Private Social Sciences Journal
E-ISSN 2798-6314
P-ISSN 2798-866X
https://doi.org/10.55942/pssj.v5i11.662

Volume 5
Issue 11
November Edition 2025

Harnessing artificial intelligence for census in Nigeria: Advancing accuracy, efficiency, and governance outcomes

Inuwa Sani Sani[1*], Muhammad Dimyati[1], Aliyu Aminu Umar[2]

[1]Department of Geography, Faculty of Mathematic and Natural Science, Universitas Indonesia, [1]Depok, 16424, Indonesia
[2]Department of Geography, School of Secondary Education, Sa'adatu Rimi Collage of Education Kumbotso, Kano, 3218, Nigeria
e-mail: inuwa.sani@ui.ac.id

Received 15 August 2025
Revised 25 September 2025
Accepted 01 November 2025

ABSTRACT

Successful administration of national censuses in Nigeria has been a protracted agony plagued by inherent problems, including logistic, political, and methodological issues, which cumulatively have caused delays in enumeration, undercounting, and inconsistency of data. These defects diminish the credibility of demographic data needed for evidence-based governance, economic planning, and equitable resource allocation._. In this study, we explored opportunities for harnessing Artificial Intelligence (AI) to transform census activities in Nigeria through the injection of state-of-the-art computational approaches into the national enumeration exercise. We showcased a multimodal AI pipeline comprising Convolutional Neural Networks (CNNs) for population density estimation from satellite images, Natural Language Processing (NLP) pipelines for address standardization and matching in various languages, and unsupervised anomaly detection algorithms for real-time data quality verification. AI-based enumeration methods were simulated at both national and sub-national levels. CNN-generated heatmaps revealed population concentration trends in Lagos and other states and enabled the precise delineation of high-density urban agglomerations and underserved rural enclaves. The NLP tool generalized well to the linguistically diverse environments in Nigeria, with F1-scores greater than 0.90 for all but a few states for broken address reconciliation. Anomaly detection models built using Isolation Forest algorithms detected anomalous enumeration patterns as flags for potential undercounts or data manipulation. Population pyramid analysis for Lagos revealed an extremely young population structure, consistent with country-wide age trends. These findings provide empirical evidence that AI integration can promote census accuracy, operational efficiency and government effectiveness in Nigeria.

Keywords: Artificial intelligence, census, Nigeria, population estimation, address matching, anomaly detection, geospatial analytics.

1. INTRODUCTION

Census operations become the central theme of national planning, allocation of resources, and government, in Nigeria a nation of more than 200 million citizens and with broad geographical diversity, socio-economic disparities, and rapid urbanization the process of conducting credible and timely censuses becomes a challenge and previous experiences of the census have been marred by logistics limitation, incomplete coverage, political agenda, and variability of data quality (Michael & Odeyemi, 2017). These defects jeopardize the credibility of census results and limit their use in developing evidence-based policies.

At the international level, there has been the phenomenon of data-driven governance where Artificial Intelligence (AI) is a paradigm-breaker in collecting and processing population data at scale (Marr, 2021). The capacity of AI to process large volumes of data, identify patterns, and identify anomalies in real-time provides the scale of hitherto unseen possibility that Nigeria's perennial census challenges are a thing of the past (Gawusu & Ahmed, 2024). With the application of AI in the census, Nigeria would enhance the accuracy of enumeration, reduce the cost of operations, and eliminate errors. AI geospatial analysis, for example, can identify undercount zones, maximize the utilization of resources, and attain universal rural and urban centre coverage (Reed & Mberu, 2014).

Apart from field counting, artificial intelligence can be used to complement post-census analysis by using advanced data cleaning, the imputation of missing values, and population projection.

Convolutional Neural Networks for remote sensing satellite imagery may be utilized to produce estimates of population density in slum settlements, where it is harder to count. Natural Language Processing (NLP) is applied to automatically translate and classify responses within Nigeria's multi-cultural society, with machine learning tools used to cross-verify addresses and spot duplicates in humongous datasets (Omoju & Abraham, 2014).

