

Evaluating Health Facility Distribution and District-Level Disease Prevalence in Rwanda: A Data-driven Approach

Brandone Fonya, Irene Busah, Michaella Rugumbira, Nchofon Tagha
bfonya, ibusah, mrugumbi, ntaghagh
Carnegie Mellon University Africa.

Abstract—Access to healthcare is a fundamental right, yet disparities in healthcare facility distribution persist in Rwanda. Existing allocation strategies predominantly rely on population density or administrative boundaries, often overlooking actual disease burden—particularly for high-morbidity illnesses such as Malaria, Tuberculosis (TB), and HIV. This study addresses that gap by combining geospatial analysis with machine learning to assess how well healthcare facilities align with regional epidemiological needs. Using national survey data, health facility locations, and demographic statistics, we estimate district-level disease burden and evaluate its spatial correspondence with existing infrastructure. To move beyond descriptive mapping, we developed predictive models—trained entirely on synthetic data—to simulate disease prevalence, population, and mortality for classifying districts based on the urgency of additional healthcare facility allocation. We explored Neural Networks, Random Forest, and an XGBoost–Random Forest ensemble, with the ensemble model achieving the highest accuracy at 96%. To ensure interpretability, we applied SHAP analysis, which identified malaria incidence as the most influential predictor of healthcare demand, followed by HIV prevalence and population density. Our findings offer actionable insights for improving healthcare equity and demonstrate the value of data-driven methods for resource planning in resource-constrained settings like Rwanda.

Index Terms—Healthcare facility distribution, disease prevalence, resource allocation, and geospatial analysis.

I. INTRODUCTION

A. Literature Review

Efficiently allocating healthcare resources is crucial for ensuring equitable access to medical care and improving public health outcomes[1]. Traditionally, healthcare resource distribution, like in Rwanda, has been focused on administrative divisions and predominantly on population density rather than the prevalence of major illnesses in different regions[2]. This approach, however, fails to account for the variation in disease burden across different geographical locations. This leads to inefficiencies in healthcare delivery. Some regions may have a higher prevalence of a certain major illness and require more targeted healthcare facilities. The inability to match healthcare resources with actual disease prevalence results in under-served populations and increased patient travel time to access these facilities, especially in the event of an outbreak [3].

Advancements in data-driven decision-making, geospatial analysis, and machine learning now enable a more precise

approach to healthcare planning[4],[5]. Over the years, various studies have explored different approaches to optimize healthcare resource allocation. One common method involves calculating Euclidean distance to the nearest healthcare facility to assess accessibility and its impact on health outcomes. For example, a study in Ethiopia found that increased distance to healthcare facilities significantly reduced maternal healthcare utilization and increased neonatal mortality rates, underscoring the role of geographic accessibility in healthcare efficiency [6]. Another approach involves GIS-based spatial accessibility analysis, which evaluates healthcare service availability in rural regions by integrating population density and healthcare infrastructure data[7]. This paper explores the direct relationship between disease burden and healthcare facility placement.

By integrating disease prevalence data into healthcare resource distribution models, policymakers can make more data-informed decisions to optimize the allocation of medical specialists and healthcare facilities. While Rwanda’s Workload Indicators of Staffing Need (WISN) report provides critical insights into staffing pressures, it does not fully incorporate variations in disease prevalence, thereby necessitating an improved approach[2].

Motivated by all these, this research aims to bridge the gap between healthcare demand and healthcare facilities allocation by examining the relationship between district-level disease prevalence and distribution in Rwanda to identify mismatches in healthcare resource allocation. This research focuses on the top three causes of death in Rwanda, identified to be tuberculosis, malaria, and HIV[2]. By developing a data-driven framework, this study will contribute to a more equitable distribution of healthcare, ensuring that the regions with the highest disease burden receive adequate medical resources. The findings will support Rwanda’s transition to a more efficient healthcare system and provide a scalable model for other countries facing similar challenges.

B. Research Questions

To have a better understanding of the context for coordinated action, these questions were asked:

- 1) What is the distribution of population and health facilities in Rwanda?
- 2) How are the main causes of death (Tuberculosis, Malaria) distributed throughout Rwanda?

- 3) What is the correlation between Rwanda's disease prevalence and resource allocation?
- 4) How can healthcare resource allocation be optimized to better address local needs?

II. DATA SOURCES

This study uses multiple publicly available data sources to analyze the distribution of healthcare facilities, and Rwanda's main disease burden.

- 1) Rwanda Demographic and Health Survey (DHS): These are publicly available data provided by the Rwandan government that holds information on health indicators, disease prevalence, and demographic characteristics.
- 2) Fifth Population and Housing Census (2022): Publicly available data providing population density statistics for all 30 districts in Rwanda.
- 3) OpenStreetMap data on healthcare facilities in Rwanda: Contains precise locations of hospitals, health centers, clinics, and pharmacies.
- 4) The Rwanda GeoPortal: A GeoJSON file delineating the boundaries of Rwanda, its provinces, and districts was downloaded from this portal.

In addition to these data sources, we had informal surveys through conversations to enhance our understanding of the ground truth distribution of health facilities across Rwanda.

III. METHODOLOGY

The data were extracted from the listed sources, processed, and analyzed to draw insights as follows:

A. Data Extraction and Preparation

The primary dataset utilized in this study was sourced from the Rwanda DHS. The population data was available at the district level, providing a granular representation of demographic distribution. This data was manually extracted and structured into a Comma-Separated Values (CSV) file to facilitate computational analysis. Utilizing this dataset, the proportion of each district's population within its respective province was determined by computing the ratio of the district population to the total provincial population. This calculation serves a critical role in subsequent analyses, particularly in estimating district-level disease burden with improved accuracy and regional specificity.

