Assessing Correlation Between Night-Time Light and Road Infrastructure: An Empirical Study

Satabdi Das
Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Bangladesh
satabdi58@student.sust.edu

Sharmin Ahmed
Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Bangladesh
sharmin91@student.sust.edu

Summit Haque
Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Bangladesh
summit-cse@sust.edu

Sabir Ismail
Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Bangladesh
sabir-cse@sust.edu

Abstract

The inadequacy of spatially explicit and accessible data portals continues to be a substantial barrier for policymakers and concerned authorities in the least developed countries. The purpose of this study is to determine the potentiality of night-time light (NTL) data to measure spatial road infrastructure development. The Day-Night Band (DNB) NTL data from the Visible Infrared Imaging Radiometer Suite (VIIRS) as well as Google Maps highways road data (RD) were used in this research. In order to analyze the correlation between VIIRS NTL and RD for two least developed countries, we performed the Chi-square test of independence, which revealed that the variables are dependent on one another. Following that, we computed the Cramer's V test as a correlation coefficient to determine the strength of the association for both countries. Our findings revealed a correlation value of 0.334 in Bangladesh and a correlation value of 0.299 in Rwanda, demonstrating that VIIRS NTL and RD are strongly correlated. Following the discovery of a statistically significant correlation, we utilized the data to do more exploratory analysis.

1 Introduction and Related Research

World Bank, United Nations, and many organizations invest billions of dollars each year in accelerating the economic growth of the least developed countries. Ideally, capital aids are offered to the countries provided the assurance of proper utilization. In order to identify those countries, extensive investigation of reliable data is crucial. However, most data are acquired in the least developed countries through surveys with long intervals due to their expense and time requirements [12]. The long intervals as well as political instability make the data inconsistent and biased.

Remote sensing data can reduce these traditional data acquisition costs in such remote locations simultaneously improving the dependability, objectivity, and consistency of data. Especially, NTL data with its eminent socio-economic predictive features opened up many research opportunities in the least developed countries. However, lower spatial resolution, lack of on-board calibration, shorter radiometric range characteristics created obstacles to in-depth research in road infrastructure related studies with previous DMSP-OLS NTL data [6].

Recent studies demonstrate VIIRS NTL data have a larger pixel footprint, better resolution, and radiometric range which made it possible to overcome those limitations of DMSP-OLS NTL data. Rahman et al. [10] used VIIRS NTL to classify the major cities in Bangladesh to address inefficient infrastructure, power crisis due to unplanned and uneven development of the country. Xu et al. [11] found VIIRS NTL data to be negatively correlated with the distance of the primary road intersections. Chang et al. [6] introduced Highway Nighttime Traffic Prosperity Index (HNTPI) to accurately measure the highway traffic which showed a strong connection with socioeconomic factors like GDP, consumption per capita, population. The day-time satellite imagery does not portray how a road impacts the surrounding subjects as it only records any human-made impervious structure in the land-cover layers [5]. As VIIRS NTL data have already shown good relations with many socio-economic factors in prior studies, the existence of a strong correlation between VIIRS NTL and road infrastructure can help to integrate NTL data in practical road infrastructure planning in the least developed countries, reducing the dependency on outdated, inconsistent survey data.

Therefore, in this paper, we intend to evaluate the potentiality of the VIIRS NTL data to measure spatial road infrastructure development in Bangladesh and Rwanda as two least developed countries[1]. First, we would find the correlation between NTL and RD data, and based on our finding, we would address 3 research questions:

- **RQ1:** How is the night activity structure in the least developed country?
- **RQ2:** Are road infrastructures well distributed in the least developed country?
- **RQ3:** Is there any reflection of regional inequality from the data?

