International Journal of Innovative Computing 16(1–2) 271–278

UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Integrated GenAI Data Observatory Framework for Evidence-based Analytics of Malaria Burden

Ifiok J. Udo*
Department of Information Systems
University of Uyo, Nigeria
Email: udojames@uniuyo.edu.ng

Emmanuel A. Dan
Department of Computer Science
University of Uyo, Nigeria
Email: emmanueldan@uniuyo.edu.ng

Idara I. James
Department of Computer Science
Akwa Ibom State University, Ikot Akpaden, Nigeria
Email: idarajames@aksu.edu.ng

Utibe Ofon
Department of Microbiology
University of Uyo, Nigeria
Email: utibeofon@uniuyo.edu.ng

Submitted: 15/9/2025. Revised edition: 28/11/2025. Accepted: 5/1/2026. Published online: 28/7/2026
DOI: https://doi.org/10.11113/ijic.v16n1-2.701

***Abstract*—Minimal size of data can degrade the performance of machine learning models, leading to overfitting. This limitation hinders the transferability and generalizability of the models to different contexts or datasets. In resource-constrained settings, the lack of adequate data inhibits the development of robust, data-driven systems capable of delivering accurate insights. However, Generative Artificial Intelligence (*GenAI*) leverage Large Language Models (LLMs) to create and synthesize a wide array of meaningful content, including text, images, audio, video, and multimedia data. These models generate synthetic data that replicate the statistical properties of real datasets, thereby augmenting existing data and improving the overall quality of inputs available for machine learning tasks. In this paper, an integrated *GenAI*-enabled data observatory framework designed to enhance the storage and analytical processing is proposed to improve the quality of evidence of predictive analytics tasks. This framework employs advanced data warehouse technology combined with the capabilities of LLMs to facilitate a more effective data management system. To validate the machine analytics layer, data is sourced from the repository of Malaria Indicator Survey (MIS), which provide in-depth insights into the malaria burden faced by rural communities across Nigeria. The machine analytical layer of the proposed framework demonstrates percentage increases in overall accuracies with Decision Tree (DT) and Extreme Gradient Boosting (XGBoost) algorithms, achieving improvements of 23.81% and 15.38%, respectively, when compared to using the MIS data alone. This substantial improvement underscores the potential of integrating *Gen*AI with traditional data management approaches to drive precise evidence for effective decision-making and policy formulation in malaria control and interventions.**

***Keywords*—Large Language Model, Machine Learning, Malaria Burden, Data quality, Data warehouse, Predictive analytics**

I. INTRODUCTION

Insufficient dataset can significantly undermine the performance of machine learning models, leading to overfitting; a condition wherein the model becomes excessively tailored to the training data. This challenge hinders the ability of the model to generalize effectively to new, unseen instances [1]. However, Generative AI (GenAI) takes a step further by leveraging large language models (LLMs) to generate and synthesize impactful content across a myriad of formats, including text, images, audio, video, and various forms of multimedia [2] to increase the quantity of the limited data size. GenAI can also accommodate increasing dynamic features associated with disease burden in diverse regions over time given its capability for generalizability.

As machine learning algorithms (a sub-component of AI) analyze historical data, relevant data patterns are extracted, and insights (otherwise called evidence) essential for informing decision-making and strategic policy formulations are derived to enhance efficiency, optimize resources, and ultimately drive innovation and growth in different areas of applications. Examples of cutting-edge GenAI technologies include ChatGPT [3] generates coherent and contextually relevant text responses, DALL-E [4], produces a high-quality, imaginative images based on textual prompts, and Google's Bard [5], offers conversational abilities.

Over the years, related data of diseases outbreaks, impacts and consequences in resource constrained environments is often collected, stored, analyse and disseminated by international organizations. Examples of such organizations are, WHO, USAID, UNICEF and more. In light of the recent withdrawal of funding from non-governmental organizations responsible for critical data collection, analysis, and reporting of disease burden by international health agencies in sub-Saharan Africa, it is imperative that modern technology is adopted to meet urgent information needs. This shift is essential for enhancing the quality of life and ensuring good health for all, in alignment with the targets set forth by the Sustainable Development Goals (SDGs). Embracing innovative solutions will not only fill the gaps left by diminished funding but also drive progress towards sustainable health advancements in the region [6]. As the disease burden increases within communities and households, the resulting health impacts, economic losses, and disruptions to social structures become evident, particularly through elevated mortality rates among vulnerable groups such as young children under 5 years and pregnant women. This burden also diminishes overall productivity, leading to increased absenteeism in both workplaces and educational institutions [7]. Furthermore, the strain on healthcare infrastructure diverts resources away from other crucial areas of development. For example, in Nigeria, malaria has been recognized as the disease with the second highest mortality rate following cholera, according to recent WHO data [7].