Data on Tuberculosis and Malaria was similarly obtained from the Rwanda Demographic and Health Survey (DHS) and structured into a Comma-Separated Values (CSV) file for systematic analysis. At this stage, data pertaining to HIV remained under review. It is important to note that the DHS provided sample-based data, with granularity at the provincial level rather than at the district level. This proportional allocation is a necessary assumption to simulate district-level disease burden in the absence of more granular data, although it may not fully capture localized variations in disease incidence. The Tuberculosis and Malaria data were categorized based on multiple demographic and socioeconomic factors, including age group, gender, province, education level, and income level.

However, for this study, we focused on the aggregated data at the provincial level.

To estimate the number of Tuberculosis and Malaria cases in a given region—defined here as the population exhibiting symptoms of the disease (referred to as positive cases)—the prevalence ratio (i.e., the proportion of positive cases within the total surveyed sample) was applied to the total provincial population. Subsequently, this estimate was further refined to the district level using the previously computed district-to-province population ratio, ensuring a more precise distribution of disease burden across districts. The population and disease density was obtained by dividing the toll by the area of the district obtained from the already prepared shape file. Also, the facility location file from OpenStreetMap required no further processing.

To estimate the total number of positive cases for Tuberculosis and Malaria in each province, the prevalence ratio derived from the DHS sample was applied to the total provincial population:

$$C_p = P_r \times N_p \quad (1)$$

where:

- i. C_p is the estimated number of positive cases in the province,
- ii. P_r is the prevalence ratio:

$$\left(\frac{\text{number of positive cases in sample}}{\text{total sample size}} \right)$$

- iii. N_p is the total population of the province.

To distribute the estimated cases from the provincial to the district level, the district share of the provincial population was used:

$$C_d = C_p \times \frac{N_d}{N_p} \quad (2)$$

where:

- i. C_d is the estimated number of positive cases in the district,
- ii. N_d is the total population of the district.

In simpler terms, we assumed that disease cases are distributed across districts in proportion to their share of the provincial population. For example, if a district represents 20% of its province's population, we estimate that it also accounts for 20% of the province's disease cases. This proportional allocation allows us to estimate district-level disease burden even when data is only available at the province level.

The healthcare facility dataset includes keywords[8] that are used to categorize the facilities. Of those, the following facilities met our inclusion criteria - facilities that can offer some form of help to positive cases:

- i. Alternative: Practitioner or place for alternative/complementary medicine outside standard categories.
- ii. Centre: Healthcare facility with unspecified characteristics.
- iii. Clinic: May be used in addition to amenities.
- iv. Doctors: May be used in addition to amenities.

- v. Hospitals: May be used in addition to amenities.
- vi. Health Post: Health centre.
- vii. Pharmacy: A source of medication.
- viii. Laboratory: Medical lab analyzing body fluids (blood, urine, faeces, etc.), also collects samples.
- ix. Yes: General healthcare facility without known specifics.

B. Data Analysis

The district-level population toll heatmap was plotted with the population density heatmap as seen in Fig. 1 and 2. This visualization offers a more accurate representation of the population distribution across Rwanda. The spatial distribution of healthcare facilities across Rwanda is illustrated in Fig. 3. Similarly, the heatmap of the prevalence and density of the disease in this study was plotted side-by-side. Using the filtered healthcare keywords, the health facility distribution by type across Rwanda was plotted.

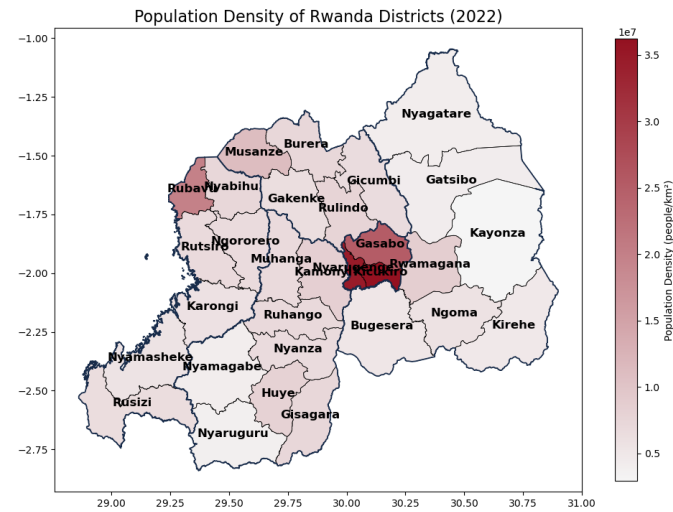


Fig. 1. Population Density Across Rwanda.

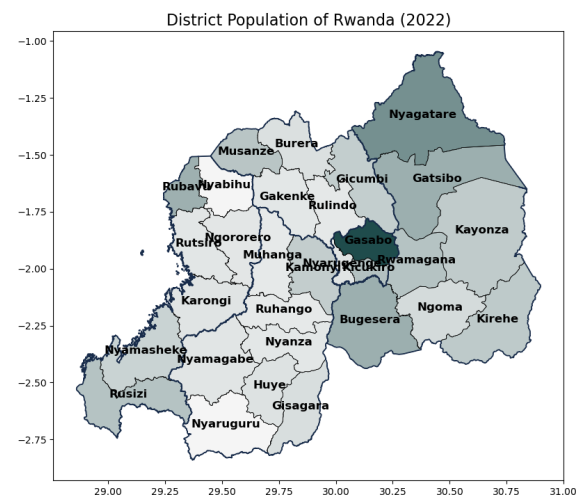


Fig. 2. Population Toll Across Rwanda.