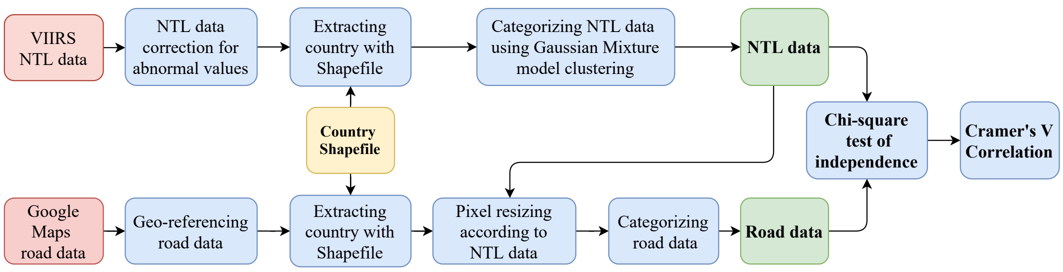

Figure 1 Workflow of the research

2 Data acquisition and preparation

In our research, we selected Bangladesh and Rwanda as two of the least developed countries. Despite being listed as least developed country, both countries have growing economy. In

[1] https://unctad.org/topic/least-developed-countries/list

2019, Bangladesh and Rwanda has GDP 302.57 billion (USD) and 10.34 billion (USD), respectively [3]. Focusing on the growing economy, it is important to understand the current state of road infrastructure to form a better development policies for the countries [12].

The source of VIIRS NTL data is NOAA Earth Observation Group (EOG) [1]. We selected the monthly VIIRS NTL data of April 2019 to maintain consistency with RD. Next, we used Google Maps to extract the road infrastructure of the selected countries. We downloaded the road maps from Google Static Maps API at zoom level 7 in April 2020. In the retrieval process from the API, we turned off every feature in the image except the highways. The resulting images were sized as 640×640 pixels. The resolution of the downloaded images was high enough to cover all the highways of the countries under study. We geo-referenced the downloaded images from Google Maps and resized the pixels to match with VIIRS NTL data. Then, we clipped the country extents from Google Maps images and VIIRS NTL rasters using shapefiles. As we only wanted to identify pixels with roads, we categorized the Google Maps image pixels into 0 or 1 based on the existence of roads. If a pixel contained a pixel of road, then it was assigned to 1. The rest of the pixels were set to 0. In addition to Google Maps, we used OpenStreetMap (OSM) primary and secondary road shapefiles of the countries under study as another road data source. For our analysis, we rasterized the OSM shapefiles so that the pixels can be categorizd as Google Maps RD. Our NTL data had some pixels containing negative values. For NTL data correction we replaced these negative values with NULL following previous studies [7]. The workflow summary is illustrated in Figure 1.

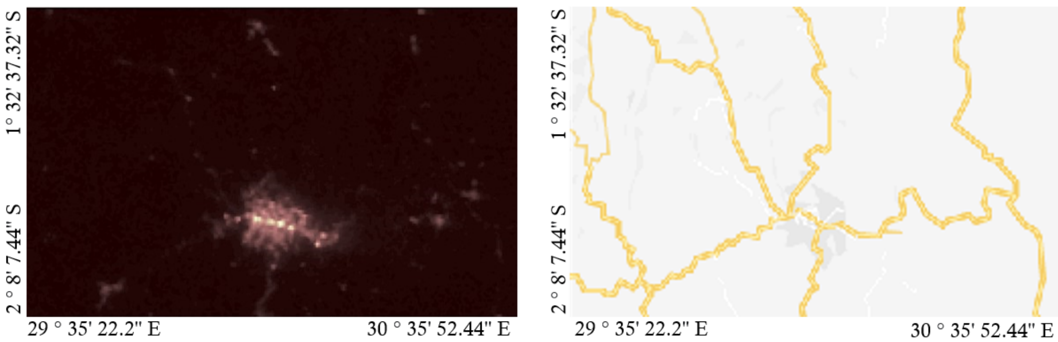

Figure 2 Visual evaluation between NTL and road of Rwanda

3 Experiment and Result

3.1 Correlation

We selected random areas from NTL and RD and compared them side by side to visually evaluate the correlation. In Figure 2, we noticed that NTL pixels are more luminous along the road sides, whereas pixels surrounding the intersections formed bright radiance clusters.

