

Strategic Planning of Rural Telecommunication Infrastructure: A Multi-Source Data Fusion and Optimization Model

Ferdaous Chaabane¹, Safa Réjichi¹, Housseem Ben Salem¹, Hassen Elmabrouk¹, Florence Tupin²

¹ COSIM laboratory, SUP'COM, Carthage University, Tunisia

² Department of Image and Signal Processing, Telecom ParisTech, France

Keywords: Rural telecommunications planning, deep embedded clustering, network optimization, infrastructure topology, multi-source data fusion.

1. Abstract

The United Nations has underscored the critical role of digital connectivity by integrating it into several sustainable development goals, with the ambition for nations worldwide to achieve comprehensive access by 2030. Over thousand million people, primarily in rural areas, are disconnected from the digital world, highlighting the urgent need for viable and sustainable telecommunications solutions. These areas are characterized by sparse populations and lower economic levels, presenting great challenges for connectivity.

This work introduces a strategy for enhancing rural telecommunication planning using geospatial and remote sensing data, deep learning-based clustering techniques, network graphs, and terrain analysis. The objective is to develop an optimal network topology and identify prime locations for telecommunications infrastructure, such as towers or relay stations.

The methodology begins with the application of an adapted Deep Embedded Clustering (DEC) technique to identify community boundaries accurately. Then, it combines geospatial data (such as roads, terrain slope, flatness, etc.) and remote sensing data (vegetation, waterways, etc.) through an optimization algorithm. This process aims to determine the most suitable sites for infrastructure placement and the best network topology for connecting these areas. The study focuses on the region of Congo, offering a detailed case study on the application of this approach. Experimental results are presented to demonstrate the effectiveness of the proposed telecommunications expansion strategy.

2. Introduction

The critical importance of telecommunications connectivity in rural areas of developing countries cannot be overstated, serving as a cornerstone for comprehensive development and enhanced quality of life. Numerous rural communities across the African continent face significant challenges due to inadequate or non-existent digital connectivity, impeding progress in various sectors. As Africa aims for sustainable development, the expansion of telecommunications infrastructure in rural areas emerges as a vital step in closing the developmental divide and fostering a more equitable and prosperous future for all.

However, extending telecommunications networks in rural African settings is fraught with challenges, particularly when relying on traditional planning techniques that lack precise optimization, environmental consideration, and scalability. Overcoming these obstacles requires a shift towards more

sustainable and efficient solutions, leveraging new technologies and advanced multi-source data.

Recent advancements in telecommunications planning have introduced sophisticated methodologies to optimize the distribution and placement of cell towers, focusing on cost reduction while ensuring service quality. Hoomod et al. proposed a technique leveraging self-organizing maps for optimal cell tower distribution, emphasizing the balance between coverage quality and cost-efficiency [1]. It clusters the subscribers (nodes) by using a traditional clustering technique to find minimum appropriate number of cell towers and optimizing their distribution. Besides, an algorithmic approach, suggested in [2], leverages satellite imagery and GIS (Geographic Information System) data to identify strategic cell tower placements, highlighting the importance of topographical considerations in minimizing infrastructure costs. This research aims to find a simple implementable algorithm which effectively determines the strategic positions of the cell towers. Given a satellite image and population density, and obtaining topographical information from GIS (Geographic Information Systems), potential tower locations can be determined.

With the same vision, Tayal et al. [3] proposes a GIS based methodology for assessing the site suitability of Base Stations of BSNL (Base Station Network Layer) cellular radio networks with the objective of optimizing and automating the process of network planning. Geographic information, such as satellite images, topographical maps, municipal digital maps, Aster DEM, site parameters of existing BSNL towers (such as Latitude and Longitude, antenna height, frequencies) of study area are collected from different sources to locate the suitable sites.

Furthermore, Bharadwaj et al. [4] demonstrate the potential of integrating 3D terrain data through LiDAR (Light Detection And Ranging) to enhance signal quality across diverse topographies, underscoring the technological advancements in pinpointing optimal cellular tower locations. However, such approaches introduce significant cost implications, particularly for extensive rural areas. Similarly in [5] a LiDAR-based technique is attempted using the point cloud data to extract building, ground, tree, etc. Then boundary or edges inside them coming in the path between potential transmission tower to the user are tried. The approach is extended to determine the best location of transmission tower for the region of study.