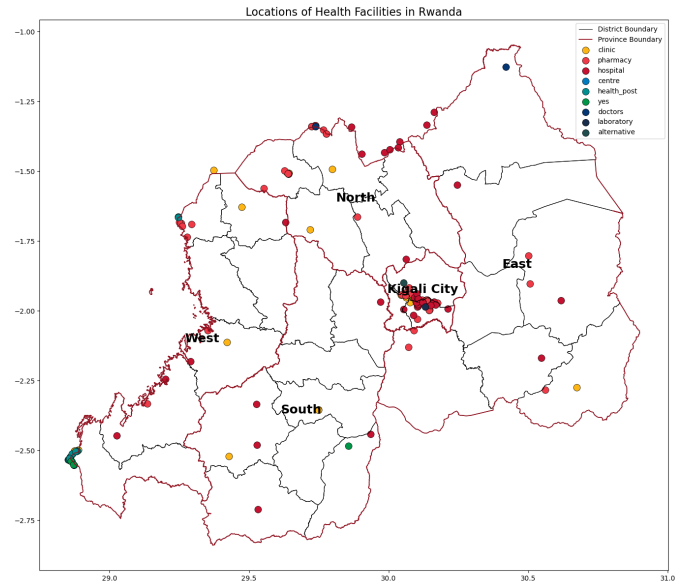


Fig. 3. Facility Distribution by Type Across Rwanda.

To comprehensively assess the spatial distribution and burden of Tuberculosis and Malaria across Rwanda, disease prevalence maps and density heatmaps were generated. These visualizations provided insights into the geographic variations in disease incidence and helped identify high-burden regions, as seen in Fig. 4, 5, 6, and 7. Additionally, a normalized comparative analysis was conducted by plotting disease prevalence against healthcare facility distribution.

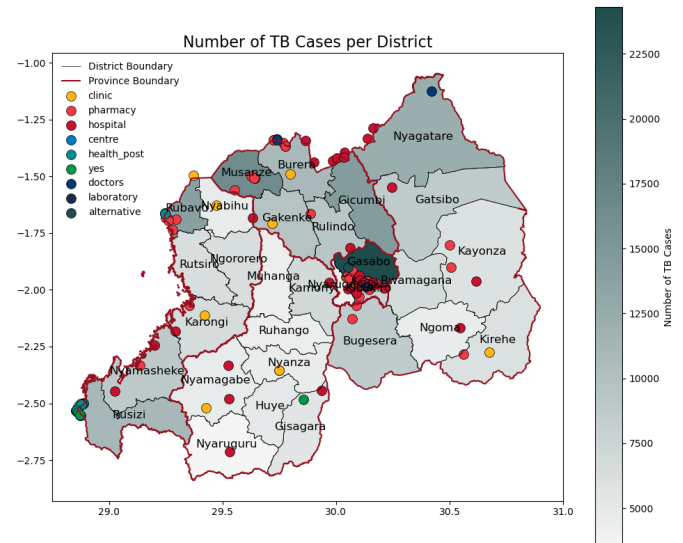


Fig. 4. Tuberculosis Toll Across Rwanda.

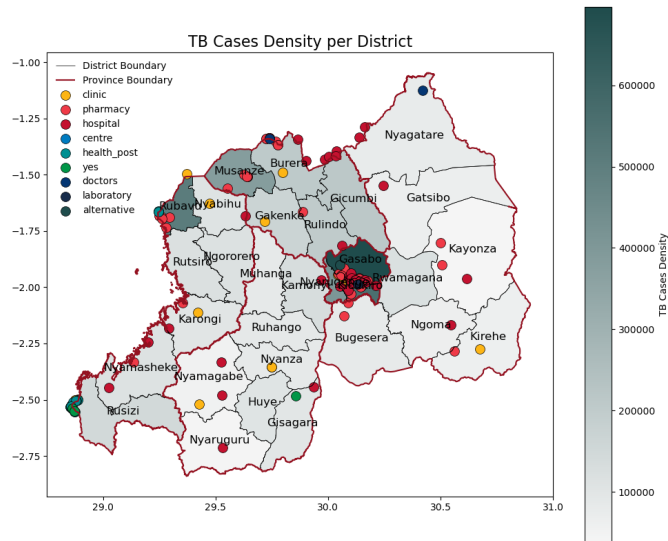


Fig. 5. Tuberculosis Density Across Rwanda.

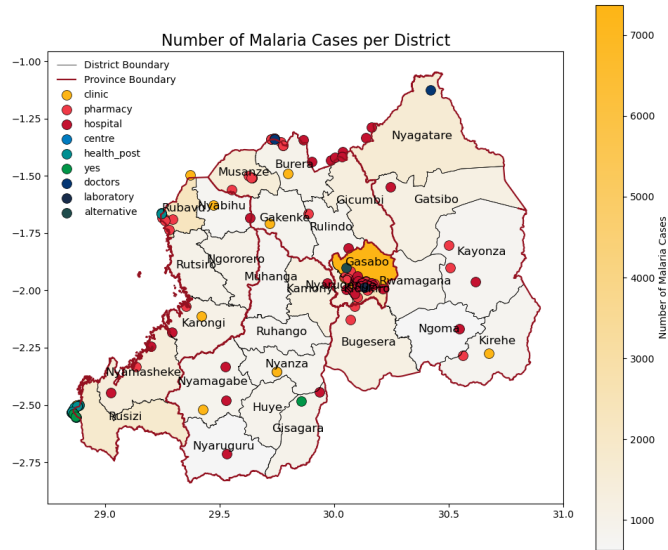


Fig. 6. Malaria Toll Across Rwanda.

