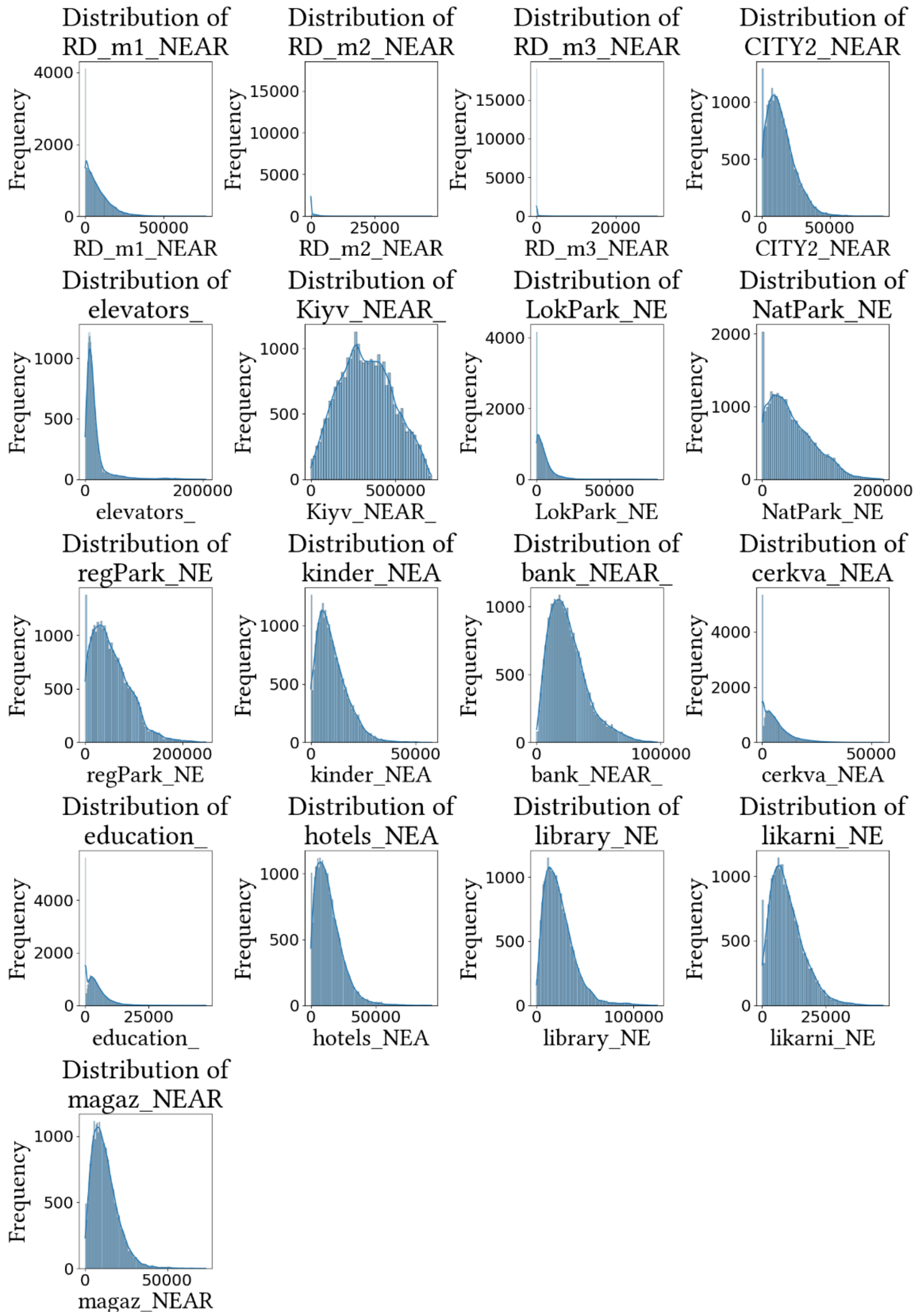

Figure 4: Graphical representation of the POIs for each of the received distance descriptors.