Malaria is a serious illness caused by Plasmodium parasites and is primarily transmitted to humans through the bites of infected female Anopheles mosquitoes. It remains a leading cause of mortality and is a major public health issue, particularly in sub-Saharan Africa, where pregnant women and children under five are most affected. The impact of malaria can be observed at various levels, including households, localities, states, regions, and nations. Even though malaria is both preventable and treatable, its economic burden which includes treatment costs and lost productivity represents about 1.3% of a country's gross domestic product (GDP) each year. In resource-constrained environments, the high cost of malaria treatment and limited access to healthcare facilities exacerbate the disease's prevalence. To address this, existing studies adopted GeoAI and machine learning analytics as technologically-driven tools for generating precise evidence [8]. GeoAI merges geospatial data systems with AI analytics, enabling real-time or historical observation, analysis, and interpretation of spatial phenomena. AI facilitates machine learning analytics to extract insights from data, helping in evidence-based decision-making that relies on accurate information derived from available data to support effective interventions and optimal resource allocation. The workflow of architecture of GeoAI framework for malaria control is illustrated in Fig. 1.

To mitigate these challenges, existing research has also proposed strategies for validating data at each stage of the machine analytics pipeline, or for defining a limited study area to facilitate effective data validation in health-related research [9]. However, these approaches may impede the efficiency of data analytics tasks and constrain the technological capabilities required for managing extensive data analytics projects. This situation prompts several inquiries within this study, such as:

(a) How can unstructured data of malaria burden be stored and analysed to ensure improve quality in machine learning analytics tasks?
(b) How could integration of data from diverse sources and augmented synthetic data guarantee high performance accuracy?

This paper utilizes data from [10], as well as structured responses from *ChatGPT* and machine learning algorithms, to design a framework for integrated data observatory that stores and analyses malaria data directly from households. By leveraging large language models and AI tools, the specific objectives are as follows:

(a) To curate data on malaria burden indicators from rural regions of Nigeria using the repository of malaria indicators and the demographic health surveys from 2021.
(b) To design a data dictionary for the multidimensional data of malaria indicators in rural Nigeria.
(c) To evaluate the quality of evidence generated from the data managed by the proposed integrated observatory framework.

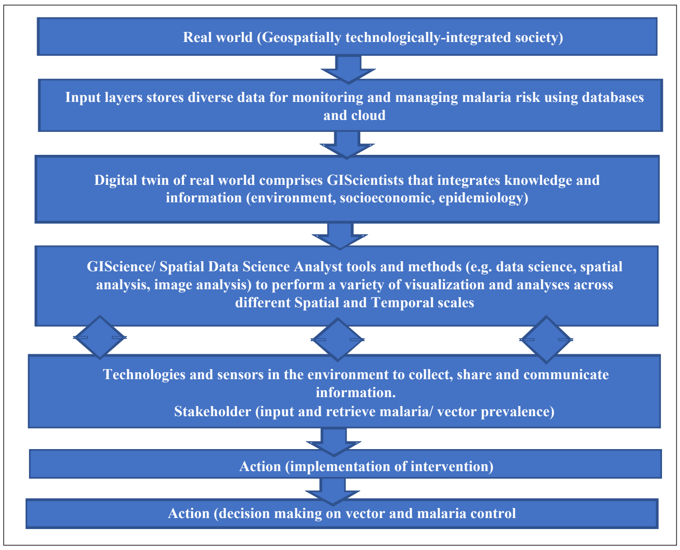

Fig. 1. The workflow of architecture of GeoAI of managing malaria control in technological-driven society (Adapted from [8])

To ensure the findings contribute meaningfully to malaria control and interventions efforts, machine learning algorithms are implemented in the data analytics engine to tune the proposed framework to derive insights and inform decisions towards future interventions aimed at eradicating malaria prevalence and burden on infected households.

II. RELATED WORKS

Related studies on development of global databases also play a crucial role in tracking the prevalence of malaria parasite, as highlighted by [11], while [12] designed observatories that connect climate data with health outcomes. Their research investigates the potential benefits and challenges of an environmental health observatory aimed at understanding the influence of climate change on public health, utilizing comprehensive Brazilian climate and health data sets to draw connections and insights.

