Leveraging Artificial Intelligence and Mobile Data Analytics for Enhancing Public Health Monitoring and Service Delivery in Kajiado County, Kenya

Jeff Wafubwa[1], Mapema Nelson[2], Okwiri Saad[3], Benard Wesonga[4]

[1]Umma University, Kajiado, Kenya
[2]In-depth Data Institute, Nairobi, Kenya
[3]Umma University, Kajiado, Kenya
[4]Masinde Muliro University of Science & Technology, Kakamega, Kenya

ABSTRACT

Kenyan government agencies routinely collect large volumes of health data, yet much of this information remains underutilized because of weak analytical infrastructure, resulting in disease surveillance that is largely reactive rather than proactive. Kajiado County, a vast pastoralist arid and semi-arid land (ASAL) region, exemplifies this challenge. To address this gap, a cross-sectional mixed-methods study was conducted across all five sub-counties of Kajiado between January 2025 and April 2026. The study analysed four complementary datasets comprising 450 records each: routine health facility disease records, demographic household survey data, mobile network connectivity data, and socioeconomic profiles. Statistical analyses, including descriptive statistics, independent t-tests, one-way ANOVA, Pearson correlation, and chi-square tests, were performed using Python 3.11. The findings identified malaria as the leading disease burden, with a mean of 4.62 cases per facility-month (SD = 2.51; total = 2,077 cases) and a statistically significant seasonal peak in April compared with July ($t = 2.251$, $p = 0.028$). County-wide childhood vaccination coverage averaged 83.38%, remaining below the national target of 90%, while Kajiado North recorded the lowest sub-county coverage at 82.79%. Low-income households accounted for 41.8% of the sample, and no significant relationship was observed between household income and consumption ($F = 0.497$, $p = 0.609$). Anomaly detection thresholds established at the 95th percentile were 9 cases per month for malaria, 6 for pneumonia, and 7.5 for diarrhea. These findings demonstrate that an AI-driven framework integrating Isolation Forest anomaly detection, LSTM time-series forecasting, and K-means spatial clustering within the existing DHIS2 infrastructure could transform routinely collected health data into actionable decision intelligence. With a projected return on investment of 42–58% over 36 months, the proposed framework presents a strong economic case for a 12-month pilot in Kajiado County and offers a scalable model for strengthening health surveillance across other ASAL health systems in Sub-Saharan Africa.

KEYWORDS

Artificial Intelligence, Mobile Data Analytics, Public Health Monitoring, Predictive Analytics, Arid and Semi-Arid Lands, Kenya

I. BACKGROUND

Across Sub-Saharan Africa, health systems routinely collect substantial epidemiological, demographic, and service-utilization data, yet the analytical infrastructure to convert these data into timely, actionable intelligence remains underdeveloped and fragmented [1, 2]. Disease outbreaks are detected late, resources are misallocated, and preventive interventions target populations defined by administrative boundaries rather than epidemiological need. Digital health technologies - mobile health platforms, AI-assisted analytics, and interactive decision-support dashboards - have demonstrated capacity to improve data completeness and reporting timeliness in low-and-middle-income countries (LMICs) [3, 4]. However, the specific challenges of arid and semi-arid land (ASAL) settings - mobile pastoralist populations, sparse infrastructure, unreliable connectivity, and climate sensitivity - have received limited scholarly attention.

Kajiado County in Kenya’s southern Rift Valley (population 1,117,840; area 21,901 km²) combines among the highest disease vulnerability rankings in Kenya with some of the weakest digital health analytical infrastructure [5, 6]. Five major public health facilities and approximately 1,200 Community Health Volunteers (CHVs) across five sub-counties generate monthly data that flows into the national District Health Information Software version 2 (DHIS2) platform but is rarely subjected to trend analysis, anomaly detection, or predictive modelling [7]. The predominant Maasai pastoralist population experiences above-average rates of malaria, diarrheal disease, and vaccine-preventable illness, amplified by seasonal climate variability and mobility patterns that render fixed-facility surveillance inadequate [8, 9].

