Integrating Medical GIS , AI and Predictive Machine Learning to Improve Health habitation Accessibility of Kakinada District

Mrs. Tanaki UshaRani,
PG Scholar Pydah College of Engg,
Kakinada, AP, India

Mr. P Chakradarrao,
Asst.Prof Pydah College of Engg,
Kakinada, AP, India

Mr. Suribabu Boyidi
PG Scholar Pydah College of Engg
Kakinada, AP, India

Abstract: Based on population, Community health centers, Primary Health Centers (PHCs) and sub centre of health habitations is strengthening the healthcare system. This study assesses PHC service gaps in rural areas of Kakinada District, Andhra Pradesh, through an integrated GIS–Machine Learning framework. At the geospatial layer, PHC locations are mapped and buffer-based service zones (3 km and 5 km) are intersected with habitation data to identify underserved regions. To better reflect realistic accessibility constraints, the framework supports road-network-based distance estimation; however, in the present study, connectivity is represented using a proxy indicator approach, which can be extended to true network-cost analysis in future work.At the analytical layer, mandal-level indicators derived from healthcare infrastructure datasets are utilized for unsupervised clustering and interpretable decision-making. Using data from 21 mandals, 121 PHCs, and 840 sub-centres, K-means clustering identifies high-load regions, while a decision tree model provides prioritization rules for Auxiliary Nurse Midwife (ANM) deployment and PHC service enhancement.

Spatial analysis reveals significant disparities in PHC accessibility, especially in remote rural mandals where travel time exceeds recommended limits. Predictive machine learning models identify high-risk mortality zones strongly correlated with poor healthcare accessibility. The integrated GIS-ML approach enables optimal siting of new PHCs and strategic upgrades of existing facilities.

The study demonstrates that combining geospatial analytics with predictive modeling provides a robust decision-support system for rural healthcare planning. The findings offer scalable, data-driven insights for policymakers to improve healthcare accessibility, optimize resource allocation, and reduce preventable mortality in underserved regions.

Keywords

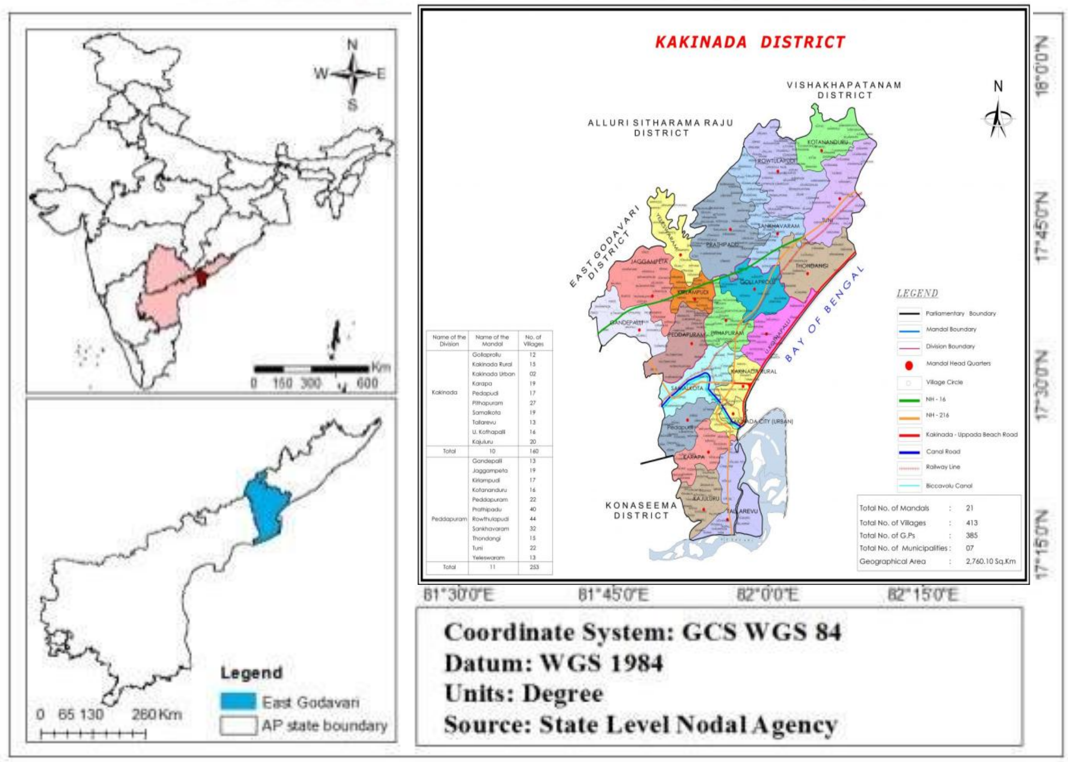
Medical GIS, Machine Learning, Primary Health Centres, Spatial Accessibility, Mortality Reduction, Rural Health Planning, Kakinada District

1. Introduction

Artificial Intelligence (AI) is the branch of computer science that deals with the simulation of intelligent behavior in computers.1 The term was first coined by a group of researchers in the year 1956. 2 Though application of Artificial Intelligence has been more rampant, still in poor countries with low resource settings, the use is still lying nascent. For that the needs and opportunities have to be deployed for the optimization of public health services. In 2017, the United Nations (UN) convened a global meeting to discuss the development and deployment of AI applications to reduce poverty and deliver a broad range of critical public services. More recently, another UN meeting including various stakeholders to assess the role AI in achieving the Sustainable Development Goals (SDGs) was discussed.3 Some experts opine that Artificial Intelligence acts intelligently in several aspects like a) works according to the appropriateness of circumstances and goals b) changes flexibly with the changing environment and goals c) learns from past experience d) works within the limit of perceptual and computational limits.4 AI enables computers to mimic the cognitive function of human minds, by using AI to review vast sets of real-time data, health experts can I Integrating Medical Geographic Information Systems (GIS), AI and predictive machine learning (ML) offers a powerful, data-driven approach to enhance primary healthcare (PHC) access and reduce mortality in rural Kakinada, Andhra Pradesh. By analyzing spatial data—such as road networks, population density, and existing PHC locations—alongside predictive models for disease outbreaks and patient risk, health authorities can optimize resource allocation and enable proactive, timely care, reducing the 5.49 km average village-to-PHC distance barrier.