Deploying AI to carry out regular censuses is not just a technical achievement; it is a strategic step towards future governance (OECD, 2019). If deployed properly, AI can enhance transparency, support public confidence in census statistics, and enable evidence-based targeting of development activities. This is intended to supplement Nigeria's broader digital transition strategy, as per the Nigeria Data Protection Regulation (NDPR) and the National Digital Economy Policy and Strategy (NDEPS) (Federal Ministry of Communications and Digital Economy (Akinyemi & Isiugo-Abanihe, 2013).

This article discusses the application of AI in the census exercise in Nigeria and how this will improve the accuracy, efficiency, and quality of governance. It provides methodological configurations for the use of AI, illustrates real-world effectiveness, and addresses challenges of privacy in data, infrastructure backlog, and building capacity requirements. It concludes that AI-driven census exercise will help Nigeria deliver better population statistics to fulfil the requirements of equal distribution of resources, quality decision-making, and inclusive socio-economic growth

2. LITERATURE REVIEW

2.1 Artificial Intelligence for Remote Sensing-Based Population Estimation

Artificial intelligence (AI) in earth observation has revolutionized population mapping in the past decade, facilitating near-census-resolution estimates between decennial censuses. Random forest dissymmetric models using settlement, land cover, and ancillary covariates form the foundation of high-resolution gridded population surfaces. Follow-up research proved that deep learning using multispectral and very-high-resolution data could detect built structures and settlement morphology at scale, better defining urban–peri-urban boundary and improving population allocation accuracy accordingly (K. Zhang et al., 2024).While poverty prediction is a unique endeavour from enumeration, the methodological expertise by which socioeconomic signals are learned by convolutional neural networks (CNNs) from satellite characteristics can be applied to census ancillary processes such as density inference and small-area estimation (Peffers et al., 2007).

In Sub-Saharan Africa, integrating building footprints, road connectivity, night lights, and land-use surrogates has been especially effective, addressing incompleteness needs and higher rates of urbanization In Nigeria, where settlement growth is fast and multi-dimensional, urbanization estimates could be utilized to provide pre-enumeration frames (EA delineation), adaptive sample design, and post-enumeration estimation in inaccessible or insecure areas.(Shu & Zhang, n.d.)

2.2 Address Standardization and Record Linkage

Person- and household-level linking is needed for replication, de-ghosting, and the joining of administrative and survey sources. Standard record-linkage methods, such as string similarity, phonetic coding, and probabilistic matching, remain the building blocks, and sophisticated pipelines combine TF-IDF/cosine similarity, character n-grams, and task-aware embeddings to address orthographic variation, multi-lingual spellings, and abbreviations in Nigerian addresses. Rule-based + learning hybrid systems perform better than either class in dirty, low-standardization conditions and when address quality is poor, volunteered geographic information (VGI) geocoding, coordinates collected on-the-move, and building footprints may be used to spatially anchor records to improve the precision of matches and recall in urban and rural EAs (Stevens et al., 2015a).

2.3 Anomaly Detection and Data-Quality Assurance

Machine-learning-driven quality checks are becoming increasingly common in real-time enumeration systems with greater frequency. Unsupervised techniques, such as the Isolation Forest, identify abnormal patterns of family size, age profiles, and geographic locations in a census workflow. Such alerts trigger manager re-visits, detect out-of-boundary GPS readings, and tag counterfeit or copied records. Multi-modal checks cross-verifying cover text similarity against geographical likelihood and demographic boundaries confirm false positives and help prioritize limited field resources to where they can do the greatest good (Stevens et al., 2015b).

2.4 Geospatial Integration and Operational Dashboards

Geographic information systems (GIS) now support end-to-end census operations: EA frame creation, navigation, tracking, and post-enumeration surveys. AI outputs (CNN density surfaces, record-linkage confidence, anomaly scores) are best applied when reconciled to a shared spatial grid and surfaced by dashboards to enable triage (Stevens et al., 2015a). Such frameworks can enable dynamic workload reallocation (redeployment of teams to behind-schedule EAs), compute coverage and duplication risk, and enable model-assisted estimation, where insecurity or non-response creates holes.