Further analysis, as shown in Figure 5, confirms our predictions. The box plots clearly depict distances and their interrelations.

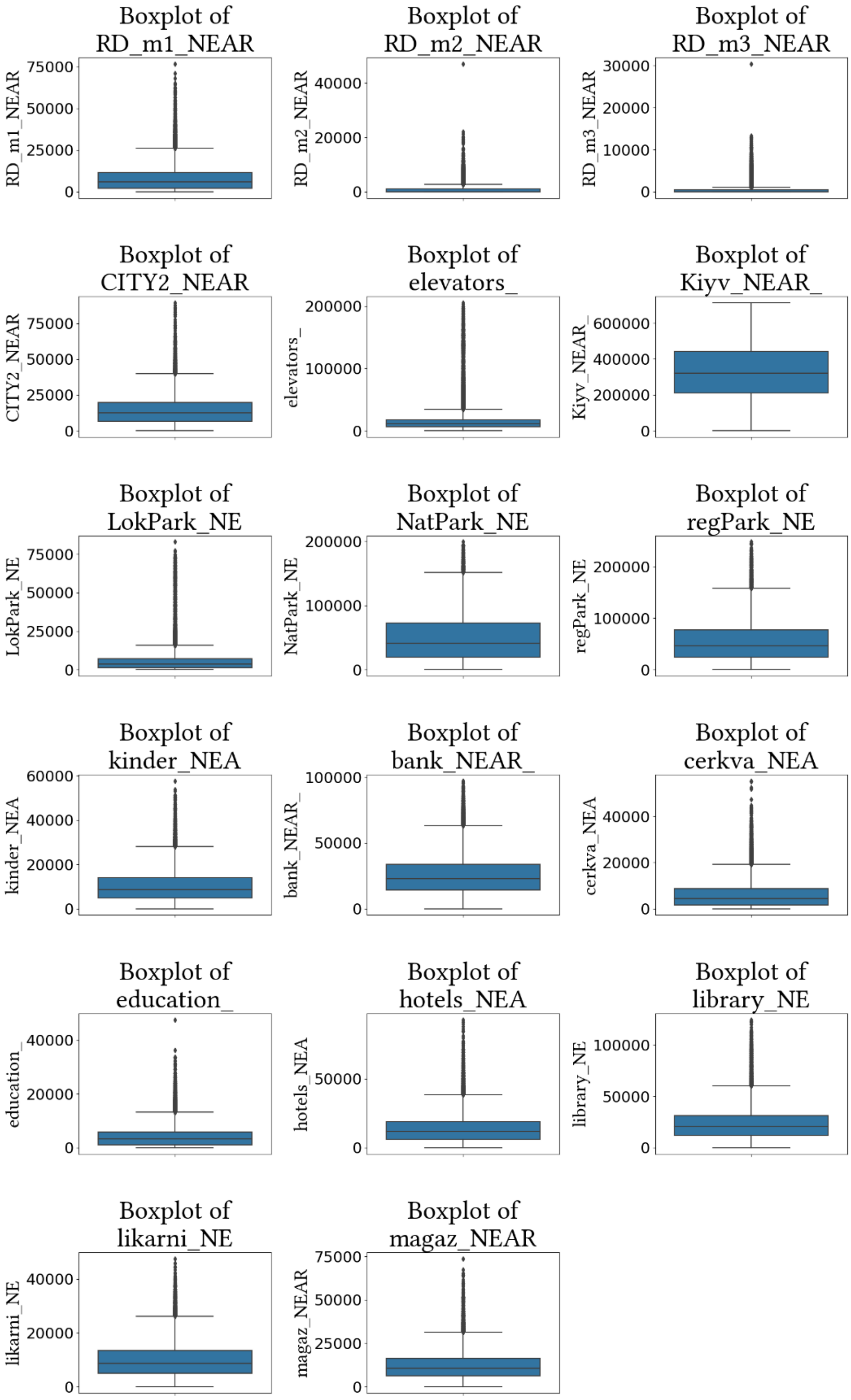

Figure 5: Boxplots of the researched columns.