This study addresses three research questions: (i) what are the patterns and seasonal or sub-county variations in communicable disease burden across Kajiado County’s health facilities? (ii) How do demographic, socioeconomic, and connectivity factors shape household health vulnerability? (iii) What statistical evidence supports the feasibility and necessity of deploying AI-assisted analytics and mobile data tools to improve public health monitoring in this setting? The study provides the first comprehensive, multi-source statistical characterization of the health information landscape in Kajiado County and makes the evidence-based case for a proposed 12-month AI-assisted pilot integrated within the existing DHIS2 infrastructure.

II. METHODS

A. *Study Design and Setting*

A cross-sectional, mixed-methods analytical design was employed, situated within a pragmatist research paradigm [10]. The study covered all five Kajiado sub-counties (Kajiado North, Central, East, South, and West) and five public health facilities: Kajiado County Referral Hospital, Kitengela Sub-County Hospital, Ongata Rongai Health Centre, Oloosokonoi Dispensary, and Kibiko Health Centre. No experimental manipulation was involved; the study analysed existing routinely generated and survey-collected data.

B. Data Sources

Four complementary datasets of 450 records each were compiled for the period January 2025 to April 2026: (1) Routine Health Facility Dataset - monthly facility records documenting malaria, pneumonia, and diarrhea case counts; (2) Demographic Household Survey Dataset — structured records capturing household size, water access, childhood vaccination rates, and maternal age across all five sub-counties; (3) Mobile Connectivity Dataset - tower-level records from five network towers capturing unique subscriber counts, call volumes, and mean population mobility (km); (4) Socioeconomic Dataset - household records profiling income group, social cohesion scores (1-5), and consumption scores (0-100). Health facility records were drawn by comprehensive enumeration of DHIS2-submitted monthly reports. Household survey participants were selected by stratified random sampling proportional to sub-county population size. Mobile tower records covered all five operational towers. Socioeconomic records were matched to the household survey sample by household identification number.

C. Statistical Analysis

All analyses were conducted in Python 3.11 (panda's v2.0, NumPy v1.25, SciPy v1.11). Descriptive statistics (means, standard deviations, frequency distributions) were computed for all variables. Group comparisons used independent-samples t-tests (two groups) and one-way ANOVA (multiple groups). Bivariate relationships were assessed by Pearson correlation. Chi-square tests evaluated categorical associations. Statistical significance was defined at $\alpha = 0.05$ (two-tailed). Anomaly detection thresholds were set at the 95th percentile of disease case distributions.

D. Ethical Considerations

All data comprised anonymised routine administrative health records or survey data collected under institutional ethical approvals compliant with NACOSTI guidelines. No personally identifiable health information was accessed. Household survey data were pseudonym zed; the pseudonymization key is held by the county data protection officer. The study complies with Kenya's Data Protection Act (2019) requirements for secondary analysis of health administrative data.

III. RESULTS

A. Routine Health Facility Disease Burden

Across the 450 facility-month records, malaria was the dominant communicable disease burden (mean 4.62 cases/report; SD 2.51; total 2,077 cases), followed by diarrhea (mean 3.61; SD 2.16; total 1,622) and pneumonia (mean 2.39; SD 1.81; total 1,076). Table 1 presents the disaggregated burden by facility.

Table 1: Routine Health Facility Disease Burden by Facility (N = 450 facility-month records, January 2025 – April 2026)

Facility	n	Malaria Mean (SD)	Total	Pneumonia Mean (SD)	Total	Diarrhea Mean (SD)	Total
Kajiado County Referral Hospital	97	5.10 (2.53)	495	2.59 (1.88)	251	3.68 (2.11)	357
Kibiko Health Centre	84	4.30 (2.48)	361	2.20 (1.60)	185	3.77 (2.25)	317
Kitengela Sub-County Hospital	105	4.71 (2.44)	495	2.42 (1.87)	254	3.64 (2.07)	382
Oloosokonoi Dispensary	83	4.61 (2.53)	383	2.30 (1.72)	191	3.63 (2.16)	301
Ongata Rongai Health Centre	81	4.23 (2.56)	343	2.41 (1.86)	195	3.27 (2.20)	265
Overall	450	4.62 (2.51)	2,077	2.39 (1.81)	1,076	3.61 (2.16)	1,622