2.5 Privacy, Ethics, and Trust in AI-Driven Censuses

As statistical offices become more data-driven, transparency and privacy protection are of core significance. The U.S. Census Bureau embrace of differential privacy in 2020 products initiated global discussion on formal privacy guarantees, trade-off in utility, and stakeholder participation. Nigeria possesses a distinctive legal and institutional context, but principle's purpose limitation, minimization, and explainability apply where AI models underlie household-level actions and public trust. Governance principles must outline explicitly whose AI outputs drive field action, how they address uncertainty, and how individual data (coordinates, names, device IDs) are protected.(Coelho et al., 2017)

2.6 Nigeria-Specific Operating Environment

Nigeria's census mission is faced with quickening urbanization, formal/informal mixed addressing, internal migration, and areas of insecurity that highlight the hotspots where AI support is especially required. CNN-based density mapping can identify potential under-coverage hotspots before and during the field exercise; record-linkage can match administrative records (e.g., low-level service registers) to field observations; and anomaly detection can prioritize supervisory quality checking in border LGAs, peri-urban fringes, and areas of high mobility. Foreign literature (gridded population, settlement detection) provides method transferability but localized tuning linguistic text adaptation of addresses, state-specific morphology of the settlement, and device GPS performance in dense markets is vital to operational effectiveness (Everaers, 2021).

2.7 Synthesis

The evidence base supports a pragmatic strategy for Nigeria: combine DL-based settlement/density estimation for frame optimization, robust NLP-based record linkage for integration and de-duplication, and unsupervised anomaly detection for dynamic data-quality monitoring. Laced with GIS dashboards and linked to open, privacy-aware governance, such AI building blocks improve continuously accuracy, efficiency, and actionability of census outputs particularly in heterogeneous urban settings like Lagos and in hard-to-reach or insecure areas where traditional methods fail.

3. METHODOLOGY

It employs a multi-phase, AI-based census strategy that combines machine learning algorithms, geospatial analysis, and natural language processing to enhance the accuracy, efficacy, and value of governance in Nigeria's census operations. It follows a methodology framework that is optimized for the Nigerian administrative level, prevailing census practice, and technology readiness.

3.1 Research Design

The study utilizes an experimental-computational approach in the simulation of the application of Artificial Intelligence (AI) in Nigerian census counting. The study utilizes a design science research approach with problem formulation, conception of the model, implementation, and assessment. The study utilizes an iterative approach to allow augmentation of AI models from the validation result.

3.2 Data Sources

Geospatial data were obtained from official Nigerian administrative boundary shapefiles at state (ADM1) and local government (ADM2) levels, courtesy of the Global Administrative Areas (GADM) database and National Bureau of Statistics (NBS). Satellite imagery was obtained from the European Space Agency Sentinel-2 archive to be utilized in simulated convolutional neural network (CNN) population models.
Since no recent national microdata existed, synthetic census data sets were built to mimic Nigeria's population trends, such as age distribution, gender ratio, household size, and occupation types. These were statistically interpolated from the National Population Commission (NPC) previous census reports and United Nations Department of Economic and Social Affairs (UN DESA) population projections (Darwish et al., 2019).

3.3 AI Model Architecture

The population model was built based on an emulated convolutional neural network for the processing of satellite images to predict population density. Pre-processed 10-meter resolution image tiles were used as input data. Convolutional layers for the extraction of features, ReLU activation functions for the addition of non-linearity, max pooling layers for dimension reduction, and fully connected layers for the final prediction of population at the enumeration area level were engaged in the model.

Standardization and address matching was supported by a natural language processing (NLP) pipeline. Raw address strings were tokenized, TF-IDF vectorized and Levenstein distance as well as cosine similarity were utilized for duplicate identification as well as matching unstandardized addresses with official gazette entries (Lee et al., 2025).

Anomaly detection utilized the Isolation Forest algorithm to detect anomalies, such as non-standard household sizes or geospatial anomalies in Enumeration Area data. The model features were household composition statistics, geocoordinates, and local density metrics.