An important observation is the correlation of facilities such as kindergartens, banks, churches, and educational services to the proximity of the nearest city, which logically follows since cities typically offer more and better facilities than villages.

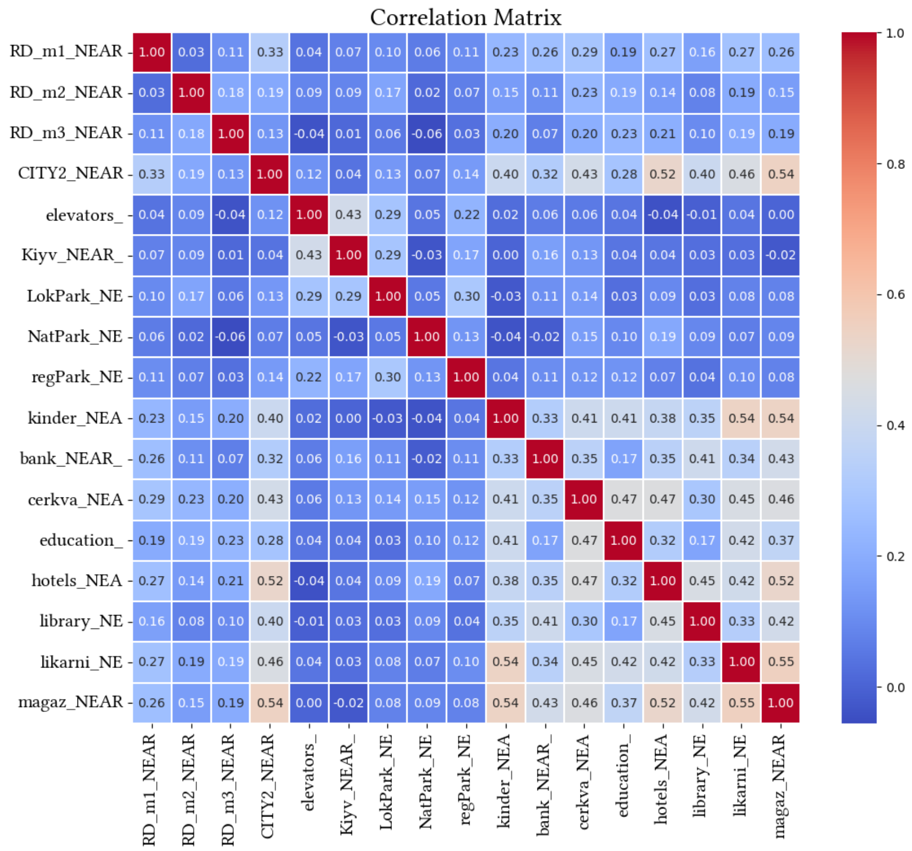

Figure 6: Correlation matrix of the columns reviewed earlier.