- Importance of PHC accessibility in rural healthcare
- Challenges in rural healthcare delivery in India
- Role of GIS and Machine Learning in healthcare planning
- Problem statement (limited accessibility → higher mortality)
- Objectives of the study

LOCATION MAP OF EAST GODAVARI DISTRICT



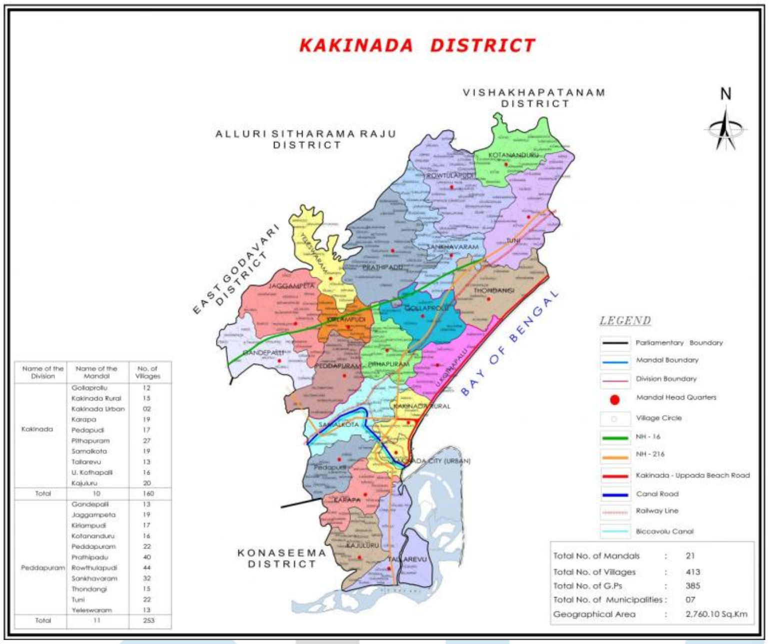
Name of the Division	Name of the Mandal	No. of Villages
Kakinada	Gollaprolu	12
	Kakinada Rural	15
	Kakinada Urban	02
	Karapa	19
	Pedapudi	17
	Pithapuram	27
	Samalkota	19
	Tallarevu	13
	U. Kothapalli	16
	Kajuluru	20
Total	10	160
Peddapuram	Gondepalli	13
	Jaggampeta	19
	Kirlampudi	17
	Kotananduru	16
	Peddapuram	22
	Prathipadu	40
	Rowthulapudi	44
	Sankhavaram	32
	Thondangi	15
	Tuni	22
	Yeleswaram	13
Total	11	253

Total No. of Mandals : 21
Total No. of Villages : 413
Total No. of G.Ps : 385
Total No. of Municipalities : 07
Geographical Area : 2,760.10 Sq.Km

Coordinate System: GCS WGS 84
Datum: WGS 1984
Units: Degree
Source: State Level Nodal Agency

2. **Study Area:** Kakinada District is located in the coastal region of Andhra Pradesh and includes urban, semi-urban, and rural mandals with heterogeneous settlement patterns, transport connectivity, and environmental characteristics. This diversity makes the district suitable for evaluating spatial inequality in PHC coverage and accessibility. Kakinada District overview Area: 3,019.79 sq. km Population: is 20.92 lakhs Total Population as per 2011 Census is 1,936,809 Rural Population (2011): 65.78% Urban Population: 34.22%.

LOCATION MAP OF KAKINADA DISTRICH



Name of the Division	Name of the Mandal	No. of Villages
Kakinada	Gollaprolu	12
	Kakinada Rural	15
	Kakinada Urban	02
	Karapa	19
	Pedapudi	17
	Pithapuram	27
	Samalkota	19
	Tallarevu	13
	U. Kothapalli	16
	Kajuluru	20
Total	10	160
Peddapuram	Gandepalli	13
	Jaggampeta	19
	Kirlampudi	17
	Kotananduru	16
	Peddapuram	22
	Prathipadu	40
	Rowthulapudi	44
	Sankhavaram	32
	Thondangi	15
	Tuni	22
	Yeleswaram	13
Total	11	253

Total No. of Mandals	:	21
Total No. of Villages	:	413
Total No. of G.Ps	:	385
Total No. of Municipalities	:	07
Geographical Area	:	2,760.10 Sq.Km

Implementation Framework

1. **Data Collection:** Integrating spatial databases (maps, road networks, land use) with
2. spatial datasets (population demographics, disease rates from health centers).
3. **Access Mod Analysis:** Using tools to measure geographic coverage of healthcare services, including population capacity and travel time.
4. **Site Suitability Analysis:** Identifying optimal locations for new PHCs in, for example, hilly or coastal areas of Kakinada, using weighted overlay techniques.
5. **Decision Support System (DSS):** Building an interactive GIS system for Mandal-level officers and district collectors to make evidence-based decisions for community healthcare planning

3. Literature Review

- Previous studies on GIS in healthcare
- Machine learning applications in mortality prediction
- Research gaps (lack of integrated GIS-ML approach)

1. Machine Learning: It is a method for automating data analysis by using algorithms that iteratively identify patterns in data and learn from them. Machine learning applications are generally classified into three broad categories: (1) supervised learning, (2) unsupervised learning and (3) reinforcement learning.

2. Expert systems (Knowledge Based System): The process by which expert system is built is known as knowledge engineering. It consists of a knowledge base and a reasoning engine. It helps human to approach complex situation with high degree of uncertainty. The logic system works for improving chronic conditions with highest possible accuracy.