In a forward-looking study, [13] developed predictive models to assess the future abundance of malaria-carrying mosquitoes. This work incorporated Satellite Earth Observation data alongside in-situ geo-spatial data and advanced machine learning algorithms, providing a nuanced understanding of the environmental conditions that favour mosquito proliferation.

Further advancing the field, [14] introduced an innovative data evaluation framework designed to assist risk modelers in determining the adequacy and reliability of various data sources. Their framework leverages a facility-based database containing crucial data points, including global population counts and CHIRPS (Climate Hazards Group InfraRed Precipitation with Stations) precipitation estimates, enabling researchers to make more informed decisions based on robust data analytics.

Additionally, other significant research related to malaria and data observatories has focused on the identification of biogeographical variables that influence the abundance and distribution of malaria vectors. This analysis utilized time-series data from cost-free Earth Observation (EO) sources in the Democratic Republic of Congo by [15]. This underscores the importance of utilizing accessible satellite technologies for public health research.

A comprehensive overview of these related works can be found in Table I, which summarizes key findings and methodologies. Collectively, these studies demonstrate the application of machine analytics to address specific challenges in malaria control and prevention. They emphasize the need for information that can help combat malaria parasites and the mosquito prevalence. However, a critical gap exist; many of these approaches struggle to adapt to resource-constrained settings where rural populations face significant barriers to accessing healthcare facilities or lack the technological infrastructure to benefit from remote services.

To the best of our knowledge, there is yet to be a study that explores the concept of an integrated data observatory utilizing data from disparate repositories and LLMs for the generation of valid evidence to guide interventions of malaria burden on households in resource-constrained setting [18]. Such an initiative could augment the volume and variety of public health data available, ultimately enhancing the generalizability and effectiveness of machine learning models aimed at addressing malaria and other public health challenges in resource-limited settings.

III. PROPOSED GEN*AI* DATA OBSERVATORY FRAMEWORK

This paper describes an integrated GenAI-enabled data observatory framework specifically designed to enhance the storage, and analysis of data for overall improvements in the quality of data critical for machine analytics tasks as shown in Fig. 2.

The framework utilizes data warehouse technology, seamlessly integrated with the capabilities of large language models (LLMs). The key components include data source layer, Extraction, Transformation and Loading (ETL) layer, Data warehouse layer [19], machine analytics, output and visualization dashboard, which interacts together to create a cohesive framework for effective data storage and analytics.

TABLE I. SUMMARY OF RELATED STUDIES ON MALARIA DATA OBSERVATORIES

Author	Objective	Methodology / Data	Findings	Focus
[11]	To assemble a global database	Relational database and GIS	3,351 spatially independent PR estimates since 1985	Prevalence of malaria parasite
[13]	To predict mosquito abundances for future	Satellite Earth Observation and other in-situ geo-spatial data and machine learning	Mean Absolute Error of the predictions did not deviate more than 14%	Prevalence of mosquito
[14]	To introduce a data evaluation framework to assist risk modelers in evaluating data adequacy.	Malaria Atlas Project, a healthcare facility database		Adequacy of biogeographical malaria data
[15]	To identify the biogeographical variables driving the abundance and distribution of three malaria vectors	Random Forest using Time series of cost-free Earth Observation (EO) data in DRC	higher rainfall implies higher abundance of mosquito	Prevalence of Malaria vector
[16]	To develop targeted approach to intervention deployment at the local government area (LGA) level as part of the high malaria burden to high impact response.	An agent-based model of Plasmodium falciparum transmission was used to simulate malaria morbidity and mortality in Nigeria's 774 LGAs	Subnational data collection systems are required to allow increased confidence in predictions at sub national level.	Assessment of malaria intervention
[17]	To examine the evolution of malaria in Nigeria over twenty years (2000–2020) and exploring the impact of environmental factors, develop a predictive model	Malaria indicator Survey data, spatial analysis and statistical methods	Regional variations in malaria prevalence over time	Prevalence of malaria