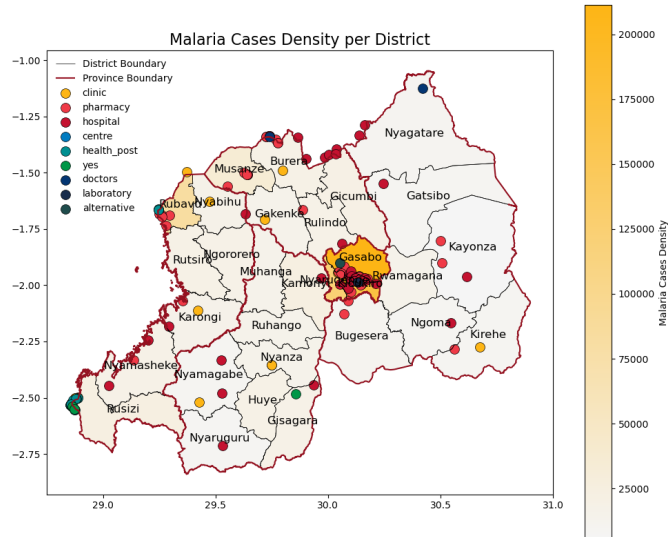


Fig. 7. Malaria Density Across Rwanda

Correlation analyses were performed to quantify the relationship between disease burden and healthcare accessibility, both in terms of absolute case counts and population density as seen in Fig. 8 and 9. The findings from these analyses serve as the basis for the discussions and interpretations presented in the subsequent section.

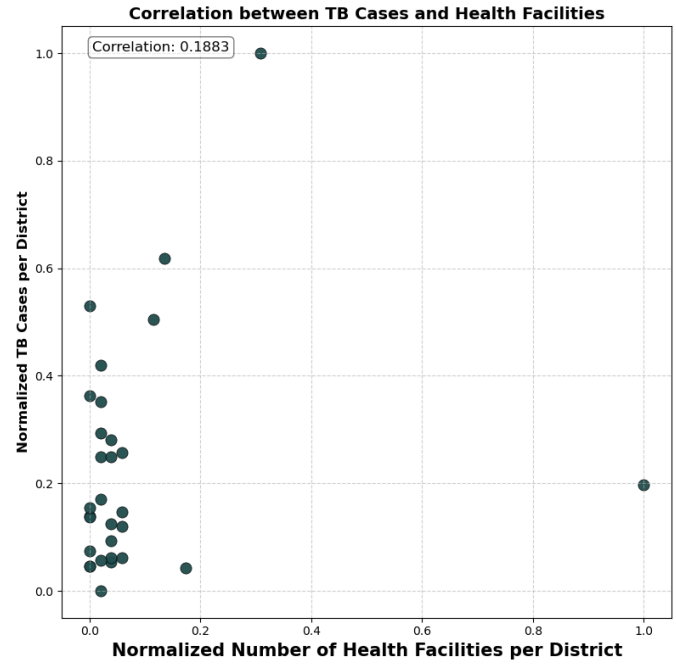


Fig. 8. Correlation in TB Burden and Facility Distribution

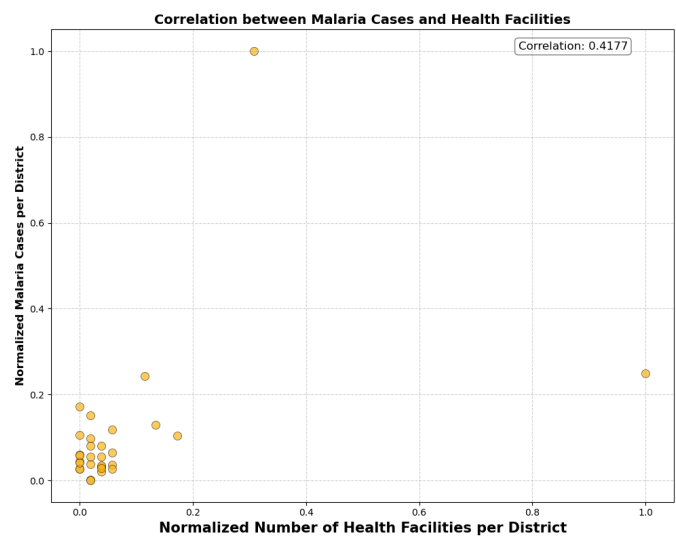


Fig. 9. Correlation in ML Burden and Facility Distribution.

C. Machine Learning

To complement the spatial and statistical analysis, a supervised machine learning pipeline was implemented to classify Rwandan districts into healthcare priority levels: Low,

Medium, and High—based on disease burden and demographic indicators.

Due to limited district-level data, a fully synthetic dataset was generated to simulate disease burden and healthcare demand for proof-of-concept modeling.

We synthesized the dataset using mortality statistics from a World Bank report covering disease-related deaths between 1990 and 2022. To enhance the representativeness of the data, we augmented the sample using a generative large language model (LLM).

1) *Feature Engineering*: A structured feature matrix was created for each district using epidemiological and demographic variables. These included: population size, number of healthcare facilities, estimated cases of Malaria, Tuberculosis, and HIV, reported mortality, and the year of data collection. All features were normalized to ensure comparability and to improve model training performance.

2) *Model Selection and Training*: Three classification algorithms were considered: a neural network, a random forest classifier, and an ensemble model combining XGBoost and Random Forest. The dataset was split into training and testing sets using an 80/20 ratio with stratified sampling to maintain class balance. Models were trained using cross-validation, and performance was evaluated using standard classification metrics including accuracy, precision, recall, and F1-score.