One-way ANOVA revealed no statistically significant between-facility differences in malaria ($F = 1.767$, $p = 0.134$), pneumonia ($F = 0.718$, $p = 0.580$), or diarrhea burden ($F = 0.860$, $p = 0.488$), indicating a geographically diffuse rather than concentrated disease pattern consistent with dispersed pastoralist populations. A statistically significant seasonal malaria peak was observed in April (mean 5.63 cases) compared to July (mean 3.97; $t = 2.251$, $df = 50$, $p = 0.028$), consistent with the long rains (March–May) expanding Anopheles breeding habitat. Empirical 95th-percentile anomaly thresholds were derived as 9.0 cases/month for malaria, 6.0 for pneumonia, and 7.5 for diarrhea.

B. Demographic and Vaccination Findings

The demographic household survey covered 450 households across all five sub-counties. County-wide childhood vaccination coverage averaged 83.38% (SD 3.32%), below the national 90% target. Kajiado North recorded the lowest sub-county mean (82.79%; SD 3.55%); Kajiado East the highest (83.75%). Water access was confirmed for 80.0% of households, ranging from 78.0% (Kajiado South) to 81.6% (Kajiado West). Mean household size was 5.98 persons (SD 2.34) and mean maternal age was 27.54 years (SD 5.16). A comparison of vaccination rates between households with and without water access was marginally non-significant ($t = -1.957$, $p = 0.051$), suggesting that water access and vaccination represent analytically distinct determinants of child health. Table 2 presents the full demographic profile by sub-county.

Table 2: Demographic Profile by Sub-County (N = 450 households)

Sub-County	N	HH Size (Mean)	Water Access (%)	Vaccination Mean (%) (SD)	Maternal Age
Kajiado Central	78	5.96	80.8%	83.56 (3.18)	28.00
Kajiado East	91	5.91	79.1%	83.75 (3.31)	27.75
Kajiado North	92	6.10	80.4%	82.79 (3.55)	27.02

Kajiado South	91	5.82	78.0%	83.21 (3.42)	26.96
Kajiado West	98	6.08	81.6%	83.61 (3.11)	28.02
Overall	450	5.98	80.0%	83.38 (3.32)	27.54

C. Mobile Connectivity and Socioeconomic Profile

Mean daily unique subscribers per tower were 200.85 (SD 13.91) and mean daily calls were 998.50 (SD 30.33). Mean population mobility was 2.56 km/day (SD 2.65 km), with the remote-pastoral TOWER005 zone recording the highest mean mobility (2.81 km/day). Pearson correlation between population mobility and call volume was not significant ($r = -0.046$, $p = 0.329$), establishing that these are analytically independent signals that must both be incorporated into transmission modelling.

Low-income households constituted 41.8% of the socioeconomic sample ($n = 188$), medium-income 39.6%, and high-income 18.7%. One-way ANOVA revealed no statistically significant difference in consumption scores across income groups ($F = 0.497$, $p = 0.609$), reflecting the limitations of consumption indices in a pastoralist economy dominated by livestock ownership and non-monetized exchange. Chi-square analysis identified a marginal water access–income association ($\chi^2 = 5.633$, $df = 2$, $p = 0.060$). Table 3 summarizes the inferential statistical findings.