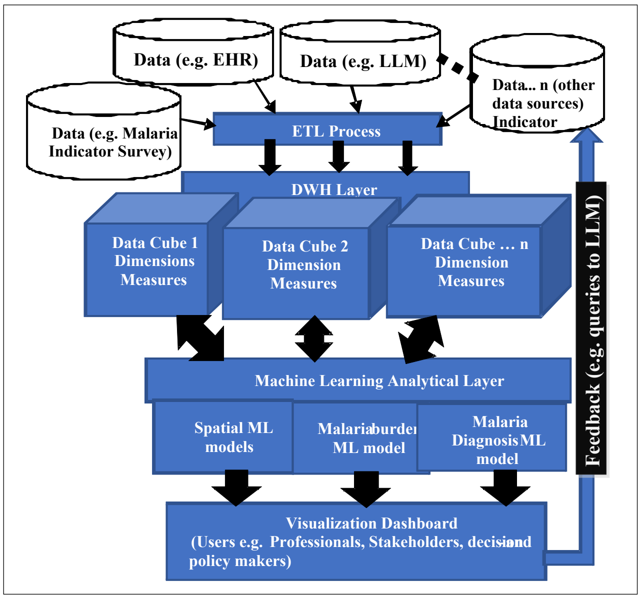

Fig. 2. Workflow of integrated GenAI data observatory framework for evidence-based analytics

There is also a provision for professionals to update the source(s) of data based on the evidence generated on the dashboard via a feedback loop.

A. Data Source Layer

The analysis within this study is grounded in secondary data obtained from two distinct sources: the Nigeria Malaria Indicator Survey 2021 (accessible via https://microdata.worldbank.org/index.php/catalog/5763). These datasets provide comprehensive coverage of rural communities across all thirty-six states of Nigeria, focusing on various key malaria control indicators. Fig. 3 shows the states and the number of rural communities covered in the data. Specifically, they include metrics such as prevalence rates, treatment coverage, and preventive measures implemented in these communities and more. The integration of such diverse data sources enhances the robustness of the analysis, allowing for targeted insights into malaria control efforts and effectiveness within different demographic contexts in Nigeria. However, synthetic data from LLM (e.g. *chatGPT*) is obtained using user queries

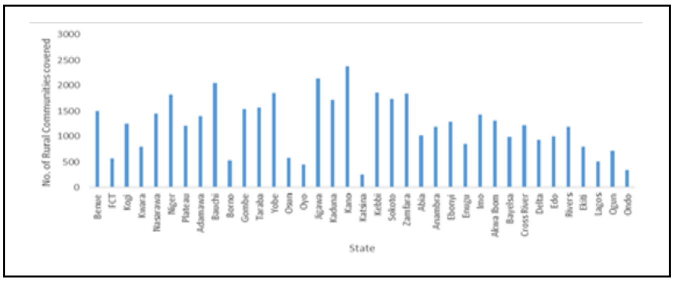

Fig. 3. Number of rural communities covered per state in Nigeria

B. Extraction, Transformation and Loading (ETL) Process

In this phase, data pertaining to the burden of malaria is retrieved from various sources (i.e., extraction), followed by a process of cleaning and integration into a standardized format (i.e., transformation). The subsequent loading process involves transferring the data into a data warehouse [19], facilitating multidimensional analysis.

C. Data Warehouse Layer

This layer consists of data cubes. Data cube is a multidimensional data structure used in data warehousing and OLAP (Online Analytical Processing) systems that enables efficient organization, storage, and retrieval of aggregated data across multiple dimensions. In this case a few of the data cube for malaria burden is summarized as follows:

Data Cube 1: Malaria Prevention and Control Cube features depicting burden on households

Dimension 1: Location, time, age group, intervention type, and incidence

Measures 1: Case counts, positivity rates, mortality rate, rate of malaria occurrence

Dimension 2: Facility, treatment type, insecticide type, parasite prevalence

Measures: Number treated, treatment success/failure rates, frequency of spraying, proportion of population infected with parasite

Dimension 3: households covered, vector abundance

Measures: Number of households covered, density of vectors in the area

Dimension 4: Climate and Vector Control Cube

Dimension 5: Location, time, environmental variables

Measures: Rainfall, temperature, vector density

Dimension 6: Intermittent preventive treatment, blood test

Measures: Number of treatment times, number of children tested and more.

1) Machine analytical layer

In this layer, machine learning models are implemented using the data cubes to generate actionable insights and trends into the malaria incidences. The ML models specific models required are:

- *Malaria burden ML Model*: Predicts and classify malaria burden/ incidence, assess intervention impact on households' class of burden.
- Spatial ML model: This model predicts malaria hotspots using spatial covariates
- Malaria treatment ML model: Predicts the drugs and outcome of diagnosis.

This paper is focus on the framework that develops malaria burden ML models to classify the burden on malaria-endemic households using the integrated data stored in the data observatory framework.