3) *Model Interpretability*: To ensure model transparency and support policy relevance, two interpretability techniques were employed. First, feature importance was computed using F-scores from tree-based models to assess the contribution of each variable to model predictions. Second, SHAP values were used to provide both global and local interpretability of model outputs, highlighting the direction and strength of each feature's influence on district-level healthcare demand predictions.

4) *Healthcare Priority Mapping*: The predicted priority levels were associated with district-level geospatial data to generate a healthcare priority map. This visualization aids in translating model outputs into actionable insights for policymakers, enabling more targeted resource allocation and planning.

IV. RESULTS

A. Population and Healthcare Facility Distribution

The population distribution across Rwanda exhibits significant regional variation. The districts within Kigali are characterized by high population densities, whereas Nyagatare, despite its large population, is more spatially dispersed. In contrast, the Western and Southern districts, particularly Nyamasheke, Rutsiro, and Gisagara, have relatively low population densities.

Healthcare facility distribution is highly concentrated within Kigali, with many rural districts having either sparse or low-capacity healthcare resources. The correlation between population density and facility distribution is weak, with a computed correlation coefficient of 0.28, indicating a misalignment between population needs and healthcare infrastructure.

B. Tuberculosis Burden and Facility Distribution

The highest Tuberculosis burden is observed in districts such as Gasabo, Musanze, and Gicumbi. However, except for Gasabo, most high-burden districts lack adequate healthcare facilities to meet the demand. The correlation between Tuberculosis prevalence and facility distribution is negligible, with a normalized correlation coefficient of 0.18. Notably, many districts either lack specialized Tuberculosis treatment facilities or have only low-level healthcare resources, exacerbating accessibility challenges.

C. Malaria Burden and Facility Distribution

The relationship between Malaria prevalence and healthcare facility distribution is generally weak but varies across regions. A normalized correlation analysis yielded a coefficient of 0.18, suggesting minimal alignment between disease burden and facility locations in many areas. However, a broader assessment of disease burden and facility distribution resulted in a slightly higher correlation of 0.42, indicating moderate alignment in certain regions. However, these results can appear deceptive due to outlier districts like Kigali. Despite this, numerous districts still lack adequate healthcare infrastructure, limiting effective disease management and intervention efforts.

D. Machine Learning Modeling

While the spatial analysis provided insights into how healthcare facilities and disease burden are distributed across Rwanda, it did not quantify how well these factors predict actual healthcare demand. To address this, we incorporated a machine learning approach.

This section presents the experimental results from the healthcare facility optimization using machine learning. It involved predicting healthcare demand and classifying district-level healthcare priority in Rwanda based on disease prevalence and demographic indicators, highlighting districts most in need of additional healthcare facilities.

1) *Model Performance and Accuracy*: Three models were evaluated: a Neural Network, a Random Forest classifier, and an ensemble model combining XGBoost with Random Forest. As shown in Figure 10, the ensemble model significantly outperformed the others, achieving an accuracy of **96%**, compared to 92% for the Neural Network and 89% for Random Forest. This suggests that ensemble methods are highly effective in capturing complex patterns in healthcare-related datasets with multiple influencing factors such as disease cases, deaths, and facility distribution.

Model	Accuracy
Neural network	0.92
Random Forest (RF)	0.89
Ensemble (XGBoost + RF)	0.96

Fig. 10. Model accuracy comparison between Neural Network, Random Forest, and an Ensemble of XGBoost and Random Forest

2) *Confusion Matrix Analysis:* The confusion matrix in Figure 11 shows the performance of the classifier in assigning districts to the correct healthcare priority categories: *Low*, *Medium*, and *High*. The classifier correctly identified 48 out of 50 low-priority districts, 46 out of 48 medium-priority districts, and 89 out of 100 high-priority districts. Misclassifications mostly occurred between neighboring classes (e.g., high misclassified as medium), indicating good overall sensitivity and minimal severe misclassification errors.

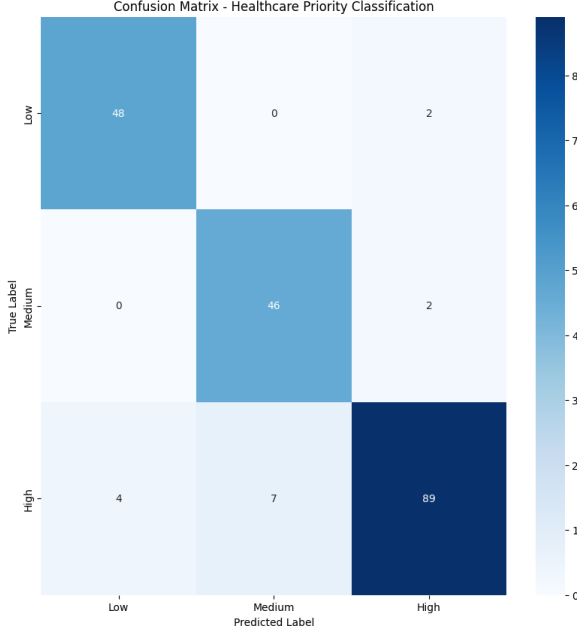


Fig. 11. Confusion Matrix for the Healthcare Priority Classification Task

3) *Feature Importance and Interpretability:* Figure 12 highlights the importance of various features in predicting healthcare demand using Ensemble. Feature f_3 (**Malaria Cases**) had the highest influence with an F-score of 1636, followed by f_5 (HIV Cases) and f_1 (Population). This indicates that malaria prevalence is the most significant driver of healthcare demand at the district level in Rwanda, aligning with national health concerns.