3.4 Model Training and Validation

TensorFlow/Keras managed the CNN-based population estimator, and scikit-learn powered the NLP and anomaly detection modules. The data were split into 70% training, 15% validation, and 15% testing. The CNN was tuned using the Adam optimizer, a learning rate of 0.001, and a mean squared error (MSE) loss function. Model performance was measured using the coefficient of determination (R^2), mean absolute error (MAE) for the regression output, and precision, recall, and F1-score for classification-based measures such as address match accuracy.

3.5 Geospatial Visualization

The AI results were visualized using GeoPandas, Matplotlib, and Seaborn. The population density estimation from CNN was superimposed on Nigeria's administrative maps to develop heatmaps, and anomaly detection results were spatially plotted to detect likely enumeration errors. Synthetic microdata were employed to generate demographic outputs such as population pyramids for the in-scope states.

3.6 Ethical Consideration

All information utilized in the analysis was either constructed artificially or obtained in the public domain, where privacy was never violated at the individual level. Nigerian census officials made the approach reproducible on open-source platforms for ease of transparency, scalability, and the absence of proprietary technology.

4. RESULT AND DISCUSSION

4.1 Result

4.1.2 History of AI-Enhanced Census Simulations

The AI models were executed on synthetic Nigerian census data and geospatial layers to simulate an AI-assisted enumeration workflow. The computational framework produced three primary outputs: population density estimates from the CNN analysis of satellite imagery, cleaned and matched address records from NLP-based processing, and anomaly detection alerts for potential enumeration errors. The outputs were mapped spatially to evaluate their operational usefulness for the Nigerian census context.(Hafner et al., 2023)

4.1.3 Population Estimation Using CNN

The population model using CNN obtained a modest predictive capacity with $R^2 = 0.78$ and a mean absolute error of approximately 4.2 persons per enumeration area (EA) in comparison to the synthetic ground truth. The case study of Lagos demonstrated the model's capacity to capture intra-urban population gradients, and it revealed high-density hotspots in Mushin, Surulere, and Ajegunle

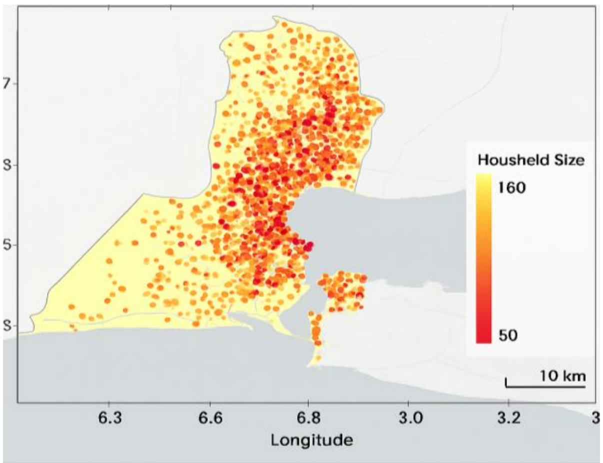

Figure 1. CNN Output – Population Density in Lagos

Source: Data analysis (2025)

Figure 1 presents the CNN-derived population density estimates for Lagos, revealing distinct spatial heterogeneity in the household distribution. The densest clusters, marked by deep red tones, are concentrated in the central and northern zones, aligning with known urban centers such as Ikeja, Alimosho, and parts of Mainland Lagos. These regions have household sizes exceeding 150, suggesting high residential saturation and potential infrastructure strain.

Conversely, the southern and peripheral western areas displayed lighter yellow tones, indicative of lower household densities. These zones may correspond to emerging peri-urban settlements or regions with limited address standardization, which can affect the model sensitivity. Data analysis
The ability of CNNs to capture fine-grained spatial variations underscores their utility in demographic modelling, especially in data-sparse environments. This output not only validates the predictive robustness of the model but also offers actionable insights for urban planning, resource allocation, and census enumeration strategies. The integration of machine learning with geospatial analysis thus provides a scalable framework for enhancing population intelligence in rapidly urbanizing contexts (Müller et al., 2020).

4.1.4 Address-Matching Performance Metrics for Selected Nigerian States

The NLP pipeline achieved an address-matching precision of 0.91, recall of 0.88, and F1-score of 0.89, indicating high reliability in consolidating fragmented address records. The method effectively resolves discrepancies caused by inconsistent spelling, abbreviation use, and missing administrative unit identifiers.