D. DatavVisualization layer

This layer provides interactive maps and charts showing malaria burden, treatment coverage, and predictive alerts for professionals, stakeholders and policymakers for informed decision-making and policy formulations. Based on the output of the framework, the data source could be updated via feedback and queries to the LLM. The significance of this layer among others include: enhancement of interpretability of evidences, reporting and communication of the evidences to guide strategic interventions and optimal allocation of resources.

The preliminary findings from the curation and preprocessing of the MIS 2021 data sets, which were gathered from various rural communities across selected states in Nigeria, are presented in Fig. 4. These results support the conclusions drawn by [17], which highlighted regional disparities in malaria prevalence over time, with a notable concentration of cases in northern Nigeria. In Fig. 4, the rural communities of FCT and Kogi represent selected Northern states, while Kwara illustrates the Western state. Additionally, Akwa Ibom and Delta rural communities represent the Southern state While Enugu rural communities represent Eastern states with moderate or lower malaria burden in comparison to the rural communities in the Northern Nigeria.

TABLE II. A SUMMARY OF DATA DICTIONARY THAT DEFINES MALARIA INDICATORS ON HOUSEHOLDS

Data element	Type	Description	Scale	Constraint	Dimension
Household_coverage	float	The percentage of households within a specific area that have been sprayed with insecticide.		≥ 0	Indoor residual spraying (IRS)
Insecticide_type	integer	The specific insecticide used in the IRS program to note varying durations of effectiveness as well as mosquito species		{0, 1, 2, 3}	
Frequency_of_spraying	integer	The timing of IRS application (e.g., before the rainy season) can be crucial for maximizing its impact on mosquito populations and malaria transmission		≥ 0	
Parasite_prevalence	integer	The proportion of the population with malaria parasites in their blood, indicating the level of malaria infection	1-positive 2-Negative 3-don't know	{1, 2, 3}	
Incidence	float	The rate at which new malaria cases occur within a population		≥ 0	
Anemia		The prevalence of anemia, as IRS can reduce malaria and thus reduce anemia, especially in children	1-positive 2-Negative 3-don't know	{1, 2, 3}	
Vector_Mosquito_ Abundance	float	The population density of mosquito vectors in the area	Mosquitos per sq. km	≥ 0	
Household_income	float	Total earnings of the household	Unit Amount	≥ 0	Socioeconomic factors
Housing_characteristics	integer	Types of roof, walls, floor material		{0,1,…,18}	
Type-of-mosquito_nets_slept	integer	Type of mosquito nets uses by the households		{0, 1, 2, 3}	
Insecticide_Treated_Net _Use		ITN ownership and use			
Sex	integer	identifying eligible individuals for specific questionnaires e.g. male, female	1- Male 2- female	{1, 2}	Household members' characteristics
Age		analyzing malaria prevalence by age group		{1,2,3,4}	
Relationship to_the_household head	integer	household structure that potentially identify vulnerable groups within the household		{1, 2, 3,…, 15}	
Education_level	integer	linked to malaria knowledge, preventative behaviors, and access to resources, which can influence malaria risk	0-no education 1-incomplete primary 2-Complete primary 3-incomplete secondary 4-complete secondary 5- higher	{0, 1, 2, 3, 4, 5}	

Data element	Type	Description	Scale	Constraint	Dimension
Usual_residents_visitors	integer	This distinction helps in accurately assessing the population at risk and understanding the impact of different living situations on malaria transmission	1-resident 2-visitor	{1,…, 2}	
Household_dwelling_characteristics	integer	water sources, sanitation facilities, and housing materials, sleeping rooms. Livestock, mosquito		{1,…, 19}	
Ownership_of_household_asset	integer	Helps categorize households based on their economic status, which are correlated with other factors related to malaria risk and intervention coverage. E g. Electricity, telephones, refrigerators, motorcycles, cars, truck, radios, indoor plumbing	1-poor 2-poorer 3-middle 4-richer 5-richest	{1, 2, 3, 4, 5}	Wealth index
Malaria_prevention_and_treatment_resources	integer	different access to and utilization of malaria prevention and treatment resources	1-accept medicine 2-refined 3-others	{1, 2, 3}	Risk and intervention
Time_of_treatment	Date	Time of treatment of high fever by children under 5 years and pregnant women	ISO 8601	YYYY-MM-DD	Malaria treatment
Intermittent_preventive_treatment		Intermittent preventive treatment during pregnancy	0-no 1-yes 3-don't know	{0,1,2}	
Place_first_sought_treatment	integer	Place of treatment	1-hospital 2-health centre and more	{1, 2,…,13}	

The summary of the multidimensional data sourced from various integrated platforms for evidence-based analysis of malaria burden in rural communities in Nigeria is presented in Table III. This data cube encompasses six dimensions as shown in Table II, and each dimension is defined by specific measures and facts.