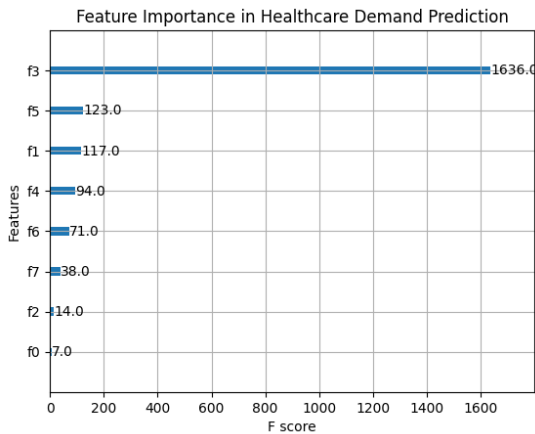


Fig. 12. Feature importance based on the ensemble model's F-score in the healthcare demand prediction task.

Feature mapping: Feature 0 (f_0) = District, Feature 1 (f_1) = Population, Feature 2 (f_2) = Healthcare_Facilities, Feature 3 (f_3) = Malaria_Cases, Feature 4 (f_4) = TB_Cases, Feature 5 (f_5) = HIV_Cases, Feature 6 (f_6) = Deaths, Feature 7 (f_7) = Year.

4) *SHAP Analysis for Model Explainability:* To provide further interpretability, SHAP (SHapley Additive exPlanations) values were computed and are shown in Figure 13. Features such as **Malaria Cases**, **TB Cases**, and **HIV Cases** had the most significant impact on the model output. Malaria and TB values had both negative and positive contributions depending on their magnitude, highlighting the non-linear relationship between disease prevalence and healthcare needs. This interpretability is essential for stakeholders to understand how model predictions relate to actual epidemiological variables and disease spread.

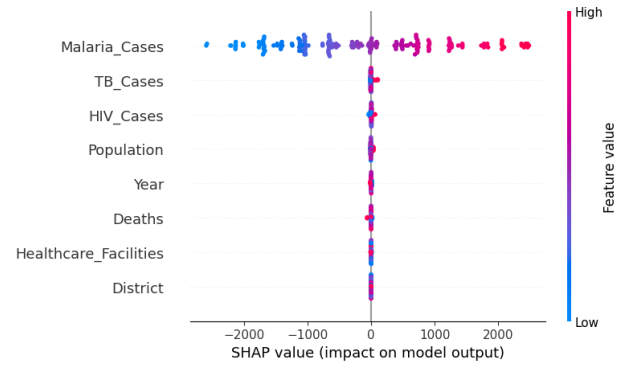


Fig. 13. SHAP summary plot showing the impact of each feature on model output for healthcare demand prediction

5) *Feature Interaction Analysis: Population and Malaria Cases:* To further interpret the model's behavior, we analyzed SHAP interaction values between key features. Figure 14 shows a dependence plot of Population (x-axis) versus its SHAP contribution (y-axis) to the healthcare demand prediction, with data points colored by the standardized value of Malaria_Cases.

From the plot in figure 14, we observe the following key insights:

- For districts with low to moderate population (left side of the x-axis), the SHAP values for Population are clustered around zero or slightly negative. This suggests that population alone does not significantly increase predicted healthcare demand in less populated areas, unless modified by other variables such as disease prevalence.
- As the population increases beyond one standard deviation above the mean (right side of the plot), the SHAP values for Population rise sharply—reaching values above 30. This indicates that in highly populated districts, population becomes a major positive contributor to the model's prediction, reinforcing the intuitive understanding that densely populated regions are more likely to experience higher healthcare demands.
- Importantly, the color gradient (representing Malaria_Cases) reveals that the impact of population is modulated by malaria prevalence. High population

districts with high malaria incidence (magenta points) tend to have the highest SHAP values, suggesting a synergistic effect where population amplifies the burden of malaria on healthcare systems. Conversely, high population areas with low malaria cases (blue points) contribute less to predicted demand.

This interaction underscores the importance of considering both demographic and epidemiological variables in tandem when prioritizing healthcare resource distribution. Population size alone is not a sufficient predictor of demand; it is the intersection with disease prevalence, particularly malaria, that intensifies the urgency for intervention.

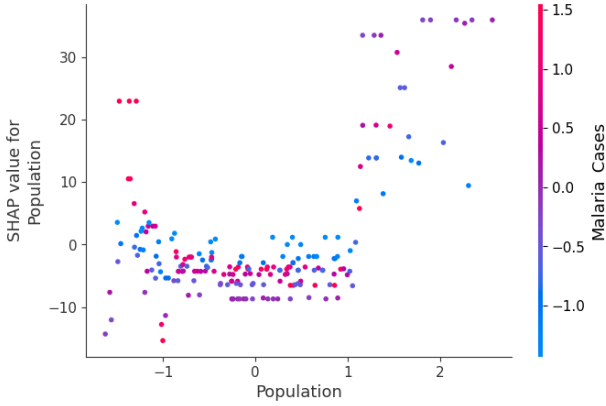


Fig. 14. SHAP dependence plot showing how Population impacts the model's prediction, colored by Malaria Cases