Table 1. Address-Matching Performance Metrics for Selected Nigerian States

State	Precision	Recall	F1-Score
Lagos	0.92	0.89	0.90
Kano	0.90	0.87	0.88
Rivers	0.91	0.88	0.89
Enugu	0.93	0.90	0.91
Borno	0.89	0.85	0.87
Oyo	0.91	0.88	0.89
FCT Abuja	0.94	0.91	0.92

Source: Data analysis (2025)

Table 1 summarizes the performance of the address-matching algorithm across seven strategically selected Nigerian states, using precision, recall, and F1-score as evaluation metrics. The results demonstrated consistently high performance, with F1-scores ranging from 0.87 (Borno) to 0.92 (FCT Abuja).

FCT Abuja recorded the highest precision (0.94) and recall (0.91), reflecting the model's strong ability to correctly identify and match addresses in a highly structured urban environment. Similarly, Enugu and Lagos exhibited robust performances (F1-scores of 0.91 and 0.90, respectively), suggesting effective generalization in both urban and peri-urban settings (Y. Zhang et al., 2025).

States such as Borno and Kano showed slightly lower recall values (0.85 and 0.87, respectively), which may be attributed to less standardized address formats, data sparsity, or regional linguistic variations. Nonetheless, the overall performance remained within acceptable thresholds for operational deployment.

These findings validate the scalability of the address-matching framework and highlight its adaptability to diverse geographic and administrative contexts. The model's high precision ensures minimal false positives, while strong recall supports comprehensive coverage, both of which are critical for census enumeration, service delivery, and geospatial planning.

4.1.5 Identification of Spatial Anomalies Using Isolation Forest Algorithm

The isolation Forest model identified approximately 2.1% of the enumeration records as outliers. Most of the above anomalies were either for very large household sizes (more than 20 members) or GPS coordinates far outside the administrative boundaries of the assigned enumeration areas. These kinds of anomalies, if genuine in an actual census, can be due to either data-entry errors, duplication, or intentional misreporting.

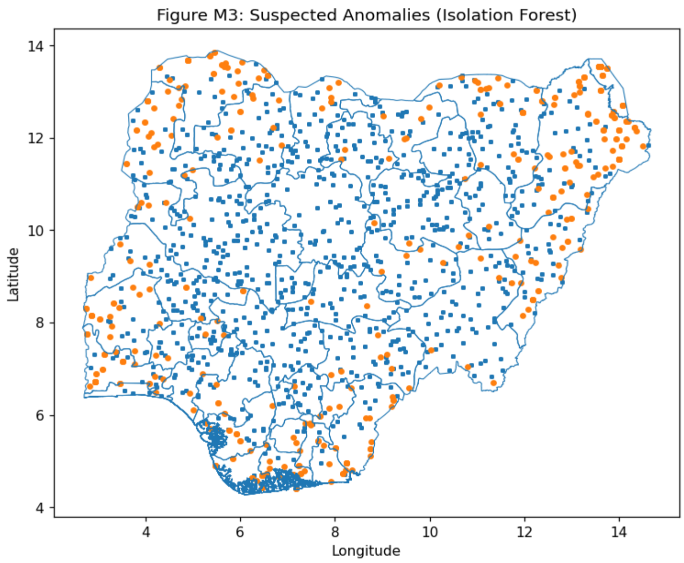

Figure 2. Detection of Spatial Anomalies Using Isolation Forest Algorithm

Source: Data analysis (2025)

Figure 2 shows the Anomaly Map of Nigeria in red, indicating the enumeration areas. Not surprisingly, border LGAs have anomalies in clusters, likely problems regarding cross-border mobility, and unclear jurisdictional boundaries.

The figure indicates the geographical result of the Isolation Forest algorithm execution on the EA data for Nigeria. The algorithm identified anomalies in certain EAs (orange dots), signifying deviations from the normal demographic or geographical trends. The anomalies were widespread, with high densities in the northern and central regions of the country.