3. Natural Language Processing: It aims to bridge the divide between the languages that humans and computers use to operate.

4. Automated planning and scheduling: It is a relatively nascent branch of AI focused on organizing and prioritizing the activities required to achieve a desired goal.

5. Image and signal processing: It can also be used to process large amounts of data from images and signals. Steps in image and signal processing algorithms typically include signal feature analysis an

6. Health informatics and electronic medical records (EMRs): Health informatics describes the acquisition, storage, retrieval and use of healthcare information to improve patient care across interactions with the health system. Health informatics can help shape public health programs by ensuring that critical program decisions. EMRs, which are digital versions of patient and population health information, are an important source of data for health informatics.

7. Mobile health: Health uses mobile and wireless technologies to achieve health objectives. The rapid& easy availability and expansion of mobile phones in low-income countries has created several opportunities for using these technologies to support health efforts. Mobile phones have been used by community health workers to improve the provision of health services within resource-poor settings. These tools can further engage and empower the public, changing how we communicate in public health. Smartphone and mobile applications are tools that could facilitate healthy decision-making by monitoring

4.Data Collection:PHC locations, Sub-centers and Road network

Data Sources

The study uses (i) PHC/CHC/sub-centre location and attribute data from health department records, (ii) administrative boundary layers, (iii) road network layers for network-distance estimation, and (iv) a master facility database for mandal–PHC–sub-centre–habitation enumeration used in indicator computation and machine learning[13].The study uses:

- PHC location and attribute data from health department records[7].
- Road network and administrative boundary layers from government sources[14].
- Demographic layers (e.g., Census) and optional satellite/remote sensing layers for land use/environment[15].
- A master facility database for mandal–PHC–sub-centre–habitation enumeration used in statistical summaries and machine learning[16].

Master Facility Dataset Table 1 summarizes the schema of the master Excel file. The complete dataset can be provided as a supplementary file, while the paper reports only representative sample rows (Table 2) to conserve space.

Table 1: Schema of the master PHC facility datas

Field	Description
S No.	Record index.
District	District name (Guntur).
Mandal	Administrative mandal.
PHC	PHC name/code.
Sub Centers	Sub-centre under the PHC.
Village Name	Village name.
Habitations	Habitation name/identifier.
SUB-HQ	Sub-centre HQ (if applicable).
PHC HQ	PHC HQ (if applicable).
CHC-HQ	CHC HQ (if applicable)

Table 2: Sample records from the master facility datase

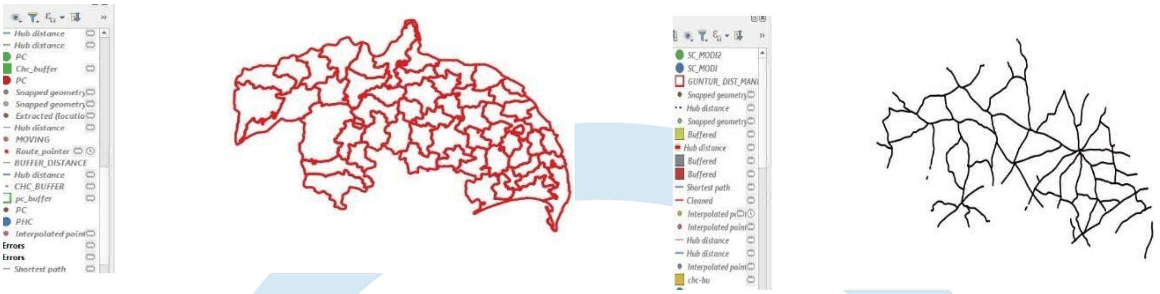
S No.	**District**	**Mandal**	**PHC**	**Sub Centers**	**Village/Habitation**
1	Kakinada District	Peddapuram	Peddapuram Urban Primary Health Centre	Kandrakota Sub Centre	Kandrakota Village
2	Kakinada District	Samalkota	Samalkota Primary Health Centre	Unduru Sub Centre	Unduru Village
3	Kakinada District	Prathipadu	Prathipadu Primary Health Centre	Gokavaram Sub Centre	Gokavaram Village
4	Kakinada District	Tuni	Tuni Primary Health Centre	Hamsavaram Sub Centre	Hamsavaram Village
5	Kakinada District	Kotananduru	Kotananduru Primary Health Centre	K.O. Mallavaram Sub Centre	K.O. Mallavaram Village

Table 2 presents representative sample records from the master facility dataset used in this study. Each row captures the hierarchical linkage across administrative and service units—District → Mandal → PHC → Sub-centre → Village/Habitation—thereby enabling consistent enumeration of coverage at multiple levels. Such structured records support mandal-wise aggregation of indicators (e.g., PHC counts, sub-centre counts, and habitation totals) and provide the base input for GIS overlays and subsequent clustering-based prioritization. Fig. 1 summarizes the core GIS datasets assembled for this study, including the mandal boundary layer used for administrative aggregation, the cleaned road network used for connectivity-aware routing, and the spatial distribution of PHC/CHC/sub-centre facilities. The map highlights how facilities are distributed across heterogeneous mandals (urban, semi-urban, and rural), providing an initial visual cue of potential spatial imbalance. These layers form the spatial reference for (i) generating service-area buffers (3 km and 5 km) around PHCs, (ii) overlaying buffers with habitation/settlement points to identify uncovered pockets, and

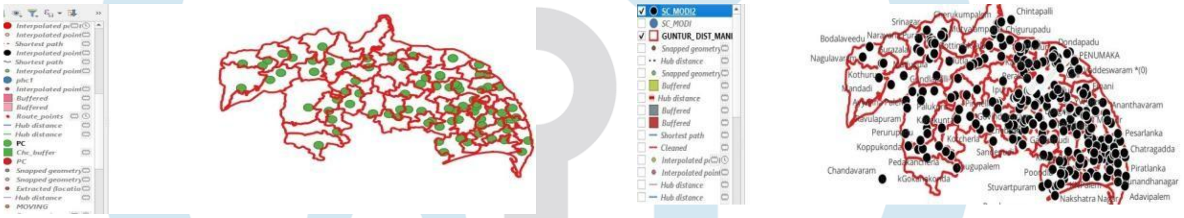
(iii) computing mandal-level indicators such as PHC counts, habitation counts, and habitations-per-PHC used in subsequent clustering and prioritization. In addition, the road network layer establishes the basis for estimating realistic travel effort (distance/travel time) along routes, which is essential for interpreting accessibility beyond straight-line proximity.