Table 3: Summary of Inferential Statistical Tests

Test	Variables	Statistic	p-value	Interpretation
t-test	Malaria (April vs. July)	$t = 2.251$	0.028	Significant seasonal peak in April
ANOVA	Malaria across 5 facilities	$F = 1.767$	0.134	No significant facility difference
ANOVA	Pneumonia across facilities	$F = 0.718$	0.580	Homogeneous facility burden
ANOVA	Diarrhea across facilities	$F = 0.860$	0.488	No significant facility difference
ANOVA	Consumption by income group	$F = 0.497$	0.609	Income group does not predict consumption
Pearson r	Social Cohesion vs. Consumption	$r = 0.004$	0.931	No significant association
Chi-square	Water Access vs. Income Group	$\chi^2 = 5.633$	0.060	Marginal association (near threshold)
t-test	Vaccination: water vs. no-water	$t = -1.957$	0.051	Marginally non-significant

IV. DISCUSSION

A. Principal Findings

The statistically significant seasonal malaria peak in April ($t = 2.251$, $p = 0.028$) provides robust empirical grounding for prioritizing malaria in an AI-assisted surveillance framework. This seasonal signal is directly exploitable by Long Short-Term Memory (LSTM) recurrent neural networks, which outperform traditional autoregressive models in climate-coupled disease forecasting in comparable East African contexts [11]. Empirically derived anomaly detection thresholds - 9 cases/month for malaria, 6 for pneumonia, and 7.5 for diarrhea at the 95th percentile - provide a locally calibrated alternative to the administratively defined Integrated Disease Surveillance and Response (IDSR) alert thresholds currently used in Kenya [7].

The absence of significant between-facility disease burden differences (malaria ANOVA: $F = 1.767$, $p = 0.134$) challenges facility-centric resource allocation models and argues for community-level, mobility-aware surveillance approaches [8]. The non-significant mobility–call volume correlation ($r = -0.046$, $p = 0.329$) extends methodological guidance on CDR-based mobility proxies [12], confirming that neither metric alone adequately characterizes population movement in this high-variability pastoralist context.

The marginally non-significant vaccination–water access relationship ($p = 0.051$) and the null consumption-income finding ($F = 0.497$, $p = 0.609$) together argue against single-indicator vulnerability targeting in favour of composite, multi-dimensional indices integrating health utilization, mobility, geographic accessibility, and socioeconomic data - implementable through K-means clustering.

B. Strengths and Limitations

This study's primary strengths are its multi-source analytical design integrating four complementary datasets, its empirically grounded AI calibration parameters derived from locally observed baseline distributions, and its comprehensive coverage of all five Kajiado sub-counties. Limitations include the cross-sectional design, which precludes causal inference; the restriction of health facility data to five of the county's 147 facilities; and the limited discriminatory power of consumption scores in a pastoralist economy. Mobile connectivity data represents tower-level aggregates rather than individual CDRs, limiting mobility characterization precision.

C. Comparison with Other Studies and Proposed AI Framework

Comparable AI surveillance deployments in LMICs have reported significant improvements in outbreak detection timeliness and health manager decision behaviour following dashboard-driven analytics deployment [3, 13]. Consistent with evidence from Kenya's Kisumu and Kilifi counties [14, 15], the study confirms that digital health tools can materially improve data completeness (a 15-40% improvement in comparable deployments) and reduce reporting delays from weeks to under 48 hours.

The proposed framework integrates four complementary data streams - DHIS2 facility records, ODK CHV-collected community data, household survey data, and mobile CDR-derived mobility data - within a unified analytics environment. Three AI techniques address distinct surveillance needs: (1) Isolation Forest anomaly detection [16] identifies statistical outliers without requiring labelled outbreak data, using the locally derived 95th-percentile thresholds as initialization parameters; (2) LSTM networks capture the non-linear seasonality documented in this study (the April malaria peak) and generate 3-month forward forecasts to pre-position commodities ahead of seasonal surges; (3) K-means spatial clustering assigns composite vulnerability rankings to community units from eight integrated variables, enabling differentiated CHV deployment. Dashboards built in Apache Superset deliver these outputs to sub-county health management teams in near-real time, consistent with evidence that dashboard access significantly improves data-informed resource allocation decisions [13].

A projected Return on Investment (ROI) analysis estimates KES 4.2-6.8 million in annual cost savings through faster outbreak detection, improved vaccination targeting, and optimized drug procurement. Total pilot implementation cost is estimated at KES 8.5-11.2 million, yielding a benefit-cost ratio of 1.42:1 and a break-even at 18-22 months.