NTL data is continuous, but the RD is categorical. First, we converted continuous NTL data to categorical data in order to determine the correlation between NTL and RD. We classified NTL values into three classes by fitting a Gaussian mixture model (GMM) to the NTL intensity [12].

In table 1, for Bangladesh and Rwanda, we got three classes low, medium, and high based on NTL intensity. Then, we performed Chi-square test of independence to find the data dependency on that categorized NTL and RD. The p-value is 0 ($P < 0.05$), which refers to

Cluster	Bangladesh (Radiance $nWcm^{-2}sr^{-1}$)	Rwanda (Radiance $nWcm^{-2}sr^{-1}$)
Low	0.0-0.81	0.0-0.47
Medium	0.81-2.41	0.47-1.76
High	2.41-118.87	1.76-64.22

Table 1 NTL Radiance range in GMM clusters

the relationship between the categorical variables; they are dependent [8]. Afterwards, for finding the correlation between NTL and RD, we performed the Cramer's V test.

Country	NTL correlation with RD	NTL correlation with OSM
Bangladesh	**0.334 (very strong)**	0.17 (strong)
Rwanda	**0.299 (very strong)**	0.15 (moderate)

Table 2 Google Maps Road data (RD) and OpenStreetMap (OSM) Correlation results

Cramer's V correlation coefficient greater than 0.25 denotes very strong relationship [2]. In table 2, for Bangladesh and Rwanda, both countries Cramer's V correlation coefficient values signify a very strong relationship with RD. While comparing OSM to RD, we found RD is regularly updated and more accurate which is the reason for different outcomes [4].

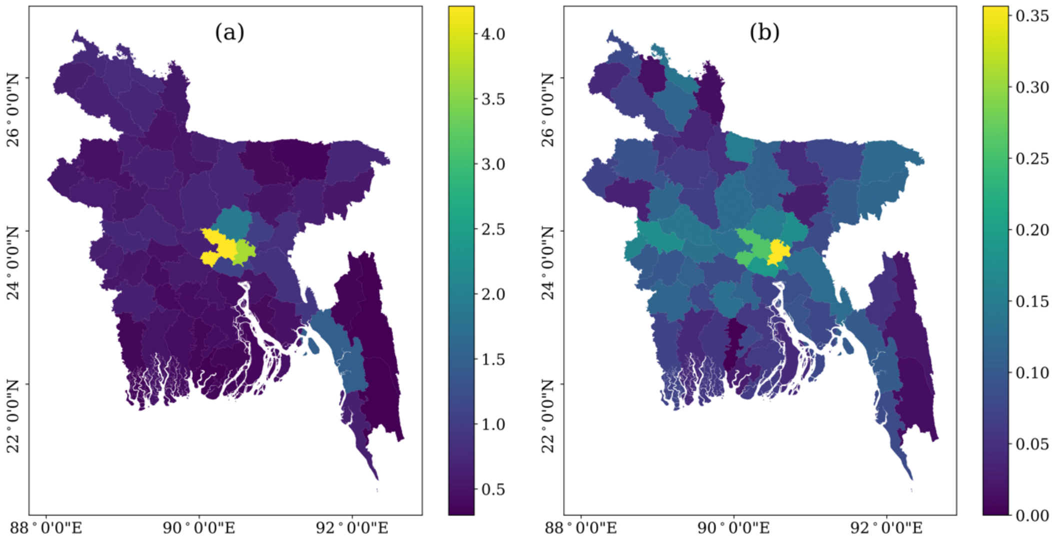

Figure 3 Bangladesh district level ratio (a) NTL; (b) road

3.2 Result Analysis

Our experiment and results shows that NTL and RD have strong correlation. Based on our initial finding we further investigate three research questions.