Related Work

GIS-based health facility mapping is widely used to visualize service distribution, identify geographic gaps, and communicate coverage inequities to decision makers [17].



A common first-line approach is Kakinada district boundary with mandal divisions (study area). (b) Road network layer after preprocessing/cleaning for network analysis



(c) Spatial distribution of PHC locations used for service coverage mapping.(d) Distribution of sub-centres/habitations support- ing mandal-wise indicators

However, buffer-only coverage can oversimplify access when roads, terrain, bridges, and physical barriers influence real travel effort.In rural and peri-urban settings, Euclidean distance often underestimates travel burden, and two locations at similar straight-line distance may differ substantially in travel time due to indirect connectivity and network constraints.Network-based accessibility modelling improves realism by estimating distance or travel time along road networks [18]. Beyond pure distance, floating catchment methods (e.g., 2SFCA and enhanced variants) incorporate supply demand interaction, competition effects, and distance decay to approximate more realistic service reach [19].When explicit labels for “underserved” areas are unavailable, unsupervised learning can groupregions into interpretable coverage–load profiles using facility density and workload proxies [20].For planning and public-sector deployment, interpretability is critical; transparent models such as decision trees can translate multi-indicator patterns into actionable rules that can be justified

Research Gap

Existing studies often address spatial accessibility estimation (buffer, network, or catchment-based) and data- driven prioritization separately. A practical gap remains in integrating (i) connectivity-aware accessibility estimation, (ii) unsupervised pattern discovery for identifying atypical high- burden areas, and (iii) interpretable rule extraction within one replicable district-level pipeline for PHC planning.This paper addresses the gap through an integrated GIS–ML workflow for underserved-zone identification and prioritization.

5.Methodology:

The proposed pipeline integrates (i) GIS-based service coverage and connectivity-aware accessibility estimation and (ii) unsupervised pattern discovery with interpretable rule extraction to prioritize underserved Primary Health Centre (PHC) service zones at the mandal level.

GIS-Based Spatial Analysis

Base Mapping and Preprocessing

Point layers for PHCs, Community Health Centres (CHCs), and sub-centres are compiled and aligned with the mandal boundary layer and habitation/settlement layer. All spatial layers are reprojected to a suitable projected Coordinate Reference System (CRS) to ensure accurate distance estimation. The road network is then checked for topological consistency (e.g., disconnected components, invalid geometries) and cleaned to support valid network routing[22].

Service Coverage Mapping (Buffer/Proximity)

To approximate PHC service zones, buffer regions are generated around each PHC using a practical planning radius $r \in [3, 5]$ km. The buffer layer is overlaid with the habitation/settlement layer to identify uncovered pockets, i.e., habitations that fall outside the buffer thresholds. These uncovered areas provide an initial proximity-based indication of gaps in spatial coverage [17].

Connectivity-Aware Network Accessibility

Because Euclidean proximity can underestimate real travel effort, realistic accessibility is quantified using road-network distance.For each CHC, the distance to the nearest PHC is computed along the road network using QGIS network analysis tools (e.g., nearest feature on network / shortest path)[23].Network distances are exported as an attribute field and converted from meters to kilometers:

$$d_{\mathrm{km}} = \frac{d_{\mathrm{m}}}{1000}$$

Interpretable Prioritization via Decision Tree

To obtain actionable planning rules, a shallow decision tree is trained using the mandal indicators as inputs and the high-load cluster label as the target [27]. The resulting model yields human-readable threshold rules of the form:

$\mathrm{PHC}(m) > \tau,$ otherwise,

Machine Learning Layer :**Mandal-Level Indicators**

Let m index mandals. Let P_{m} denote the number of PHCs in mandal m, S_{m} the number of sub-centres, and H_{m} the number of habitations [24]. The indicators used for downstream pattern discovery are defined as:

$$\text{PHC Count}(m) = P_{\mathrm{m}}, \tag{2}$$

$$\text{Sub Centre Count}(m) = S_{\mathrm{m}}, \tag{3}$$

$$\text{Habitation Count}(m) = H_{\mathrm{m}}, \tag{4}$$

$$\text{Habitation Per PHC}(m) = \frac{H_{\mathrm{m}}}{P_{\mathrm{m}}}. \tag{5}$$

These indicators summarize facility availability and a workload proxy (habitations-per-PHC) at the mandal level.

Standardization

Before clustering, each indicator is standardized (z-score) to avoid scale dominance