Besides, analogous challenges and solutions in rural electrification planning were proposed. As evidenced in the strategic methodologies of Ganguly et al. [6], Blechinger et al. [7], and Dimovski et al. [8], our study adopts a multi-disciplinary approach.

These references in electrical network planning, employing optimization techniques like particle swarm optimization and MILP for cost-effective electrification, serve as a conceptual foundation for addressing the telecommunications infrastructure deployment in rural areas. This cross-disciplinary insight encourages the exploration of similar optimization and strategic planning methodologies, tailored to the telecommunications context, aiming to devise a scalable and cost-efficient framework for network expansion.

Addressing these challenges, our paper introduces a novel framework that, unlike methods reliant on expensive LiDAR sensors, utilizes deep learning-based clustering, geospatial, and remote sensing data alongside optimization algorithms to develop an optimal network topology and identify prime locations for telecommunications infrastructure at a lower cost. This approach ensures efficient coverage and connectivity in underserved rural areas without incurring the high expenses associated with LiDAR technology. By focusing on Congo as a case study, we demonstrate the effectiveness of our proposed strategy, offering a scalable solution that can be adapted to various rural contexts worldwide.

Structured in three sections, our study first outlines the methodology, incorporating insights from existing literature and highlighting our innovative, cost-efficient techniques for network planning. We then present a detailed case study on Congo, showcasing the application of our methodology and the results achieved. Finally, we conclude with a discussion on the implications of our findings for rural telecommunications expansion, setting the stage for future research and implementation.

3. Methodology

The procedure developed in this study aims to address the limitations of traditional connectivity expansion techniques and proposes a novel methodology to establish an optimal telecommunications plan for expansive rural areas. This approach integrates various concepts from the rural telecommunications planning paradigm, such as embedded deep learning clustering, network graph optimization, and multi-source data fusion, etc.

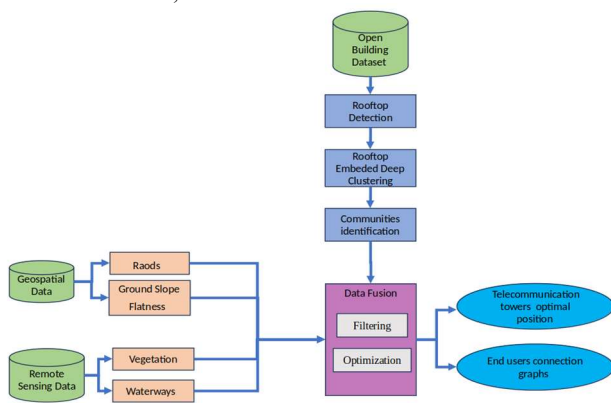


Figure1: Flowchart of the proposed methodology

Figure 1 illustrates a flowchart of the entire methodology and the data utilized in each phase. The initial step involves telecommunications network analysis and clustering, where the objective is to segment rural areas lacking connectivity. This segmentation is based on information from existing networks or satellite images showing coverage gaps. The Deep Embedded

Clustering (DEC) method is employed to define community clusters and determine the most suitable boundaries based on the spatial concentration of buildings. Following this, geospatial data and information derived from remote sensing images are used to identify the optimal locations for telecommunications infrastructure, such as towers or relay stations, within each cluster. This step also involves designing the most efficient network topology to connect these infrastructures to roads and ensure optimal coverage for the end users. A telecommunications tower or relay station is proposed for implementation within each identified community cluster.

During this phase, integrated data is used through a one-shot single optimization algorithm, aiming to maximize a score function derived from all the outputs of previous steps.

3.1. Buildings Clustering

To accomplish building clustering, we first need to apply preprocessing to VHR satellite images such as georeferencing correction and pansharpening. Then we proceed to the detection and extraction of the roof surfaces edges using Open Building public dataset which contains building footprints [9]. This large-scale open dataset contains the outlines of buildings derived from high-resolution satellite imagery and auxiliary data in order to support different types of use cases including our project.