RQ1. First, we try to understand how night activity is propagated in the least developed countries among the various administrative areas. In Bangladesh, there are three major levels of administrative areas: division, district, and sub-districts. As for Rwanda, the three major levels of administrative areas are province, district, and sector. We used both countries district level administrative areas for our analysis. Bangladesh and Rwanda have 64 and 30 districts, respectively. To understand the NTL distribution at the district level, we measured the NTL ratio for each district. We used (1) to calculate the NTL ratio. The NTL ratio result of Bangladesh districts is plotted in Figure 3a.

$$\text{NTL ratio} = \frac{\text{total NTL intensity of corresponding area}}{\text{total pixels of corresponding area}} \tag{1}$$

From Figure 3a, we can notice, NTL is concentrated only within few districts. We have found a similar pattern of NTL distribution in Rwanda as well. Only a few districts have a high NTL ratio in Rwanda. It implies that the economic growth of those countries very much centered to few districts.

RQ2. Next, to understand the road infrastructure in the those countries, we calculated the road ratio at the district level of countries. We applied the same procedure as NTL ratio calculation in RQ2. Here, we counted the pixels containing road instead of NTL value (2).

$$\text{Road ratio} = \frac{\text{total pixels having road information of corresponding area}}{\text{total pixels of corresponding area}} \tag{2}$$

The road ratio result of Bangladesh districts is plotted in Figure 3b. From the visualization in Figure 3b we can observe that even if roads are not evenly distributed, it's not as concentrated as NTL ratio in Figure 3a. For Rwanda, we observed a similar road distribution pattern. The district road ratio distribution is not well dispersed in Rwanda.

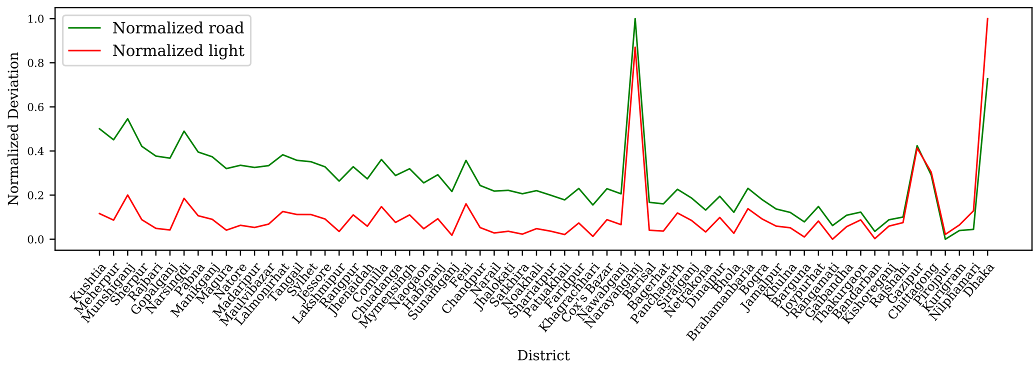

Figure 4 Normalized light-road deviation of Bangladesh

RQ3. Our next finding is there exists any regional inequality in both countries. We normalize the NTL ratio (1) and Road ratio (2) of the every district by subtracting from the mean. We found the standard deviation of the normalized NTL ratio and Road ratio of Bangladesh to be 0.637 and 0.059, and for Rwanda to be 1.04 and 0.052, respectively. For both countries NTL deviation is higher than road deviation which signifies that even in the presence of relatively better road infrastructure in the districts, electricity is not provided in that proportion. Next we again normalized NTL ratio and road ratio to same scale, Figure 4. As NTL is an economic development indicator, Figure 4 portrays that most of the districts have little or no development. It is observable that only major districts have noticeable economic growth. Rahman et al. [10] mentioned few districts are the main urban centers in Bangladesh, which are the capital Dhaka, industrial district Gazipur, Narayanganj and port city Chittagong. Most economic activities are concentrated around these cities.

Other districts are very much lagging behind. Thus, the inequalities between the regions are reflected in the data. In Rwanda, we also observed a similar kind of distribution. Only Gasabo, Kicukiro, and Nyarugenge districts have notable economic growth, which are under one of the fastest growing provinces, Kigali city [9].