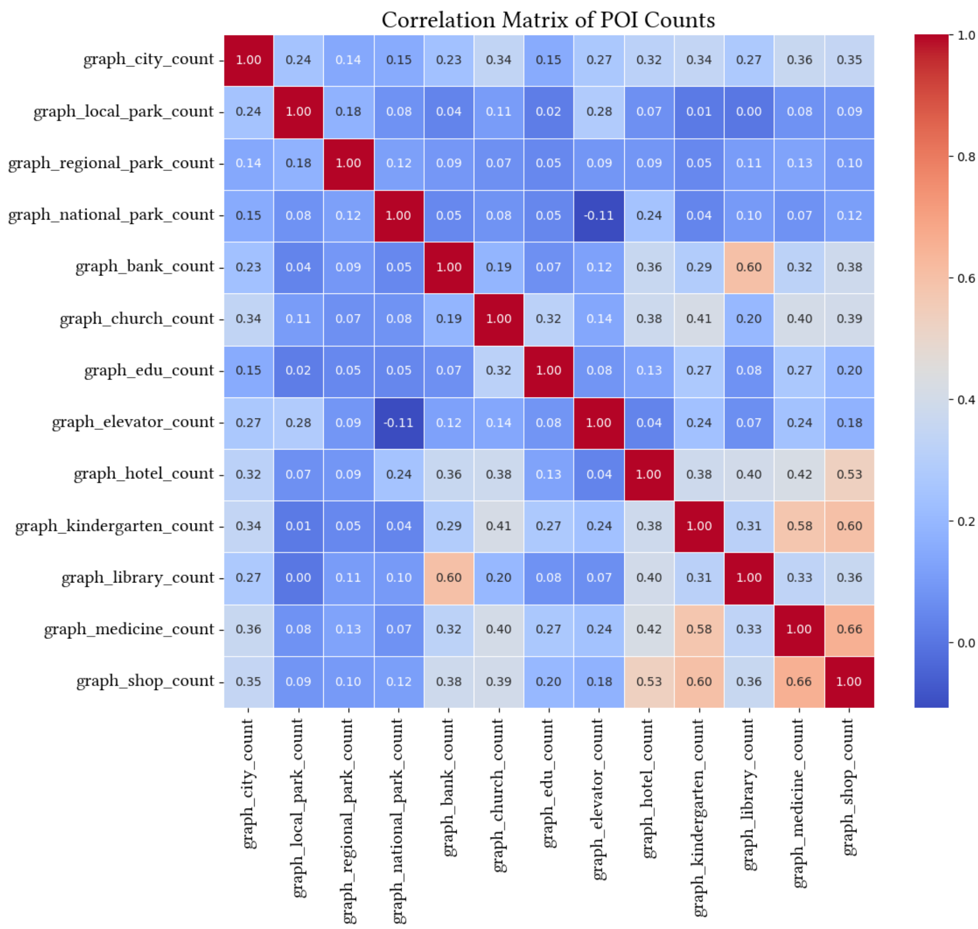

Figure 7: Correlation matrix of the newly created columns.

This brings us to our initial statistical description of the data in this paper. It is crucial that our data does not lose integrity or exhibit widely varying distributions. Figure 7 confirms the continued relevance of social facilities' accessibility. Next, we will verify that the data distributions remain consistent or nearly so by constructing distribution plots, as shown in Figure 8. Although these plots are extensive, we observe that the overall distribution has not significantly altered in our new data description method.