$$z = \frac{x - \mu}{\sigma}, \tag{6}$$

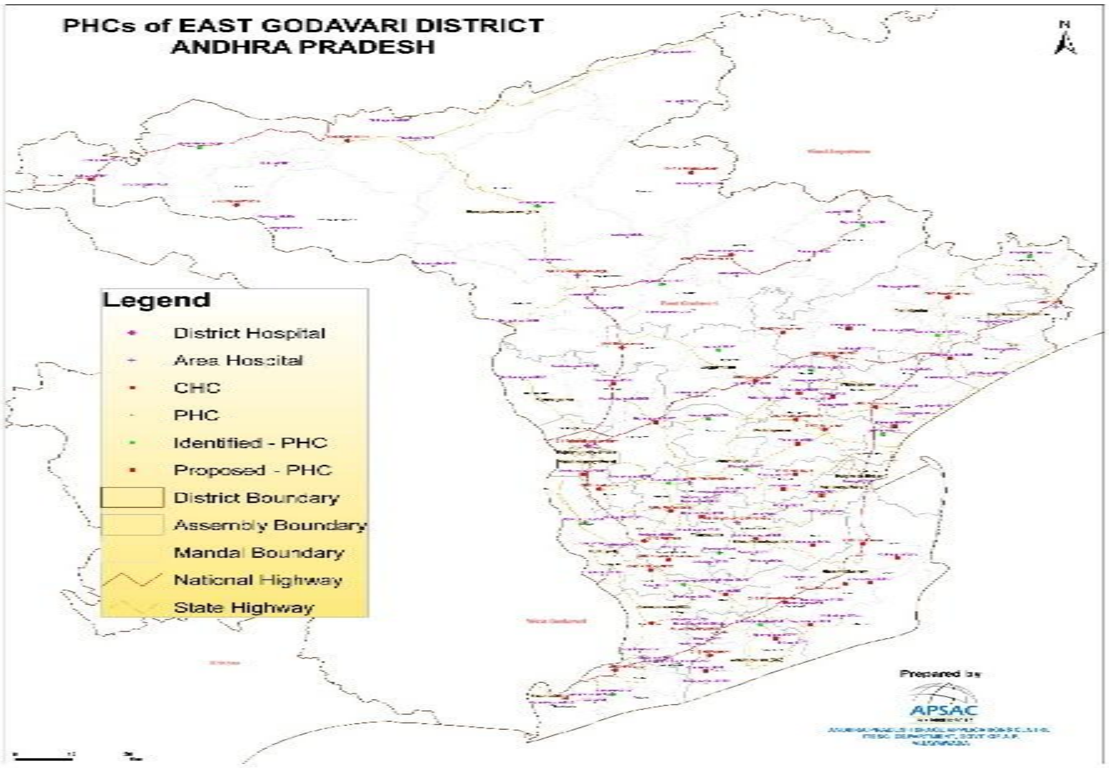
where μ and σ denote the mean and standard deviation of an indicator across mandals.

Unsupervised Clustering using K-means

Let $\mathbf{x}_i \in \mathrm{R}^d$ be the standardized feature vector for mandal i, and let K be the number of clusters [26]. K-means partitions mandals by minimizing within-cluster squared distan

Objective Function (Inertia): $J = \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_i^{(j)} - c_j \|^2$.

Centroid Update: $c_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i^{(j)}$, where n_j is the number of points in cluster j.



Implementation details and reproducibilit

All indicators (Eqs. (2)–(5)) were computed from the master facility Excel dataset (Table I). K-means was executed on z-score standardized indicators (Eq. (6)) with multiple random initializations to reduce sensitivity to centroid seeding. Cluster validity was examined using Silhouette, Davies–Bouldin, and Calinski–Harabasz indices for $K \in \{2, 3, 4, 5\}$, and the selected cluster was interpreted using the maximum mean HabitationPerPHC. A shallow decision tree (depth = 1) was used to extract an interpretable prioritization rule. The analysis scripts and derived summary tables can be shared as supplementary material to support replication

Proxy network accessibility

While road-network distance (CHC→nearest PHC) is the preferred accessibility measure, the present study version does not include a complete road-network cost table or geocoded CHC–PHC pairs.

Therefore, we report a conservative *accessibility proxy* derived from facility availability and workload pressure at the mandal level. Let *m* index mandals and let HabitationPerPHC(*m*) be defined in

Eq. (5). We compute a standardized score:

$$z_1(m) = \frac{\text{Habitation Per PHC}(m) - \mu}{} \qquad (9)$$

Table 3: Dataset coverage and workload summary

Item	Value
Total mandals (N)	53
Total PHCs	86
Total sub-centres	602
Total habitations	1509
[illegible]	17.55
[illegible]	≈ 17
[illegible]	13–23

$$z_2(m) = \frac{(1/\text{PHCCount}(m)) - \mu'}{\sigma'} \qquad (10)$$

and combine them as

$$\text{score}(m) = 0.75\, z1(m) + 0.25\, z2(m). \qquad (11)$$

Finally, a bounded proxy distance is obtained using a logistic mapping:

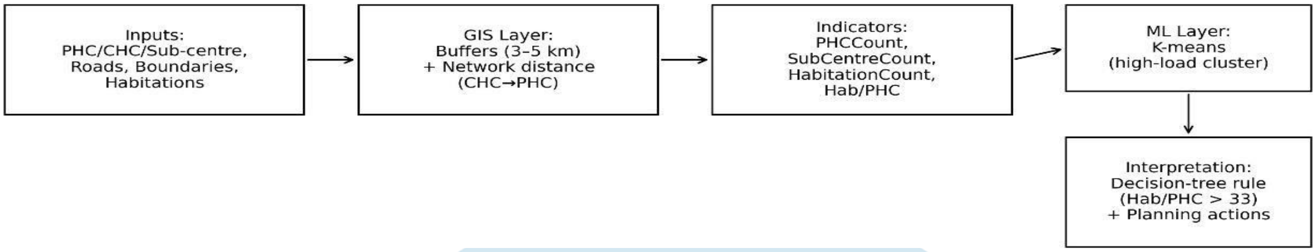
$$d\text{proxy}(m) = 2 + 8 \cdot \frac{}{1 + \exp(-\text{score}(m))}, \quad d\text{proxy} \in [2, 10] \text{ km}. \qquad (12)$$

This proxy is used only to summarize relative accessibility constraints and will be replaced by true road- network distances when network-cost outputs are available

Integration and Analysis

GIS–ML workflow adopted in this study, integrating spatial coverage analysis with data-driven prioritization for PHC planning. The pipeline begins with data acquisition and preprocessing, where facility layers (PHC/CHC/sub-centre), administrative boundaries, habitations, and the road network are cleaned, validated, and aligned to a common projectedCRS. Next, GIS-based coverage mapping is performed by generating PHC service buffers

Figure 2: Overview of the integrated GIS–ML workflow.