4 Conclusion and Future Work

The main objective of our study is to evaluate NTL data's capability to add value in assessing spatial road infrastructure development as NTL data is more reliable, inexpensive and readily available. We discovered in Bangladesh and Rwanda that NTL and RD are strongly correlated. Although in our study there is a gap between our NTL and RD's time frame, there is less possibility of any drastic change in road infrastructure within a short period. This investigation suggests that NTL data can be considered to be used in assessing road infrastructure development. Furthermore, our exploratory analysis portrayed that NTL data can reflect the road infrastructure development, power crisis, economic development, and regional inequality in the least developed country to a great extent.

RD used in this research is mostly highways and major roads. Future work includes, depth analysis with urban and local roads can be utilized to shed more light on the applicability of NTL data in road infrastructure development. Several satellite characteristics, such as road density, water data, building density, agricultural land, agricultural type, area, urban are highly used for socio-economic prediction. As satellite feature is closely connected with economic well-being, we can add more features with NTL to predict where road development is needed. Day-time satellite imagery covers more information about man-made structures, and NTL data illustrates how these structures affect the surrounding subjects. Therefore, aggregation of NTL data with day-time satellite imagery can also give us more insights. Since, the data used in the investigation is public this analysis is producible for other countries as well.

References

1 Version 1. Viirs day/night band nighttime lights, last accessed December 2020. URL: `https://eogdata.mines.edu/download_dnb_composites.html`.

2 Haldun Akoglu. User's guide to correlation coefficients. *Turkish journal of emergency medicine*, 18(3):91–93, 2018.

3 The World Bank. Data. URL: `https://data.worldbank.org`.

4 Chahat Bansal, Aditi Singla, Ankit Kumar Singh, Hari Om Ahlawat, Mayank Jain, Prachi Singh, Prashant Kumar, Ritesh Saha, Sakshi Taparia, Shailesh Yadav, et al. Characterizing the evolution of indian cities using satellite imagery and open street maps. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 87–96, 2020.

5 Kathryn Baragwanath, Ran Goldblatt, Gordon Hanson, and Amit K Khandelwal. Detecting urban markets with satellite imagery: An application to india. *Journal of Urban Economics*, page 103173, 2019.

6 Ying Chang, Shixin Wang, Yi Zhou, Litao Wang, and Futao Wang. A novel method of evaluating highway traffic prosperity based on nighttime light remote sensing. *Remote Sensing*, 12(1):102, 2020.

7 Zuoqi Chen, Bailang Yu, Yingjie Hu, Chang Huang, Kaifang Shi, and Jianping Wu. Estimating house vacancy rate in metropolitan areas using npp-viirs nighttime light composite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5):2188–2197, 2015.

8 Mary L McHugh. The chi-square test of independence. *Biochemia medica: Biochemia medica*, 23(2):143–149, 2013.

9 Gilbert Nduwayezu, Richard Sliuzas, and Monika Kuffer. Modeling urban growth in kigali city rwanda. *Rwanda Journal*, 1(1S), 2016.

10 Md Shahinoor Rahman, Hossain Mohiuddin, Abdulla-Al Kafy, Pintu Kumar Sheel, and Liping Di. Classification of cities in bangladesh based on remote sensing derived spatial characteristics. *Journal of Urban Management*, 8(2):206–224, 2019.

11 Huimin Xu and Xi Li. Evaluating spatial details of luojia-1 night-time images using road network analysis. In *Geoinformatics in Sustainable Ecosystem and Society*, pages 122–131. Springer, 2019.

12 Xizhi Zhao, Bailang Yu, Yan Liu, Zuoqi Chen, Qiaoxuan Li, Congxiao Wang, and Jianping Wu. Estimation of poverty using random forest regression with multi-source data: A case study in bangladesh. *Remote Sensing*, 11(4):375, 2019.