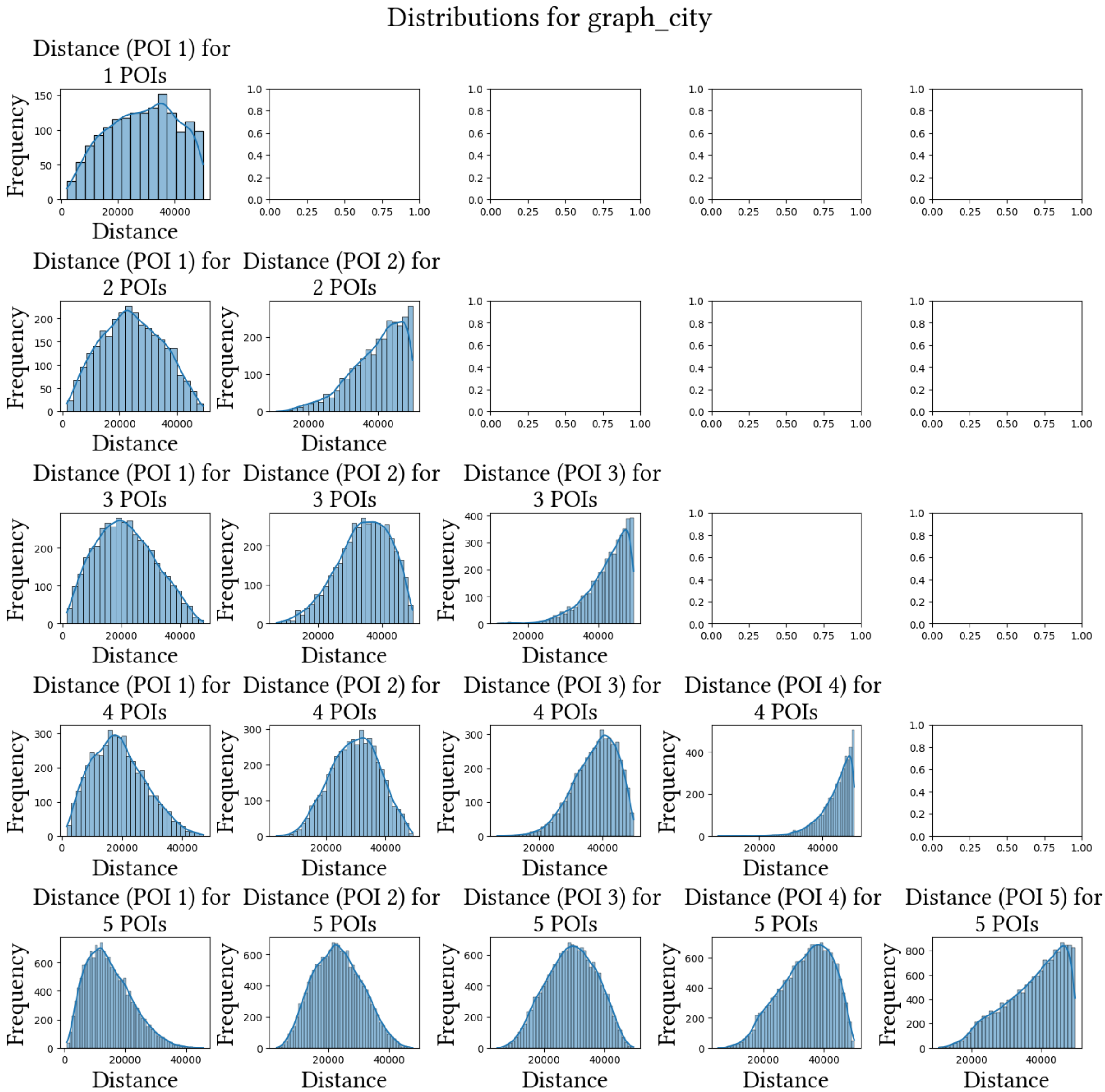

Figure 8: Distribution matrix objects in "graph city" based on the number of POIs within a radius, for each specific order of the point from this array.

6. Discussion

The comprehensive dataset and the introduction of new descriptors, as detailed in the results section, significantly enhance our understanding of rural areas. By incorporating proximity to Points of Interest (POIs) such as roads, cities, and social facilities, the graph descriptor structures — with their detailed geospatial coordinates and distance metrics — provide an in-depth examination of the accessibility and distribution of essential infrastructure.

Our statistical analysis uncovers several key trends. The right-skewed distribution of RD_m1_NEAR suggests that while many villages are connected to major roads, a substantial number are not, thus highlighting potential areas for infrastructure development. From the histogram presented in Figure 4, we can clearly see that most villages have direct access to secondary and rural roads. This access eliminates the need to allocate resources for such types of roads and facilitates ease of access, which aids in the development of infrastructure and businesses. Similarly, the log-normal distributions for distances to various POIs reveal disparities in access to vital services like education, healthcare, and retail, indicating a need for targeted improvements. Amenities such as education, churches, hospitals, and kindergartens, alongside less essential but still important facilities for people's comfort — such as local parks, shops, hotels, and national parks — display high first quartiles. This indicates that many villages possess well-developed infrastructure for these types of POIs. The triangular distribution for Kyiv_NEAR_ suggests relatively equitable access to the capital, which could benefit economic opportunities and service accessibility.