E. Key Insights from machine learning

- **Superior Performance of Ensemble Learning (96% Accuracy):** The ensemble model combining XGBoost and Random Forest significantly outperformed individual classifiers, achieving an accuracy of **96%**. This demonstrates the advantage of leveraging multiple learning algorithms to capture diverse data patterns. While the neural network achieved 92% and the standalone Random Forest reached 89%, the ensemble approach successfully reduced variance and bias by aggregating predictions, leading to a more robust and generalizable model. This result is particularly important for public health applications, where high predictive reliability is essential for real-world deployment in policy planning.
- **Malaria Prevalence is the Dominant Driver of Healthcare Demand:** Among all features, `Malaria_Cases` (Feature `f3`) emerged as the most critical factor in predicting healthcare needs, with an F-score of 1636—over 13 times greater than the next most important feature. This dominance, further corroborated by SHAP value analysis, highlights malaria as the leading contributor to healthcare burden in Rwanda. Given its strong predictive power, malaria incidence should be considered a primary indicator for prioritizing health infrastructure expansion, especially in rural and high-risk districts.
- **Consistent and Accurate Multi-Class Classification:** The confusion matrix shows that the model accurately

classified districts across all healthcare priority levels—Low, Medium, and High—with minimal confusion. Importantly, the model avoids high-to-low or low-to-high misclassifications, which are more detrimental in resource allocation decisions. The few errors observed mostly involved adjacent classes (e.g., High misclassified as Medium), indicating that the model's mistakes are tolerable and often involve borderline cases. This level of consistency adds trustworthiness to the model's predictions in a policy-sensitive context.

- **SHAP Values Enable Transparent Model Interpretability:** SHAP (SHapley Additive exPlanations) analysis provided valuable insights into how individual features influenced model predictions across districts. The SHAP summary plot revealed that in addition to malaria, TB and HIV cases also had significant influence, particularly in regions with high disease prevalence. Notably, these features demonstrated both positive and negative SHAP contributions, implying that their effects on healthcare demand are non-linear and context-specific. This level of interpretability is essential for decision-makers, allowing them to understand *why* the model predicts a certain priority level and which epidemiological factors drive the decision, fostering transparency and accountability in healthcare planning.
- **Implications for Resource Allocation and Health Policy:** The combined insights from accuracy metrics, feature importance, and SHAP interpretation suggest that disease-specific data, particularly for malaria, can be used as a proxy to guide equitable distribution of healthcare resources. The high performance of the model supports its practical application in guiding facility construction, mobile health unit deployment, and staffing priorities, especially in high-demand districts. This framework can also be extended to support dynamic health resource reallocation during outbreaks or seasonal fluctuations in disease prevalence.

These results support the feasibility of using machine learning for guiding healthcare facility allocation in resource-constrained settings like Rwanda.

V. CONCLUSION

This research underscores the critical misalignment between healthcare resource allocation and disease burden in Rwanda, particularly for Tuberculosis and Malaria. By leveraging geospatial analysis and data-driven methodologies, the study reveals significant disparities in healthcare facility distribution, with high-burden regions often lacking adequate resources. The weak correlations between disease prevalence and facility availability highlight inefficiencies in the current system, exacerbating accessibility challenges and health inequities. These findings not only provide actionable insights for Rwanda's healthcare system but also present a replicable model for other low-resource settings striving for equitable and efficient healthcare delivery. This insight suggests an overwhelming need to explore an alternative approach to healthcare resource allocation.

VI. LIMITATIONS

While this study provides valuable insights into optimizing healthcare facility distribution using machine learning and epidemiological data, several limitations must be acknowledged:

- **Limited Data Quality and Granularity:** The accuracy and robustness of the predictive models are constrained by the quality and consistency of the underlying datasets. Several district-level variables—including disease prevalence, mortality, and infrastructure—suffer from missing values, inconsistent formatting, or aggregated reporting. This restricts the ability to perform fine-grained analysis and may introduce bias in underreported districts, especially in remote or underserved regions.
- **Facility Quality and Capacity Not Considered:** This research focuses primarily on the *presence* or *location* of healthcare facilities, without accounting for their internal capacity, staffing levels, service availability, or operational quality. As a result, a facility in one district may be vastly under-resourced compared to another, yet treated equally in the analysis. This oversimplification may misrepresent actual accessibility and healthcare equity across regions.
- **Static Demographic and Infrastructure Metrics:** The population and healthcare facility data used are based on static snapshots in time and do not reflect temporal changes, such as urbanization, migration, or infrastructure expansion. This limits the model's applicability in predicting future demand or responding to dynamic public health trends. Incorporating time-series data would enhance the ability to conduct longitudinal analyses and anticipate shifts in resource needs.
- **Omission of Socioeconomic and Environmental Factors:** Although disease burden is central to the analysis, the model does not currently incorporate socioeconomic variables (e.g., income, education, insurance coverage) or environmental risk factors (e.g., climate, water access), which may significantly influence healthcare access and outcomes. Inclusion of such contextual variables could improve model fairness and policy relevance.
- **No Real-Time Model Deployment or Field Validation:** The study remains at the analytical stage and has not yet been integrated into real-time decision-making tools or validated through field implementation. Further collaboration with health ministries and local stakeholders is necessary to operationalize the model and test its recommendations in real-world planning processes.

Despite these limitations, the findings offer a strong foundation for data-driven and evidence based healthcare resource planning and allocation strategies.

VII. FUTURE WORK

Future work will focus on addressing the identified disparities in healthcare resource allocation by developing and applying machine learning models to infer optimal resource distribution based on disease burden. These models will aim to provide data-driven recommendations for equitable healthcare access. Additionally, the study will incorporate comprehensive HIV

case data, which was unavailable in the current DHS dataset, to extend the analysis to this critical disease. By integrating these advancements, the research will enhance its framework for informed decision-making, offering a more robust and scalable solution for improving healthcare resource allocation in Rwanda and similar contexts.

VIII. ANNOTATED BIBLIOGRAPHY

To build a strong foundation for our research on optimizing healthcare resource distribution in Rwanda, we reviewed key studies that explore spatial healthcare accessibility, GIS-based resource allocation, and predictive modeling for healthcare demand. These studies provide insights into the limitation-sof traditional population-based healthcare distribution and support the need for a data-driven, disease-prevalence-based approach. Below, we present an annotated bibliography of the most relevant research that informs our study.