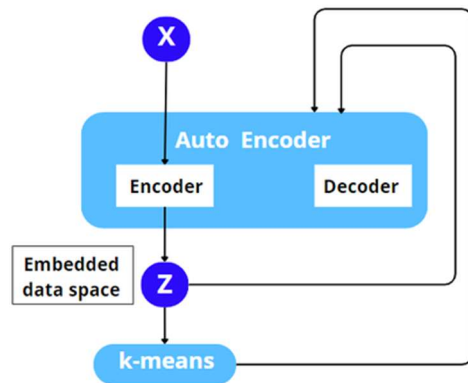


Figure2: Flowchart of the Embedded Deep Clustering Technique

Then the DEC illustrated by Figure2 is applied [10]. Its goal is to learn a cluster-friendly embedding by jointly training an unsupervised deep neural network with a clustering algorithm. Autoencoders are the most common form of unsupervised deep learning architectures used to first encode input data ($X \in \mathbf{R}^d$) into a latent space or embedding ($Z \in \mathbf{Z}^m$) where $d < m$. The input data are reconstructed from the embedding using a decoder module as (\hat{X}). The encoder and decoder involve a set of trainable parameters $\theta \in \{W_\theta, b_\theta\}$ and $\Phi \in \{W_\Phi, b_\Phi\}$, respectively. The outputs of the encoder (Z) and decoder (\hat{X}) are obtained as follows.

$$Z = f(\theta, X) \tag{1}$$

$$\hat{X} = g(\Phi, f(\theta, X)) \tag{2}$$

Here, $f(\cdot)$ and $g(\cdot)$ represent sigmoid activation functions to introduce non-linearity in embedding. The learning objective of an autoencoder is to minimize the following reconstruction loss, updating θ and Φ parameters.

$$L_{recon} = \underset{\phi, \theta}{\operatorname{argmin}} \sum_{i=1}^N \|X_i - \hat{X}_i\|^2 \quad (3)$$

The autoencoder embedding (Z) is known to retain all information about the input data to facilitate perfect data reconstruction. Instead, an ideal embedding should emphasize or retain information useful for clustering. The cluster assignments and centroids are used to compute an embedding distribution (Q) or pseudo labels as a learning target.

A target distribution (P) is mathematically derived from the embedding distribution (Q). The overall learning objective is to update the autoencoder's trainable parameters (θ, ϕ) by jointly minimizing the reconstruction loss and the divergence between P and Q distributions, as below.

$$L = L_{recon} + \gamma * L_{cluster}$$

$$= \underset{\phi, \theta}{\operatorname{argmin}} \sum_{i=1}^N \|X_i - \hat{X}_i\|^2 + \gamma \sum_{i=1}^N \sum_{j=1}^K p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (4)$$

Here, N is the number of samples, K denotes the number of clusters, and γ is the trade-off parameter between the reconstruction and clustering losses.

3.2. Communities Identification

3.2.1 Community Definition Criteria

To accurately represent the rural landscape, we define a "Community" as a cluster that adheres to the following criteria to emphasize the compactness of rural habitation :

- Contains more than 100 and fewer than 10K rooftops,
- Is characterized by proximity; rooftop located more than 300 meters from the nearest neighbour, is considered outside the cluster.

3.2.2. Spatial Indexing and Iterative Clustering

For this step, an iterative clustering process enhanced by spatial indexing to efficiently manage the vast geographical data is used.

1/ DEC Initial Cluster Formation: We start by applying the DEC to generate preliminary clusters. DEC utilizes spatial coordinates and the density of rooftops to create clusters by optimizing internal homogeneity and external separation.

2/ Spatial Indexing: After DEC formation, the R-Trees spatial indexing structures are implemented. These structures allow us to quickly query and analyze spatial relationships among the data points. When the initial clusters are established, we apply an iterative refinement phase by assessing each cluster against the community criteria. Finally, the iterative process culminates in the definitive determination of community boundaries.

3.3. Community Telecommunications Infrastructure Development Process

The foundation of the telecommunications tower or relay station positioning begins with a comprehensive gathering of multi-source data, which provides a detailed view of the terrain and is crucial for decision-making.