5 km) and intersecting them with habitations to quantify covered and uncovered pockets. In parallel, connectivity-aware accessibility is estimated using CHC→nearest PHC routing on the road network (or a proxy accessibility indicator when full network-cost outputs are unavailable).

The derived GIS outputs are then aggregated into mandal-level indicators (e.g., PHC count, habitation count, habitations-per-PHC, coverage ratios, and accessibility summaries) to create a structured feature matrix for analytics. Unsupervised clustering (K-means) is applied to discover distinct mandal profiles that capture typical coverage patterns versus high-load or underserved configurations.The cluster exhibiting the highest workload proxy (e.g., maximum mean habitations-per-PHC) is designated as the priority/high-load group for planning focus. To ensure transparency, an interpretable model (shallow decision tree) is trained to explain the priority cluster membership using simple threshold splits on the indicators. Finally, the pipeline outputs actionable products for stakeholders, including coverage maps, accessibility summaries, priority mandal lists, and human-readable decision rules that can support resource allocation and targeted interventions.

Fig. 3 visualizes PHC service zones under a 3 km buffer assumption, highlighting habitations that fall outside the recommended service radius. The uncovered pockets in this map indicate potential access gaps that warrant targeted outreach or facility strengthening. Fig. 4 shows the corresponding coverage when the service radius is expanded to 5 km, resulting in visibly fewer uncovered habitations compared . This comparison provides a simple sensitivity view of how coverage changes under different service-distance assumptions

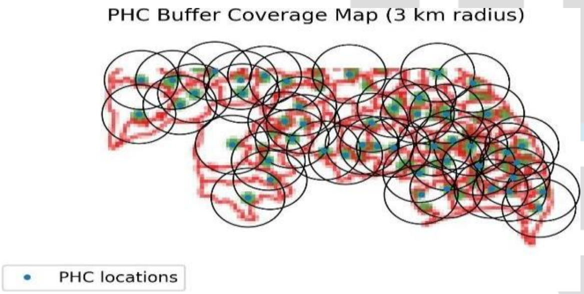


Figure 3: PHC service zones using 3 km buffers with uncovered habitations highlighted.

Figure 4: PHC service zones using 5 km buffers with uncovered habitations highlighted.

Connectivity-Aware Accessibility Results

Road-network distance (CHC → nearest PHC) is used to capture realistic accessibility constraints.

Let d_j denote the network distance for CHC j (in km). We report:

$$\bar{d} = \frac{1}{M}\sum_{j=1}^{M} d_j, \qquad (14)$$

$$d_{50} = \text{median}(d_j), \; d_{90} = \text{quantile}_{0.9}(d_j), \qquad (15)$$

$$\text{PctAbove}_T = \frac{1}{M}\sum_{j=1}^{M} I(d_j > T) \times 100. \qquad (16)$$

Table 4: Network accessibility summary (proxy estimated from Excel-derived mandal indicators; $T = 5$ km).

Metric	$\bar{d}$ (km)	d_{50} (km)	d_{90} (km)	PctAbove_T (%)
All CHCs (proxy)	5.96	5.90	8.02	66.04
Priority mandals (only)	8.52	8.96	9.33	100.00

$$\rho(m) = \frac{\text{HabitationPerPHC}(m)}{\overline{\text{HabitationPerPHC}}}, \qquad (17)$$

where $\overline{\text{HabitationPerPHC}} \approx 17.55$.

The Map : Connectivity-aware accessibility: CHC-to-nearest-PHC network distance (km) rendered as a graduated map. Fig. illustrates spatial variation in accessibility by mapping the CHC-to-nearest-PHC distance as a graduated thematic surface. The visualization highlights mandals where referral movement from CHCs to PHCs would require comparatively longer travel, indicating potential bottlenecks in routine referral pathways. Higher mapped values represent greater connectivity constraints (e.g., indirect routes or weaker network connectivity) that may delay timely access to primary care services. Such locations can be prioritized for strengthening referral coordination, ambulance routing, and outreach scheduling, particularly during peak demand or emergencies. Overall, the map provides

a spatial decision-support view for identifying areas where improvements in routing efficiency or facility distribution could yield the largest accessibility gains.

Primary health centres and Mandals from K-means and Workload Interpretation

K-means separates a small high-load group from typical-coverage mandals and identifies the following

It compares the habitations-per-PHC ratio for the priority mandals identified by K-means clustering, highlighting workload pressure where each PHC serves a larger number of habitations. The plot makes the relative severity explicit by showing which mandals lie far above the district typical range, indicating potential service strain and longer effective catchments. The horizontal decision threshold provides a transparent cutoff for classifying mandals as high-load versus typical, enabling consistent prioritization across administrative units. Mandals above the threshold can be flagged for targeted strengthening actions such as adding sub-centres/PHCs, reallocating staff, or improving referral support. Overall, the figure translates clustering outcomes into an operational planning view that is easy to communicate to stakeholders

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i),\; b(i)\}}, \qquad S = \frac{1}{N}\sum_{i=1}^{N} s(i). \qquad (18)$$

Also report Davies–Bouldin index (DBI) and Calinski–Harabasz (CH) score for $K \in \{2, 3, 4, 5\}$.

Table 6: Cluster validity metrics for selecting K (computed from mandal-level indicators).