- 1) Analysis of Spatial Distribution of Health Care Facilities and its Effects on Access to Primary Healthcare in Rural Communities in Kpandai District, Ghana[3].
This study analyzed the spatial distribution of health-care facilities in rural Ghana, revealing a misalignment between facility locations and high disease-prevalence areas. Using GIS mapping and spatial analysis, the researchers demonstrated that healthcare accessibility was limited in regions with the greatest disease burden. The findings highlight the need for disease-based healthcare planning, aligning with our research focus on optimizing resource allocation based on health needs rather than administrative boundaries.
- 2) The Effect of Distance to Health Facility on Neonatal Mortality in Ethiopia[6].
This study examined the impact of distance to healthcare facilities on neonatal mortality in Ethiopia, showing a direct correlation between increased travel distance and higher mortality rates. The researchers used demographic and health survey data to analyze how spatial barriers affect access to essential neonatal care. Their findings underscore the importance of geographic accessibility in healthcare planning, supporting our research on optimizing facility distribution to improve health outcomes in underserved regions.
- 3) Issues of Healthcare Planning and GIS: A Review[9].
This study provided a comprehensive review of GIS applications in healthcare planning, highlighting its effectiveness in optimizing facility distribution and resource allocation. The researchers examined various case studies where GIS was used to identify underserved areas and improve healthcare accessibility. Their findings support our research by demonstrating how spatial analysis can enhance data-driven decision-making for equitable healthcare distribution.
- 4) GIS in Healthcare Facility Planning and Management: A Review[7].
This study explored the transformative role of GIS in healthcare facility planning, focusing on spatial analytics, predictive modeling, and real-time monitoring.

The researchers discussed how GIS improves decision-making by identifying optimal facility locations and forecasting healthcare demand. Their analysis aligns with our research by demonstrating how geospatial tools can enhance data-driven healthcare planning while addressing challenges like data privacy and system integration.

5) The Contributions of Population Distribution, Healthcare Resourcing, and Transportation Infrastructure to Spatial Accessibility of Health Care[10].

This study quantified how population distribution, healthcare resources, and transportation infrastructure contribute to healthcare accessibility using an enhanced catchment area method. The researchers found transportation infrastructure dominates accessibility in urban areas, while resource distribution is more critical in rural regions. Their approach provides a framework for our research on integrating disease burden data (TB, Malaria, HIV) when optimizing healthcare facility placement in Rwanda beyond traditional administrative boundaries.

These papers collectively provide a strong theoretical and methodological basis for our study. They demonstrate the necessity of integrating disease prevalence, spatial accessibility, and healthcare resource distribution into healthcare planning. Our research builds upon these studies by applying similar methodologies to Rwanda, with a specific focus on tuberculosis, malaria, and HIV prevalence, to develop a data-driven framework for optimizing healthcare resource allocation.

REFERENCES

- [1] K. Ganasegeran, M. K. Kamarudin, and M. R. A. Manaf, "Modeling accessibility to public health facilities in resource-limited settings through gis and geo-ai applications," in *Advances in Geoinformatics Technologies: Facilities and Utilities Optimization and Management for Smart City Applications*, R. N. Yadava and M. U. Ujang, Eds. Cham: Springer Nature Switzerland, 2024, pp. 319–348.
- [2] (2019) Wisn final report 2019. Accessed: Feb. 14, 2025. [Online]. Available: https://moh.prod.risa.rw/fileadmin/user_upload/Moh/Publications/Reports/WISN_Final_report_2019.pdf
- [3] J. Amoah-Nuamah, W. Agyemang-Duah, G. P. Ninorb, and B. G. Ekeme, "Analysis of spatial distribution of health care facilities and its effects on access to primary healthcare in rural communities in kpandai district, ghana," *Cogent Public Health*, vol. 10, no. 1, p. 2183566, Dec. 2023.
- [4] T. Mizan and S. Taghipour, "Medical resource allocation planning by integrating machine learning and optimization models," *Artificial Intelligence in Medicine*, vol. 134, p. 102430, Dec. 2022.
- [5] (2025, Jan.) Machine learning for healthcare resource allocation. Accessed: Feb. 17, 2025. [Online]. Available: https://www.researchgate.net/publication/387723321_Machine_Learning_for_Healthcare_Resource_Allocation
- [6] G. Kibret, D. Demant, and A. Hayen, "The effect of distance to health facility on neonatal mortality in ethiopia," *BMC Health Services Research*, vol. 23, no. 1, p. 114, Feb 2023.
- [7] P. W. Biu, C. N. Nwasike, N. K. Nwaobia, C. A. Ezeigweneme, and J. O. Gidiagba, "GIS in healthcare facility planning and management: A review," *World Journal of Advanced Research and Reviews*, vol. 21, no. 1, pp. 012–019, Jan. 2024.
- [8] OpenStreetMap Wiki, "Key: healthcare - openstreetmap wiki," 2025.
- [9] B. F. Khashoggi and A. Murad, "Issues of healthcare planning and gis: A review," *ISPRS International Journal of Geo-Information*, vol. 9, p. 352, 2020. [Online]. Available: <https://doi.org/10.3390/ijgi9060352>
- [10] L. Chen, T. Chen, T. Lan, C. Chen, and J. Pan, "The contributions of population distribution, healthcare resourcing, and transportation infrastructure to spatial accessibility of health care," *Inquiry*, vol. 60, p. 469580221146041, 2023. [Online]. Available: <https://doi.org/10.1177/00469580221146041>