D. Meaning of the Study

This study demonstrates that the barrier to AI-assisted health surveillance in Kajiado County is not data insufficiency but analytics infrastructure - a technically addressable gap within feasible budgets aligned with Kenya's national Digital Health Strategy and SDG 3 commitments. By providing empirically calibrated AI parameters, the study distinguishes its proposed framework from generic, context-free AI health tools. The multi-source data integration approach, showing cross-stream patterns invisible to single-dataset analysis, provides a scalable blueprint for ASAL health systems across Sub-Saharan Africa.

E. Unanswered Questions and Future Research

The mediation hypothesis - that seasonal population mobility partially drives the April malaria peak - was supported by temporal alignment but could not be formally tested with the available cross-sectional data; this warrants longitudinal investigation using individual-level CDRs. Algorithmic bias in AI models trained on predominantly facility-based data requires prospective evaluation for differential performance across pastoral and non-pastoral sub-populations [17]. Future research should employ interrupted time-series designs with ≥24-month follow-up to evaluate the impact of framework deployment on outbreak detection timeliness and health outcomes.

F. Limitations

The cross-sectional design precludes causal inference. Health facility records represent only five of Kajiado County's 147 facilities, potentially underrepresenting dispensary-level disease burden serving the rural pastoral majority. Consumption scores exhibited limited discriminatory power across income groups, suggesting that future data collection should incorporate asset indices and livelihood diversity measures better calibrated to pastoral economies. Mobile connectivity data are tower-level aggregates, limiting individual-level mobility precision.

V. CONCLUSION

This study provides the first comprehensive, multi-source statistical characterization of Kajiado County's health information landscape, documenting a statistically significant and analytically exploitable seasonal malaria peak ($t = 2.251$, $p = 0.028$), sub-threshold vaccination coverage, a geographically diffuse disease burden pattern consistent with pastoralist population dispersal, and a predominantly low-income socioeconomic profile requiring composite vulnerability indexing. Taken together, the four datasets reveal that routinely collected health data in this resource-constrained ASAL county yields meaningful epidemiological signals when properly analyzed. An AI-driven framework combining Isolation Forest anomaly detection, LSTM forecasting, and K-means spatial clustering - integrated within DHIS2 and delivered through interactive dashboards - can transform this analytically under-exploited data landscape into actionable decision intelligence. The empirically derived calibration parameters provided by this study establish the quantitative foundation for a proposed 12-month pilot, offering a replicable model for ASAL health systems across Sub-Saharan Africa.

What is already known on this topic?

- AI and mobile health tools have improved data completeness and outbreak detection timeliness in Sub-Saharan African LMICs.
- DHIS2 provides Kenya's national health data repository, but ASAL counties consistently report data completeness below the 90% Ministry of Health benchmark.
- Pastoralist populations in Kenya's ASAL counties face above-average communicable disease burdens that are poorly captured by fixed-facility surveillance systems.

What this study adds

- The first multi-source statistical characterization of disease burden, mobility, demographics, and socioeconomic vulnerability in Kajiado County, providing locally calibrated AI initialization parameters.
- Empirical demonstration that mobility and call volume are analytically independent signals in a pastoralist ASAL context, challenging single-metric CDR proxy approaches.
- A technically grounded, economically feasible, and ethically designed AI analytics framework for DHIS2-integrated surveillance in resource-constrained pastoralist health systems with a projected BCR of 1.42:1.

A. Competing Interests

The authors declare that they have no competing interests.

B. Authors' Contributions

MN: conceptualization, methodology, statistical analysis, AI framework design, writing - original draft. OS: study setting access, data acquisition, writing - review and editing. JL: data collection, validation, writing - review and editing. BW: methodology review, supervision, writing - review and editing. All authors read and approved the final manuscript.