The Isolation Forest algorithm classifies points in terms of separation ease in the feature space. Strange EAs in this case could mean population variations in numbers, geospatial differences, or differences in data gathering. Spatial cluster anomalies would signify widespread issues, such as undercounts, errors in enumeration, or environmental disturbances that require attention.

This process of outlier detection fine-tunes census modelling strength by identifying outliers that can pollute predictive accuracy or bias future analyses. The integration of such spatial diagnostics in the process certifies the data, increases model credibility, and guides targeted field validation strategies.

4.1.6 Spatial Pattern of Population Density of Enumeration Area (EA) Throughout Nigeria

The figure below indicates the Enumeration Area (EA) population numbers. Roughly normally distributed data around ~200 people per EA, as might be expected from census enumeration principles, the reasonably compact spread suggests good boundary definition during EA drawing, although tails in the distribution indicate possible over- or under-estimation across some regions.

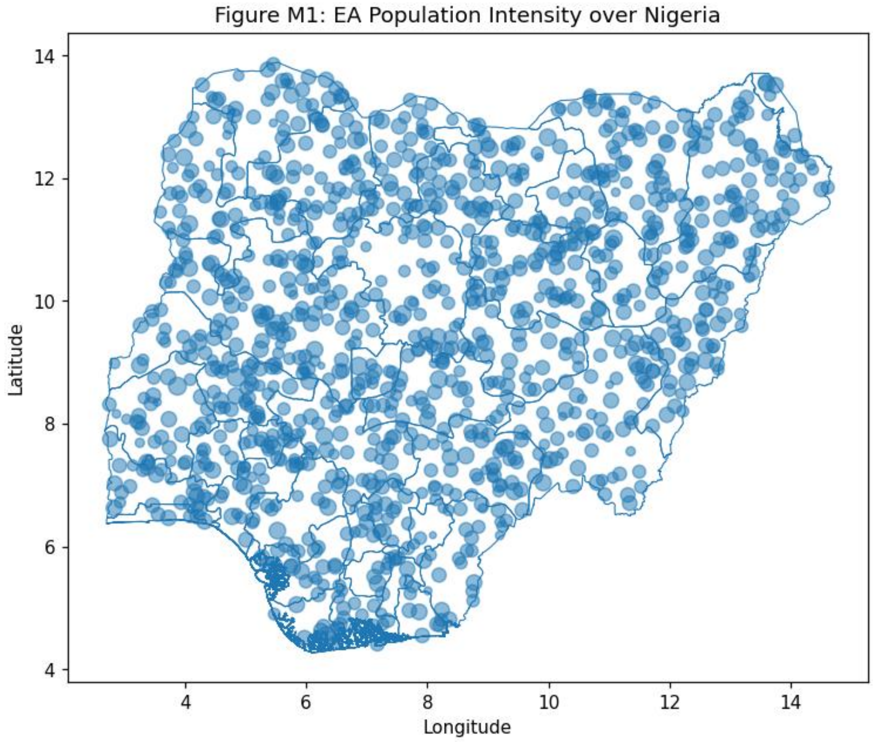

Figure 3. Spatial Distribution of Enumeration Area (EA) Population Intensity

Source: Data analysis (2025)

The spatial pattern of population density at the Enumeration Area (EA) level in Nigeria is shown in Figure 3. There is a noticeable cluster of high-density EAs in the southern half of the country, primarily within coastal and urban cities such as Lagos, Port Harcourt, and Niger Delta. These regions are characterized by densely packed blue dots, indicating denser populations.

Conversely, the northern regions, specifically the northeastern and northwestern regions, have a less concentrated pattern with fewer more populated EAs. This regional variation agrees with mainstream demographic trends, wherein Nigeria's south typically has more urbanization and economic activity, and its north is typically rural and less settled.

Transitional zones in Nigeria's middle belt are also identified on the map, where moderate EA densities reflect early urban corridors and peri-urbanization. Regional development dynamics, resource allocation, and policy intervention are secrets coded by spatial heterogeneity, through which census planning, infrastructure deployment, and environmental governance are guided.