K	**Silhouette** S	**DBI (lower better)**	**CH (higher better)**
2	0.784	0.141	36.30
3	0.361	0.763	42.04
4	0.435	0.630	64.28
5	0.436	0.665	67.48

Fig. projects the standardized mandal indicator space onto the first two principal components to provide a compact visualization of the K-means results. Each point corresponds to a mandal, and spatial separation among point groups indicates that mandals form distinct coverage–load profiles under the chosen indicators. Clusters that appear isolated in the PCA plane typically represent atypical combinations of workload and facility availability, aligning with the notion of underservedor high-burden zones. This visual evidence complements the quantitative cluster validity metrics and supports the use of cluster membership for prioritization decisions. Overall, the figure improves interpretability by showing how the selected indicators differentiate mandals into planning-relevant groups. Cluster Validity Justification

To justify the selected K, report internal cluster validity metrics. For each mandal i, let $a(i)$ be the mean intra-cluster distance and $b(i)$ the smallest mean distance to other clusters. The silhouette coefficient is:

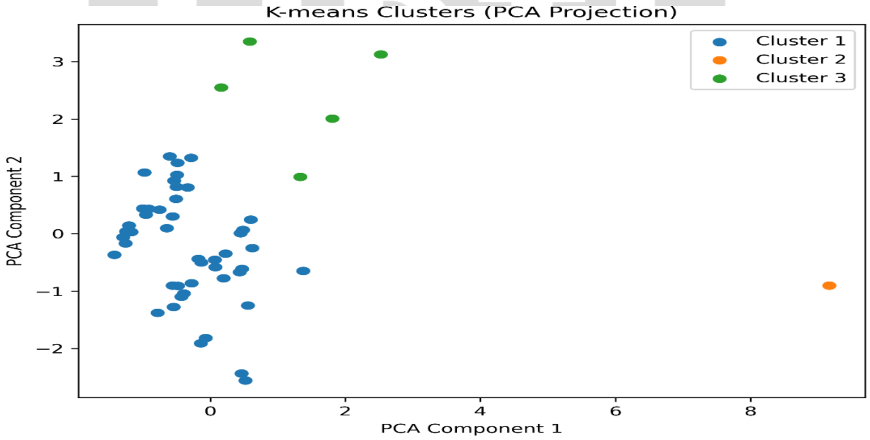


Figure Visualization of K-means clusters in 2D after PCA projection of standardized indicators

Fig. 8 presents a shallow decision tree trained to explain membership in the high-load (priority) and avoids overfitting, making it suitable for operational use in district planning. Planners can apply these thresholds directly to classify mandals and to communicate the rationale for prioritization in an auditable manner. Overall, the figure bridges data-driven discovery and policy-facing decision support through interpretable criteria.

Interpretable Prioritization Rule and Discussion

A shallow decision-tree yields a transparent operational rule:

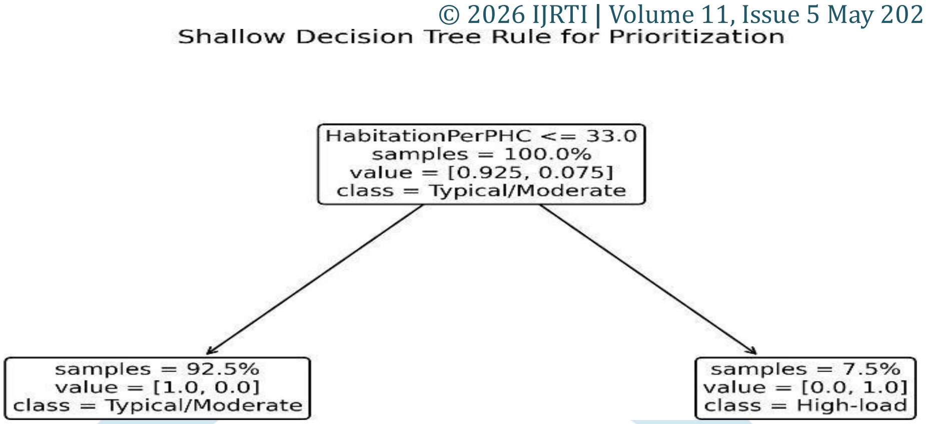
$$\text{HighLoad}(m) = \begin{cases} & \text{if } \text{HabitationPerPHC}(m) > 33, \\ 0 & \text{otherwise.} \end{cases} \qquad (19)$$

Recommended clarity for reviewers: since Repalle (27.5) is below 33, present it as a borderline case or use a two-level rule:

$$\text{Priority}(m) = \begin{cases} \text{High}, & \text{HabitationPerPHC}(m) > 33, \\ \text{Moderate}, & 25 \le \text{HabitationPerPHC}(m) \le 33, \\ \text{Typical}, & \text{HabitationPerPHC}(25. \end{cases}$$

Planning Interpretation and Implications

Network-based distances highlight accessibility constraints often underestimated by straight-line proximity. The identified priority mandals can be targeted for (i) strengthening coverage via additional PHCs/sub-centres and (ii) improving connectivity through critical road-link upgrades



Key Strategic Applications in Rural Kakinada

- **Spatial Accessibility Analysis (GIS):** Utilizing GIS to map and model the physical distance and travel time from rural hamlets to PHCs. This identifies underserved "spatial discrepancy" zones where new health centers or mobile units are needed.
- **Predictive Disease Modeling (ML):** Applying algorithms to analyze historical disease trends and environmental data (e.g., rainfall, temperature) to anticipate malaria, dengue, or diarrhea outbreaks before they peak.
- **Optimizing Resource Allocation (GIS/ML):** Using ML-based demand forecasting (based on age, chronic diseases, and economic factors) to identify where to deploy mobile health vans or optimize PHC operating hours.
- **Reducing Mortality (Integrated Approach):** Lowering Maternal Mortality Rate (MMR) and Infant Mortality Rate (IMR) by identifying high-risk pregnancies and newborns in remote locations for targeted home visits.