C. Acknowledgements

The authors acknowledge the support of UMMA University through the 2025 Research Grant Award, which funded this study. The authors further recognize the valuable contribution of the Community Health Volunteers who participated in the data collection process and made this research possible.

REFERENCES

[1] WHO Regional Office for Africa. Digital health in the WHO African region: An overview of progress and the way forward. 2020. WHO Regional Office for Africa.

[2] Kaneko M, Yamashiro H, Kitamura K. Data quality assessment in health management information systems in sub-Saharan Africa: A systematic review. BMC Health Serv Res. 2021; 21(1):1038.

[3] Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, Agarwal S. Best practices in scaling digital health in low- and middle-income countries. Global Health. 2018; 14(1):103.

[4] Manya A, Were M, Odhiambo F, Otieno G, Shipala E, Kimani J. Implementation outcomes of an integrated electronic health record system for HIV care and treatment in Kenya. Int J Med Inform. 2022; 91:71-80.

[5] Kenya National Bureau of Statistics. Kenya Population and Housing Census Volume I: Population by County and Sub-County. 2019. KNBS.

[6] Kajiado County Government. Kajiado County Integrated Development Plan III 2023-2027. 2022. Kajiado County Government.

[7] Ministry of Health Kenya. Kenya Health Policy 2014-2030 midterm review report. 2018. Ministry of Health Kenya.

[8] Ndemwa P, Mwisukha A, Mwaniki D. Community health workers: A review of their role in health promotion and disease prevention in sub-Saharan Africa. Afr Health Sci. 2019; 19(3):2611-2621.

[9] Ministry of Devolution and ASALs. National policy for the sustainable development of northern Kenya and other arid lands. 2021. Government of Kenya.

[10] Creswell JW, Plano Clark VL. Designing and conducting mixed methods research. 3rd edition. 2018. Thousand Oaks. Sage.

[11] Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals. 2020; 135:109864.

[12] Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, Bhatt S, Brownstein JS, Collier N, Myers MF, George DB, Gething PW. Global mapping of infectious disease. Philos Trans R Soc B. 2013; 368(1614):20120250.

[13] Ndung'u P, Otieno G, Were E, Nyambura S. Assessment of the uptake of the Community Health Information System among community health volunteers in Kisumu County, Kenya. Pan Afr Med J. 2020; 35:26.

[14] Odhiambo F, Murithi P, Wango RK, Nuwagaba-Biribonwoha H, Bain LE. Strengthening health information systems and data use for malaria surveillance in Kenya: Learning from Kilifi County. Malar J. 2021; 20(1):62.

[15] Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health. 2014; 14(1):188.

[16] Liu FT, Ting KM, Zhou Z-H. Isolation-based anomaly detection. ACM Trans Knowl Discov Data. 2019; 6(1):1–39.

[17] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019; 366(6464):447–453.

[18] Republic of Kenya. The Data Protection Act 2019. 2019. National Council for Law Reporting.

[19] WHO. Ethics and governance of artificial intelligence for health: WHO guidance. 2021. WHO.
[20] Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: How can AI contribute to health in resource-poor settings? BMJ Glob Health. 2018; 3(4):e000798.
[21] Moran KR, Fairchild G, Generous N, Hickmann K, Osthus D, Priedhorsky R, Hyman J, Del Valle SY. Epidemic forecasting is messier than weather forecasting: The role of human behavior and internet data streams in epidemic forecast. J Infect Dis. 2016; 214(Suppl 4):S404–S408.
[22] Hartung C, Lerer A, Anokwa Y, Tseng C, Brunette W, Borriello G. Open Data Kit: Tools to build information services for developing regions. Proc 4th ACM/IEEE ICTD. 2010.
[23] Okolo CA, Abebe R, Aruleba K. Where are we now? Examining intersectionality and culturally responsive design in African AI research. Proc FAccT. 2021; 598–609.
[24] Mehl G, Lehmann U, Prema N, Sherber N. Developing actionable digital investment cases in Africa for universal health coverage. BMJ Glob Health. 2020.
[25] Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: Where is the evidence? PLOS Med. 2016; 10(2):e1001382.