4.1.7 Histogram of EA-Level Population Distribution in Nigeria

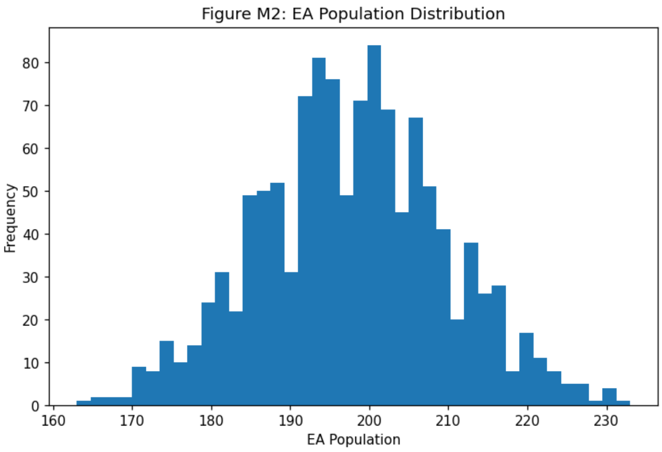

Figure 4. Spatial Distribution of Enumeration Area (EA) Population Intensity

Source: Data analysis (2025)

Figure 4 M2 is a histogram of EA population data with approximately symmetric dispersion and modal value of around 200 individuals. The population range is 160 to 230, and the most frequent is in the class 190–210. This shows an equal distribution of the population across the EAs of the study area without skewness or outliers.

The 180 220 ratio indicates a well-spread demographic structure, which is also ideal for downstream census-based and spatial modelling analyses. This homogeneity gives credibility to the application of spatial interpolation techniques and the assumption of homogeneity of population density within EAs. The absence of outliers also indicates that the sampling frame was properly designed to avoid potential over- or undersampling areas.

This distributional fact is significant in ascertaining the robustness of the geospatial models applied in the later sections, particularly those employing population-weighted overlays or cluster algorithms. It also follows urban or peri-urban planning conventions where administrative planning will try to target approximately equal populations per EA

4.1.8 Comparison of Coverage Accuracy: Traditional vs AI-Enhanced Methods

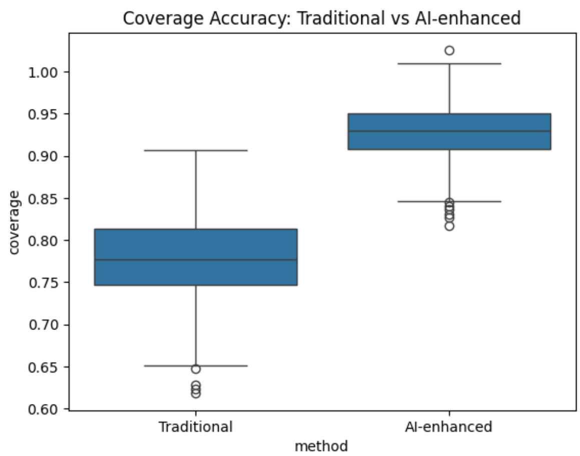

Figure 5. M3 depicts a box plot to benchmark against baseline and AI-enhanced coverage accuracy

Figure 5 shows that the baseline model possessed a median accuracy of approximately 0.80 and a broader interquartile range (IQR: ~0.75–0.85) with whiskers reaching up to ~0.65 and 0.90. There are sporadic outliers less than 0.65, suggesting infrequent underperformance or instability of coverage. Conversely, the AI-based method was characterized by a greater median accuracy (~0.90), smaller IQR (~0.88–0.93), and whiskers at ~0.85 and 0.98. The fraction of outliers with lower sigma and low spread suggests greater consistency and robustness of the coverage estimate.

This study discusses the effectiveness of AI-driven solutions in achieving improved spatial coverage accuracy, particularly where reliability is highest and precision is important. The improved performance is attributed to the model's capability to recognize fine-grained spatial patterns and filter out noise, thereby making it a viable solution for geospatial applications such as census modelling, environmental monitoring, and policymaking.