3.3.1 Buffer Zone Formation

Buffer zones around each identified community cluster are created. These buffer zones are then divided into multiple grids. The size of each grid is determined based on the estimated coverage requirements of the area. Each grid represents a potential site for the placement of a telecommunications tower or relay station, considering spatial constraints and the connectivity needs of the community.

The following table encapsulates the relationship between tower capacity and the required ground space for telecommunications infrastructure:

Table1 : Tower capacity vs. required ground space.

Tower Capacity (simultaneous users)	64	97	127	193	254
Minimum Footprint (m ²)	320	460	580	860	1150

3.3.2. Application of Filtering Techniques

Subsequent to grid formation, we applied a series of filtering techniques to refine the selection of potential tower locations.

The vegetation and the waterways areas are extracted from VHR images covering the region of study. We used the technique proposed in [11] using an SVM based classification algorithm combining textural, spatial, and spectral features to extract vegetation and water areas. The roads and the ground elevation models are extracted from OpenStreetMap.

1/ Vegetation Filter: Employing vegetation density data, we excluded grids with high vegetation coverage that would hinder construction or operation of telecommunications towers.

2/ Waterways and Water-Free Areas Filter: Grids intersected by waterways or situated in water-free areas were eliminated to respect natural water bodies and mitigate any environmental impact.

3/ Road Filter: Any grid bisected by roads was also removed from consideration to prevent disruption of existing infrastructure and to ensure safety and accessibility.

This filtering technique not only ensures that we identify the most environmentally and logistically suitable sites for the installation of telecommunications towers, but also significantly reduces the complexity of the subsequent optimization problem (concentrate the optimization efforts on a smaller set of grids).

3.3.3 Tower Site Optimization

For this step, we use the ground elevation model to extract slope statistics from each grid, followed by a scoring algorithm that assesses the suitability of locations for telecommunications tower installation.

3.3.3.1 Slope Analysis and Scoring

Using a digital elevation model, we extract the slope information for each grid to determine the terrain's flatness. The Flatness score for each grid reflects its suitability for tower construction, with flatter terrains being more favourable and thus receiving a higher score. The overall score for each grid (g) incorporates the following metrics:

Flatness : This score is based on the inverse of the average slope within a grid, favouring flatter terrains.

Proximity to Community Density: The distance score is based on the proximity to the densest point.

Road Accessibility: A road distance score evaluates the grid's accessibility, with closer proximity to roads indicating better logistical convenience for construction and maintenance.

3.3.3.2 Weighted Scoring Formula

The total score $T(g)$ for a grid g is determined by a weighted sum of these factors:

$$T(g) = W_F \frac{1}{1+S(g)} + W_D \frac{1}{D(g)} + W_{RD} \frac{1}{RD(g)} \quad (5)$$

Where $S(g)$ is the Slope, $D(g)$ is the distance and $RD(g)$ is the road distance. The W_i are the corresponding weights. The optimization algorithm seeks to find the grid with the highest $T(g)$ within each zone. Formally, for each zone z the best grid \hat{g}_z^{best} is the one that maximizes the total score:

$$\hat{g}_z^{best} = \underset{g \in G}{argmax} T(g) \quad (6)$$

This approach significantly streamlines the selection process, effectively reduces the complexity of the optimization problem and focuses resources on the most promising sites.

4. Results

The proposed approach was adapted for a telecommunications expansion project in the region of Kivu, Democratic Republic of the Congo, Africa, an area known for its challenging terrain, including mountainous landscapes and dense forests. By identifying optimal tower sites, it aimed to improve connectivity for isolated communities, demonstrating a vital step toward closing the digital gap in the area.

4.1. Communities clustering and Identification

The Open Building Dataset is used to detect the rooftops. For buildings accuracy detection, the Intersection over Union (IoU) metric was used. The IoU is equal to 0.832 which is a very good prediction of buildings boundaries.

Then the spatial indexing and Iterative DEC are applied to define community clusters. Figure3 and 4 show the difference for the fitting of building concentration between DEC and traditional techniques.