TABLE III. TABULAR REPRESENTATION OF DATA CUBE MEASURES AND ASSOCIATED FACTS FOR SELECT MALARIA HOUSEHOLDS' LOCATION, MALARIA RISKS AND INTERVENTION (A-I) USING MIS FEATURES

HoH	A	B	C	D	E	F	G	H	I …	MB
Benue_1	0.61904	0.375	0	0	0.71428	0.12	0	0	0 …	VL
Benue_2	0.61904	0.375	0.0302	0	0.71428	0.12	0	0	0 …	L
FCT_1	0.85714	0.75	1	1	0.85714	0.12	0.0080	0	0 …	H
FCT_2	0.85714	0.75	1	1	0.85714	0.12	0.0080	0.016	0.444…	VH
Kogi_1	0.57142	0.125	0.2857	0	0.57142	0.25	0.0020	0	0 …	H
Kogi_2	0.14285	0	0	0	1	0.12	0	0	0.333…	VL
Kwara_1	0.23809	0.25	0.4285	0	0.57142	0.12	0	0	0.222 …	L

Keys:

Household location (HoH), Proportion of household members who are dependents (A), Proportion of children under 5 and below over the total no of households (B), Proportion of mosquito bed nets over total number of households (C), Number of children under mosquito bed net previous night (D), Ideal number of children under 5 years over total number of children (E), Average days husband/partner worked in last 7 days/12 months (F), Average cost of treatment of fever over 12 months (G), Average number of days after fever began that advice or treatment was sought within 12 months (H), Average duration of pregnancy over 9 months (I), Class of Malaria burden on households: Very High (VH), High (H), Low (L), Very Low (VL)

1) Evaluation of preliminary results of machine analytics layer.

Data extracted from the machine analytics layer of the proposed data observatory framework and other existing data repositories were trained and tested on Decision Tree (DT) and Extreme Gradient Boosting (XGBoost) learning algorithms to classify malaria burdens on rural households in the study area. The overall percentage increases in performance accuracies recorded on DT and XGBoost algorithms are 23.81% and 15.38%, respectively. These results imply that the percentage increase in the overall performance of classification of malaria burden using integrated data from the proposed observatory outperforms performance accuracies of machine analytics using data from disparate sources. This finding aligns with the work of [16] on targeted approaches to malaria intervention deployment at the local government area (LGA) level, emphasizing the need for improved data collection and storage at subnational levels to foster confidence in malaria interventions at local government areas.

IV. CONCLUSION

In resource-constrained malaria endemic households where access to healthcare facilities is limited and costs are prohibitive, data is often insufficient for advanced machine analytics. This results in a decreased level of confidence in the outcomes of data analytics at the household level. The relatively small size of data typical in these settings can be enhanced through the use of advanced data surveillance technologies, such as GeoAI, or by integrating structured responses from Generative AI (GenAI). However, the implementation of advanced technologies is often hindered by a lack of supporting infrastructure that suits resource-constrained environments.

Integrating structured responses from large language models (LLMs), like ChatGPT, emerges as an affordable alternative with a wider coverage of web sources, significantly improving data quality and ensuring better transferability and generalization in machine analytics. Unlike other synthetic datasets created through over-sampling or under-sampling algorithms, GenAI models can generate synthetic data that retains the statistical properties of real-world datasets, augmenting existing data and enhancing the quality and quantity of inputs for malaria prevention and control.