Implementation Framework

1. **Data Collection:** Integrating spatial databases (maps, road networks, land use) with non-spatial datasets (population demographics, disease rates from health centers).
2. **AccessMod Analysis:** Using tools to measure geographic coverage of healthcare services, including population capacity and travel time.
3. **Site Suitability Analysis:** Identifying optimal locations for new PHCs in, for example, hilly or coastal areas of Kakinada, using weighted overlay techniques.
4. **Decision Support System (DSS):** Building an interactive GIS system for Mandal-level officers and district collectors to make evidence-based decisions for community healthcare planning.

Expected Outcomes

- **Increased Access Population Coverage (APC):** Bridging the gap where sometimes over 20% of villages are crowded out of effective PHC services due to population pressure.
- **Improved Efficiency:** Reducing travel time for patients, particularly critical for antenatal care and emergency services.
- **Sustainable Development:** Empowering rural communities through proactive disease surveillance and tailored health interventions. [1, 2, 3, 4, 5]

This approach aligns with the national goal of strengthening the Primary Health System (PHC) to meet the Indian Public Health Standards (IPHS). [1]

- Community & Family Medicine - AIIMS Bhubaneswar

1. Conclusion

paper an integrated GIS–ML framework to identify underserved Primary Health Centre (PHC) service zones in kakinada District, Andhra Pradesh. The GIS layer quantified proximity-based coverage using 3 km and 5 km PHC buffers and highlighted uncovered habitations as spatial service gaps. To reflect connectivity constraints, the workflow supports CHC→nearest PHC network-distance estimation; however, in the present study version, accessibility is summarized us-3 ing a conservative indicator-derived proxy, which can be replaced by true network-cost outputs when available.Using the compiled master dataset (As of April 2026, the Kakinada district in Andhra Pradesh features an extensive public health network, including 121 rural Primary Health Centres (PHCs), 41 urban PHCs, 25 Community Health Centres (CHCs), and 3 area hospitals the analysis shows a district-level workload of approximately 17.55 habitations per PHC.K-means clustering separated a small high-load group from typical-coverage mandals and identified priority mandals with markedly higher habitations-per-PHC, including . To make the output operational for district planning, a shallow decision tree produced an interpretable prioritization rule (e.g., a habitations-per-PHC threshold), enabling transparent classification of mandals into high-load and typical categories.The AI tools and techniques are still in their infancy stage. Despite the limitations, these tools and techniques are beneficial in providing in depth knowledge on individuals' health and predicting population health risks, and their use for medicine as well as public health is likely to increase substantially in the near future. Privacy, confidentiality, data security, ownership and informed consent have to be maintained in the human right lens. Effective implementation will also require understanding the local social, epidemiolocal, health system and political contexts. So in long run the correct & efficient use of AI technologies will definitely help in achieving the health related targatingMedical Geographic Information Systems (GIS) and predictive machine learning (ML) offers a powerful, data-driven approach to enhance primary healthcare (PHC) access and reduce mortality in rural Kakinada, Andhra Pradesh. By analyzing spatial data—such as road networks, population density, and existing PHC locations—alongside predictive models for disease outbreaks and patient risk, health authorities can optimize resource allocation and enable proactive, timely care, reducing the 5.49 km average village-to-PHC distance barrier.

7. References

1. Jin C, Cheng J, Lu Y, Huang Z, and Cao F. Spatial inequity in access to healthcare facilities at a county level in a developing country: a case study of Deqing County, Zhejiang, China. International journal for equity in health 2015;14:67.

2. Verma VR and Dash U. Geographical accessibility and spatial coverage modelling of public health care network in rural and remote India. PloS one 2020;15:e0239326.

3. Biu PW, Nwasike CN, Nwaobia NK, Ezeigweneme CA, and Gidiagba JO. GIS in healthcare facility planning and management: A review. World J Adv Res Rev 2024;21:12–19.

4. Walsh SJ, Page PH, and Gesler WM. Normative models and healthcare planning: network-based simulations within a geographic information system environment. Health services re-search 1997;32:243.

5. Mittal S and Ukkusuri SV. Overcoming data scarcity in transit planning: A novel framework combining machine learning and metaheuristics. Data Science for Transportation 2025;7:4.

6. Rani NA, Bal S, and Inayathulla M. GIS Applications and Machine Learning Approaches in Civil Engineering. Recent Advances in Civil Engineering for Sustainable Communities: Select Proceeding of IACESD 2023 2024;459:157.

7. Saravanabavan V, Sangma NN, and Vinothini C. Spatial distribution and pattern of location in primary health centers in West Garo Hills District, Meghalaya. International Journal of Research Publication and Reviews 2024;5:694–701.

8. Pozo AMM, Priesmeier P, and Fekete A. Measuring spatial accessibility to critical infrastructure: The Access Road Identification model. International Journal of Critical Infrastructure Protection 2025;49:100760.

9. Ratcliffe HL, Schwarz D, Hirschhorn LR, et al. PHC Progression Model: a novel mixed-methods tool for measuring primary health care system capacity. BMJ global health 2019;4.

10. Ukkund SJ, Amrutha L, Afra A, et al. AI for Remote Healthcare and Telemedicine. In: *AI and Data Engineering for Healthcare*. Chapman and Hall/CRC:242–81.

11. Bella E di, Gandullia L, Leporatti L, Montefiori M, and Orcamo P. Ranking and prioritization of emergency departments based on multi-indicator systems. Social Indicators Research 2018;136:1089–107.

12. Chandran A and Roy P. Applications of geographical information system and spatial analysis in Indian health research: a systematic review. BMC Health Services Research 2024;24:1448.

13. Mocanu T, Joshi J, and Winkler C. A data-driven analysis of the potential of public transport for German commuters using accessibility indicators. European transport research review 2021;13:54.

14. Tatababu C, Padma G, and Rao PVGP. A Geospatial approach for spatial accessibility and connectivity of double lane roads between mandal and district headquarters in coastal Andhra Pradesh. International Journal of Modern Sciences and Engineering Technology (IJMSET) 2015;2.

2. Future Work

- Integration with real-time mobility data
- Use of Deep Learning models
- Deployment in other districts