Figure3 : Clustering using traditional techniques

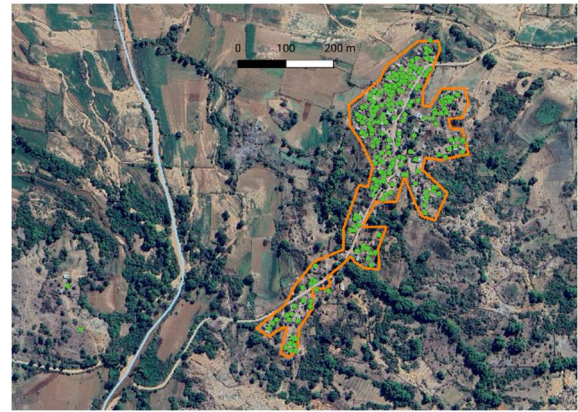


Figure4 : Deep Embedded Clustering results.

There is several metrics used to evaluate the performance of clustering techniques. These metrics provide a quantitative basis for comparison and help illustrate why DEC outperformed traditional methods. These metrics are: Inertia, Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index [12].

Table2 : Clustering metrics comparing DEC to traditional techniques.

Clustering Technique	Inertia	Sihouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
K-means	3.40	0.45	-	-
DBSCAN	3.10	0.52	4.87	-
HDBSCAN	2.15	0.56	3.65	690792.01
DEC	5.90	0.71	0.53	1090772.61

The DEC algorithm stands out with its superior performance, as evidenced by the highest Silhouette Score and Calinski-Harabasz Index, alongside the lowest Davies-Bouldin Index. These results indicate that DEC not only creates more cohesive and well-separated clusters but also excels in defining clearer cluster boundaries compared to traditional methods like K-means, DBSCAN, and HDBSCAN. The substantial difference in these metrics underscores the effectiveness of DEC in accurately grouping buildings for the purposes of telecommunication tower positioning, setting a robust foundation for subsequent planning and implementation phases.

4.2. Telecommunications Tower positioning and Network Expansion Planning

It may be technically feasible to place a telecommunications tower in forested or water regions, but the environmental, regulatory, and especially logistical challenges make it impractical and often undesirable.

Thus, the technique proposed in [11] was applied to extract the vegetation and waterways areas illustrated by Figure5. As we can see these regions surround and are nested in community clusters and need to be taking into account for the network expansion planning.

Then, the buffer zoning grid division and filtering is applied as explained in section 3.3.1 (cf. Figure 6 and 7).



Figure5: Detection of Vegetation and Waterways areas

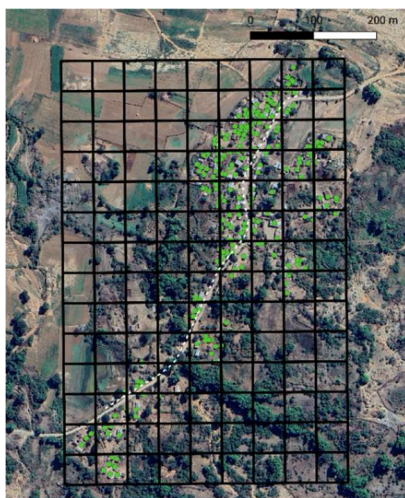


Figure 6: Buffer Zone Grid Division.

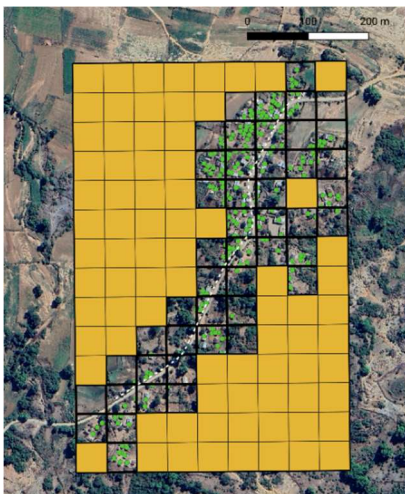


Figure 7: Grid filtering techniques applied.

Besides, the geospatial data extracted from OpenStreetMap is a mix of raster and vector layers which provide the roads and the ground elevation model from which the slope and flatness are computed.