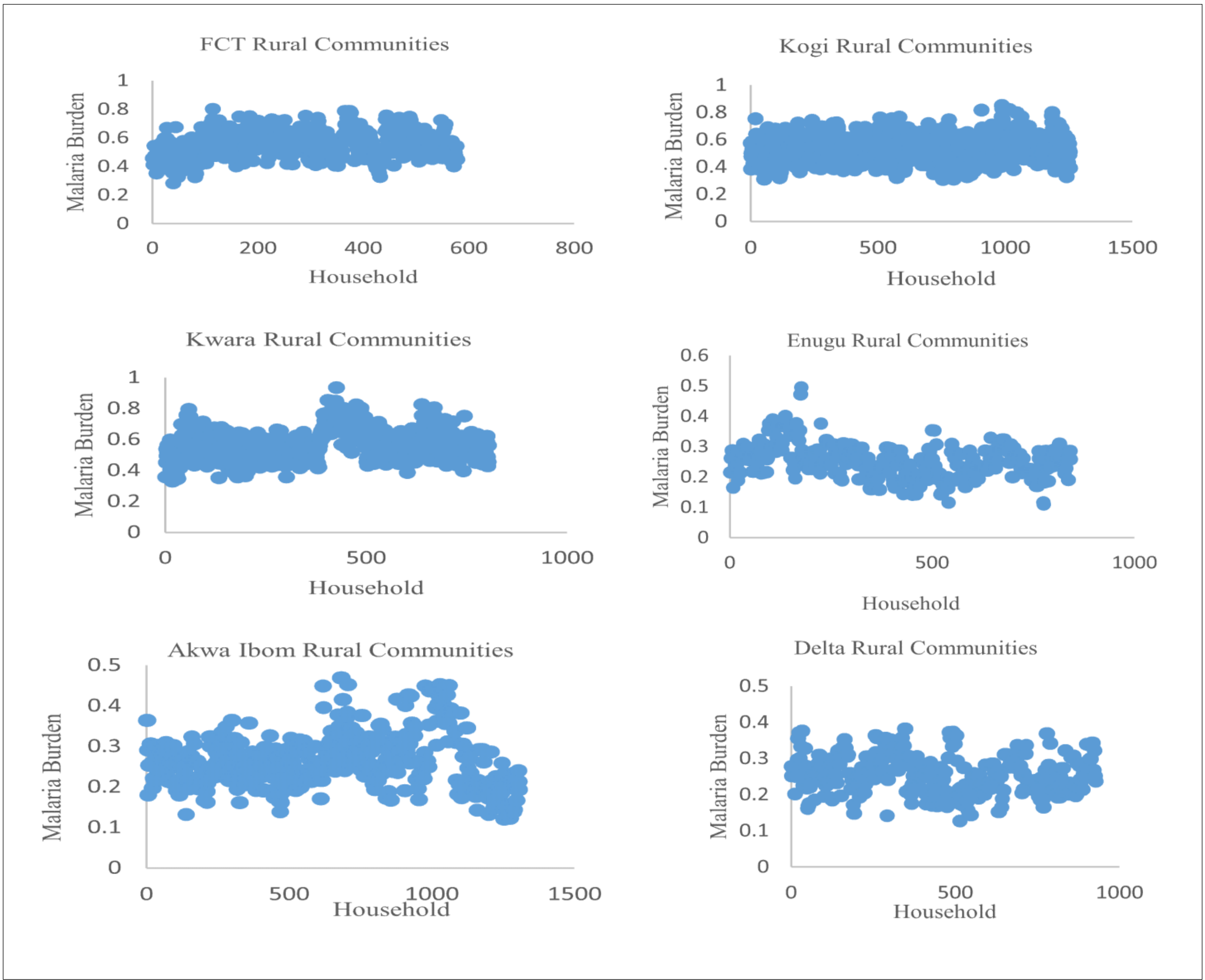

Fig. 4. Preliminary results malaria burden across rural communities across states in Nigeria from MIS (2021)

The proposed Gen*AI* data observatory framework treats malaria burden as a multidimensional data cube for storage and prompt analytics at various levels of granularity. Future work will involve the implementation and evaluation of a prototype in the study area using malaria burden indicators derived from Malaria Indicator Surveys (MIS) and Demographic and Health Surveys (DHS) integrated with Gen*AI* data. Mobile applications featuring real-time visualization and analytics will be developed with the goal of providing timely evidence that informs decision making and guides policy formulation.

V. ACKNOWLEDGMENTS/FUNDING

The authors would like to acknowledge the support of Tertiary Education Trust Fund (TETFund), Nigeria for providing financial support on this research.

VI. CONFLICT OF INTERST STATEMENT

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

VII. AUTHORS' CONTRIBUTIONS

Ifiok Udo carried out the research, wrote and revised the article. Emmanuel Dan reviewed the literature and provided the theoretical framework. Idara James implemented the programming modules of machine learning algorithms using Python. Utibe Ofon provided the background on malaria burden and reviewed the literature.