The boxplot in Figure 5 corroborates the descriptions provided above. It also offers additional insights, confirming that the distance to all types of amenities is typically under 50 kilometers, with some exceptions such as regional and national parks, or the distance to the capital. This finding is crucial in terms of society's ability to access social services quickly and easily. The analysis reinforces these insights, providing visual evidence of the distances to different facilities and identifying areas where disparities could be most effectively addressed. The correlation matrix further emphasizes the logical connection between village proximity to cities and the availability of facilities, affirming that urban centers typically offer more comprehensive infrastructure.

The alignment of the new graph descriptors with the original dataset indicates that our novel approach not only retains but also enriches the data's descriptive quality without altering its fundamental distribution characteristics. This consistency is vital for ensuring that any policy recommendations based on these descriptors accurately reflect real-world conditions. Furthermore, this research has yielded meaningful insights, such as the correlation between the presence of shops, medical facilities, kindergartens, and hotels. Given that the sectors of shops and hotels are predominantly private in Ukraine, these insights provide a clear direction for how to enhance overall village appeal. Improvements in public services such as medicine, kindergartens, and churches will likely lead to increased interest in the region, and consequently, more private investment. Accompanying the novel graph descriptors, a Figure 9 has been produced that displays the quantity of all Points of Interest (POIs) with a gradient scale, offering preliminary indications of which areas in Ukraine are the most underdeveloped and necessitate improvements. This map lays the groundwork for subsequent detailed analyses in future research endeavors.

This methodology supports the Ministry of Education and Science of Ukraine's objective to leverage technology for rural development by identifying spatial inequalities and infrastructure needs. Nevertheless, the study's recommendations must also take into account the socio-economic and political challenges inherent in implementing infrastructure improvements. These challenges include the mobilization of necessary resources and garnering political support.

Moreover, while the use of OpenStreetMap data is advantageous, it introduces concerns regarding data completeness and accuracy, as discussed by Herfort et al. (2023)[2]. These potential discrepancies in data quality, particularly in rural settings, could affect the study's conclusions and should be acknowledged as a limitation.