The advanced features generated from remote sensing data, the extracted geospatial data and the obtained communities clusters are combined through the optimization algorithm presented in section 2.3. Figure8 shows the obtained results for tower placement and network connectivity.

As mentioned, each community cluster is designated to receive a tower, based on network coverage needs and land use analysis.

As we can see the telecommunication towers are placed in regions deprived of vegetation and waterways. The optimisation algorithm takes also the ground slope and the closest roads into consideration which can't be illustrated by Figure8.

We can notice some rooftops that are not connected to the network. This is due to the extraction of the roof surfaces edges using Open Building public dataset which contains some missing building footprints.

5. Conclusion

The methodology presented is an automated tool designed for planning rural telecommunications, aiding stakeholders in decision-making by providing comprehensive expansion strategies. It utilizes open-source and multi-source data to determine the most effective locations for telecommunications infrastructure, highlighting the process's effectiveness in Congo's rural contexts. This automatic approach can be applied to any rural region from the moment that all the input data is available. It has the capacity to span expansive areas while considering time constraints related to consumption.

As a future work, we may propose to include inter-community connectivity and incorporating cost analyses to optimize broader network integration challenges.

6. References

- [1] H. K. Hoomod, I. Al-Mejibli, and A. IssaJabboory, "Optimizing SOM for cell towers distribution," 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT), Baghdad, Iraq, 2017, pp. 138-143.
- [2] R. Kashyap, M. S Bhuvan, S. Chamarti, P. Bhat, M. Jothish, K. Annappa, "Algorithmic Approach for Strategic Cell Tower Placement," IEEE 2014 Fifth International Conference on Intelligent Systems, Modelling and Simulation, 2166-0662.
- [3] T. Shikha, P. K. Garg, and S. Vijay. "Site suitability analysis for locating optimal mobile towers in Uttarakhand using GIS." Proceedings of the 38th Asian Conference on Remote Sensing—Space Applications: Touching Human Lives, ACRS. 2017.
- [4] Bharadwaj, S.; Dubey, R.; Zafar, M.I.; Tiwary, S.K.; Faridi, R.A.; Biswas, S. "A Novel Method to Determine the Optimal Location for a Cellular Tower by Using LiDAR Data." Appl. Syst. Innov. 2022, 5, 30.
- [5] S. Bharadwaj, R. Dubey and S. Biswas, "Determination of the best location for setting up a transmission tower in the city," 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing, Aurangabad, India, 2020, pp. 63-68.
- [6] S. Ganguly, N.C. Sahoo, D. Das, "Mono- and multi-objective planning of electrical distribution networks using particle swarm optimization," Applied Soft Computing Journal, vol. 11, no. 2, pp. 2391-2405, 2011.



Figure8: Telecommunications Towers Placement and Network Topology for Community Connectivity

[7] P. Blechinger, C. Cader, P. Bertheau, "Least-cost electrification modeling and planning—A case study for five Nigerian federal states," *Proc. IEEE*, vol. 107, no. 9, pp. 1923–1940, 2019.

[8] A. Dimovski, S. Corigliano, D. Edeme, M. Merlo, "Holistic MILP-based approach for rural electrification planning," *Energy Strategy Reviews journal*, vol. 49, 101171, 2023.

[9] F. Chaabane, S. Réjichi, F. Tupin, "Rooftop Surfaces and Types Identification Using VHR Satellite Images for Electrification Potential in Africa," *IGARSS 2023*: 5301-5304.

[10] J. Xie, R. Girshik, A. Farhadi, "Unsupervised Deep Embedding for clustering Analysis," *Computer Vision and Pattern*, 24 May 2016.

[11] S. Rejichi, F. Chaabane, F. Tupin, "Expert knowledge-based method for Satellite Image Time Series analysis and interpretation," *IEEE JSTARS*, vol.8, no.5, pp.2138-2150, May 2015.

[12] H. Zhong, H. Zhang, F. Jia, "Analysis and improvement of evaluation indexes for clustering results," *EAI Endorsed Transactions on Collaborative Computing*, 2020, Volume 4.