REFERENCES

[1] Safonova, G., Ghazaryan, S., Stiller, M., Main-Knorn, M., Nendel, C., & Ryo, M. (2023). Ten deep learning techniques to address small data problems with remote sensing. *International Journal of Applied Earth Observation and Geoinformation, 125*, Article 103569. https://doi.org/10.1016/j.jag.2023.103569

[2] Lv, Z. (2023). Generative artificial intelligence in the metaverse era. *Cognitive Robotics, 3*, 208–217. https://doi.org/10.1016/j.cogr.2023.06.001

[3] Welsby, P., & Cheung, B. M. (2023). ChatGPT. *Postgraduate Medical Journal, 99*(1176), 1047–1048. https://doi.org/10.1093/postmj/qgad056

[4] Marcus, G., Davis, E., & Aaronson, S. (2022). *A very preliminary analysis of DALL-E 2* (arXiv Preprint No. arXiv:2204.13807). https://doi.org/10.48550/arXiv.2204.13807

[5] Gan, R. K., Ogbodo, J. C., Wee, Y. Z., Gan, A. Z., & González, P. A. (2024). Performance of Google Bard and ChatGPT in mass casualty incidents triage. *American Journal of Emergency Medicine, 75*, 72–78. https://doi.org/10.1016/j.ajem.2023.10.034

[6] Sustainable Fund. (2015). *Sustainable development goals*. World Bank. Retrieved July 23, 2025, from https://documents1.worldbank.org/curated/en/

[7] World Health Organization. (2022). *World malaria report 2022*. World Health Organization. Retrieved from https://books.google.com.ng/books

[8] Blanford, J. I. (2024). Managing vector-borne diseases in a GeoAI-enabled society: Malaria as an example. *Acta Tropica, 260*, Article 107406. https://doi.org/10.1016/j.actatropica.2024.107406

[9] Khalid, S. (2021). A standardized analytics pipeline for reliable and rapid development and validation of prediction models using observational health data. *Computer Methods and Programs in Biomedicine, 211*, Article 106394. https://doi.org/10.1016/j.cmpb.2021.106394

[10] Malaria Indicator Survey. (2021). *Malaria Indicator Survey (MIS)*. Retrieved July 23, 2025, from https://microdata.worldbank.org/index.php/catalog/5763

[11] Guerra, A., Hay, S. I., Lucioparedes, L. S., Gikandi, P. W., Tatem, A. J., Noor, A. M., & Snow, R. W. (2007). Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. *Malaria Journal, 6*(1), Article 17. https://doi.org/10.1186/1475-2875-6-17

[12] Barcellos, C., Roux, E., Ceccato, P., Gosselin, P., Monteiro, A. M., Matos, V. P., & Xavier, D. R. (2016). An observatory to gather and disseminate information on the health-related effects of environmental and climate change. *Revista Panamericana de Salud Pública, 40*, 167–173.

[13] Tsantalidou, G., Arvanitakis, G., Georgoulias, A. K., Akritidis, D., Zanis, P., Fornasiero, D., & Kontoes, C. (2023). A data-driven approach for analyzing the effect of climate change on mosquito abundance in Europe. *Remote Sensing, 15*(24), 5649. https://doi.org/10.3390/rs15245649

[14] Petutschnig, L., Clemen, T., Klaußner, E. S., Clemen, U., & Lang, S. (2024). Evaluating geospatial data adequacy for integrated risk assessments: A malaria risk use case. *ISPRS International Journal of Geo-Information, 13*(2), Article 33. https://doi.org/10.3390/ijgi13020033

[15] Marston, C., Rowland, C., O'Neil, A., Irish, S., Wat'senga, F., Martín-Gallego, P., Aplin, P., Giraudoux, P., & Strode, C. (2022). Developing the role of Earth observation in spatio-temporal mosquito modelling to identify malaria hot-spots. *Remote Sensing, 15*(1), Article 43. https://doi.org/10.3390/rs15010043

[16] Ozodiegwu, I. D., Ambrose, M., Galatas, B., Runge, A., Nandi, K., Okuneye, A., ... & Gerardin, J. (2023). Application of mathematical modelling to inform national malaria intervention planning in Nigeria. *Malaria Journal, 22*(1), 137.

[17] Ogunsakin, R. E., Babalola, B. T., Olusola, J. A., Joshua, A. O., & Okpeku, M. (2024). GIS-based spatiotemporal mapping of malaria prevalence and exploration of environmental inequalities. *Parasitology Research, 123*(7), Article 262. https://doi.org/10.1007/s00436-024-08276-0

[18] James, I., & Osubor, V. (2025). Machine learning evidence towards eradication of malaria burden: A scoping review. *Applied Computer Science, 21*(1), 44–69. https://doi.org/10.35784/acs_6873

[19] Awoyelu, I., Omodunbi, T., & Udo, J. (2014). Bridging the gap in modern computing infrastructures: Issues and challenges of data warehousing and cloud computing. *Computer and Information Science, 7*(1), 33–41.