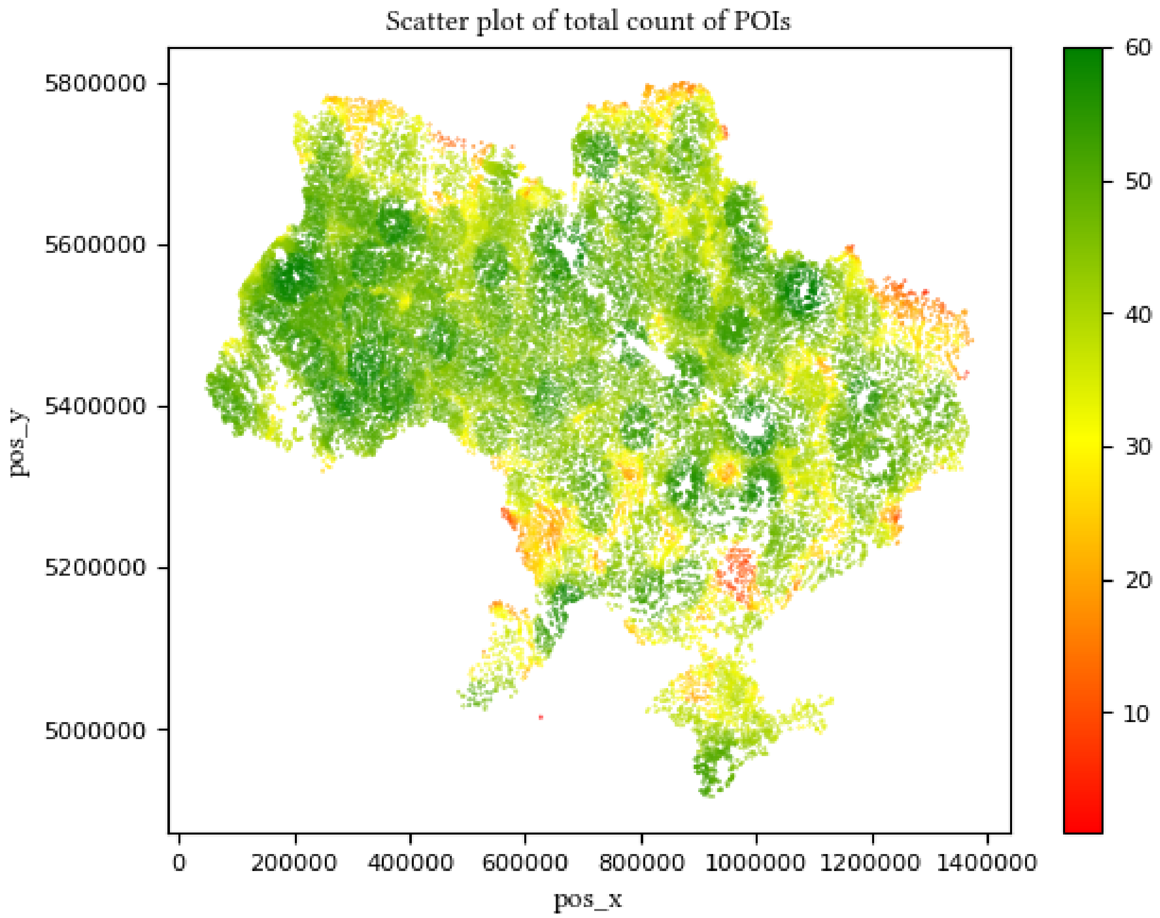

Figure 9: Total count of all types of Points of Interest (POIs) present for each village within the newly created graph structure, visualized with a gradient transitioning from green to red. In this representation, green indicates a higher availability of POIs, suggesting better accessibility for the respective villages.

7. Conclusion

This research has underscored the value of data-driven geospatial analysis in advancing rural development, especially within the uniquely challenging current context facing Ukraine. Beyond the acute infrastructural damages from the ongoing war, years of strained public spending have exacerbated rural accessibility gaps even in peaceful regions.

As such, identifying and prioritizing the highest impact revitalization investments is instrumental for balanced, sustainable recovery. Our study has paved an evidentiary path for such decisions by spotlighting pressing infrastructure deficits and access inequalities facing Ukrainian villages.

The multi-category accessibility linkage mapping to surrounding facilities provides localized actionability to rehabilitation policy. Granular quantifications also enable cost-optimization for connectivity improvements per impact on total beneficiaries. Appropriately directed rural health, education and transit upgrades promise significant welfare improvements per dollar.

Moreover, the project's backing from Ukraine's Ministry of Education and Science offers a conduit for translating these rural infrastructural insights directly into development programs under their remit. More broadly, our analytical blueprint promises transferability to guide strategic rebuilding worldwide after war involving extensive physical damage.

However, considerable challenges still stand in the way of on-ground implementation, especially financing gaps with state coffers drained by war costs. Overcoming these hurdles necessitates greater involvement of external development partners combined with private participation. Our village-tiered accessibility perspectives provide the evidence base for making competitive pitches for such inclusive cooperation.

Beyond the post-war context, this study has contributed methodologies advancing the descriptive depth of rural landscapes globally. The algorithms constructing enhanced infrastructure connectivity representations retain possibilities for myriad applications. As emerging data gathering mechanisms continue to accelerate across spheres like internet-of-things sensor networks, integrating such data streams through our graph model promises ever more fine-grained insights into infrastructural dynamics. This offers a versatile toolkit for navigating villages into more modern, equitable futures worldwide.

8. Acknowledgements

The study has been supported by the Ministry of Education and Science of Ukraine through the project “Information Technologies of Geospatial Analysis for the Development of Rural Areas and Communities" state registration number: 0123U102838. The authors are grateful to the project for providing the raw data for this analysis.

9. References

[1] Y. Liu, X. Ke, W. Wu, M. Zhang, X. Fu, J. Li, J. Jiang, Y. He, C. Zhou, W. Li, Y. Li, Y. Song, X. Zhou, "Geospatial characterization of rural settlements and potential targets for revitalization by geoinformation technology," Sci. Rep. 12 (2022) 8399. https://doi.org/10.1038/s41598-022-12294-2.

[2] B. Herfort, S. Lautenbach, J. Porto de Albuquerque, J. Anderson, A. Zipf, "A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap," Nat. Commun. 14 (2023) 3985. https://doi.org/10.1038/s41467-023-33956-z.

[3] H. Yailymova, B. Yailymov, N. Kussul, A. Shelestov, "Geospatial Analysis of Life Quality in Ukrainian Rural Areas," in Proceedings of the 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2023, pp. 1-5. https://doi.org/10.1109/DESSERT61349.2023.10416517.

[4] Maryada, A., Thatiparthi, V. L. "Geospatial technology for mapping and analysis of social and infrastructural facilities at village level: a case study of Chinnapendyala village," Modeling Earth Systems and Environment, 6, pp. 1763–1781, 2020. https://doi.org/10.1007/s40808-020-00788-9.

[5] Open Street Map. URL: https://download.geofabrik.de/europe/ukraine.html

[6] The Humanitarian Data Exchange. URL: https://data.humdata.org/

[7] Elevators Elevators in Ukraine. URL: https://elevatorist.com/karta-elevatorovukrainy

[8] Ye, X., Wei, Y. D. "Geospatial Analysis of Regional Development in China: The Case of Zhejiang Province and the Wenzhou Model," Pages 445-464, 2013. https://doi.org/10.2747/1538-7216.46.6.445.

[9] Roy, S. M., Prasad, N. S. R., Srinuvas, B., "Achieving Sustainable Village Development Through Geo-information Application," in ICT Analysis and Applications, Lecture Notes in Networks and Systems, vol. 314, pp. 833–843, 2022. https://doi.org/10.1007/978-981-16-3331-1_70.

[10] Pandey, S., Tripathi, G., "Development of a Smart Village Through Micro-Level Planning Using Geospatial Techniques–A Case Study of Jangal Aurahi Village of Gorakhpur District," in S. Kanga, V. N. Mishra, S. K. Singh (Eds.), 2020. https://doi.org/10.1002/9781119687160.ch6.

[11] Agustiono, W. "Smart Villages in Indonesia in the Light of the Literature Review," 2022 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, pp. 1-5, 2022. https://doi.org/10.1109/ICISS55894.2022.9915061.

[12] Liu Liu, M., Zhang, Q., Gao, S., Huang, J. "The spatial aggregation of rural e-commerce in China: An empirical investigation into Taobao Villages," Journal of Rural Studies, 80, pp. 403-417, 2020. https://doi.org/10.1016/j.jrurstud.2020.10.016.
[13] Bielska, A., Stańczuk-Gałwiaczek, M., Sobolewska-Mikulska, K., Mroczkowski, R. "Implementation of the smart village concept based on selected spatial patterns – A case study of Mazowieckie Voivodeship in Poland," Land Use Policy, 104, 105366, 2021. https://doi.org/10.1016/j.landusepol.2021.105366.
[14] Xu, J., Yang, M., Hou, C., Lu, Z., Liu, D. "Distribution of rural tourism development in geographical space: a case study of 323 traditional villages in Shaanxi, China," 2020. https://doi.org/10.1080/22797254.2020.1788993.
[15] A. Shelestov and L. Shumilo, "Generative Adversarial Network Style Transfer for Sustainable Urban Restoration Planning in Ukraine," in 2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2023, pp. 1-5. https://doi.org/10.1109/DESSERT61349.2023.10416492.