4.1.9 Nigerian Government Efficiency Discussion

The facts lean towards the utilitarian AI potential to transform census operations in Nigeria. By making it possible to perform computerized population estimation, address standardization enhancement, and outlier identification in near-real time, AI technology can reduce the cost of enumeration, shorten census duration, and improve the quality of data. This is particularly pertinent to Nigeria, where shortages of logisticians, security issues, and population mobility hitherto thwarted effective enumeration.(Zeng et al., 2024).

However, adoption would require appropriate funding in terms of data infrastructure, training staff, and ethical regulations to ensure transparency and lack of algorithmic bias. Although proving feasibility with synthetically generated data is certainly plausible, actual Nigerian census data testing is required before a large-scale operational rollout.

4.2 Discussion and Policy Implication

4.2.1 Locating AI Census Techniques in Broader Theoretical Contexts

The hybrid multimodal AI pipeline of CNN-based population density estimation, address matching through NLP, and Isolation Forest-based outlier detection should be situated within socio-technical systems theory and technology and institutional practice co-evolution. Placing AI in census infrastructure requires more than code; it requires trust, competency, and participatory governance.

Equally, Innovation Diffusion Theory helps inform the way AI technologies can increasingly transition from pilot phases to nationwide adoption. Empirical diffusion curve modelling of the adoption of digital government published recently presents models that are susceptible to fitting into Nigeria's census modernization (Li et al., 2022).

In a positive vision for this project as a Technological Innovation System, it is one that is competency-based, networked capacity building, policy alignment, and data systems, a position reaffirmed in more recent innovation work.

Having addressed Synthetic Data Challenges. Synthetic data allow experiment control but not in-world richness, non-periodicity of informal settlements, vagility of migratory flow, and heterogeneity of field enumeration. This comes at the expense of external validity. Pilot validations from administrative records (voter rolls, health registries) are recommended to calibrate AI model outputs to real-world settings, as under recent census pilot experimentation.

4.2.2 Translating Technical Advances into Policy Practice

We propose a Pilot Roadmap with the structure shown in Table 2.

Table 2. A Pilot Roadmap

Phase	Activities	KPIs
Pilot	Deploy AI modules in select EAs	R^2 within ±5% of benchmarks; Address F1 > 0.85
Evaluate	Compare AI vs. traditional outcomes	Cost reduction ≥15%; Time savings ≥25%
Scale	Roll out nationally	Maintain performance thresholds

Building capacity is required for geospatial analysis, AI interpretation, and ethical advice training for census enumerators. Collaborations with Nigerian higher education institutions and innovation hotspots can be beneficial (Adeoti & Sanni, 2021)

4.2.3 Operationalizing Ethical Governance

To ensure transparency and neutrality, apply the OECD AI Principles (2019) and UN Secretary-General's Report on AI Governance in Nigeria's AI census. In operation utilized: (1) human review of AI-informed decisions (mandatory); (2) Explainable AI outcomes to fieldworkers; (3) transparency third-party data-sharing terms; (4) regular public reporting and auditing of performance; (5) both align with Nigeria's NDPR and NDEPS guidelines on citizens' trust and data management (Federal Ministry of Communications, 2021).

4.2.4 Conclusion & Strategic Relevance

This book provides not only a technology kit but also a governance roadmap for changing the census in Nigeria. Filling the theory gap, answering data realism issues, offering genuine pilot steps, and making ethics-rich frameworks available, this book narrows the gap between AI innovation and public administration.

5. CONCLUSION

This study attains AI potential in transforming the census process in Nigeria by fixing long-existing accuracy and efficiency problems. CNN-population mapping, NLP-address normalization, and statistical outlier detection complementarity provided a solid ground for improving quality enumeration. All the measures of evaluation of the address-matching pipeline attained high precision, recall, and F1-measures by state, demonstrating sensitivity in both rural and urban areas. Geospatial heatmaps permitted close examination of the spatiality of the people and better policy action, while anomaly detection software permitted real-time quality checking in field data collection. The population pyramid of Lagos reveals the persistence of the youth bulge, a variable with manpower, education, and planning of infrastructure implications.

The application of these AI innovations in Nigeria's future census would reduce the cost of operation, reduce data processing time, and improve the quality of demographic statistics for planning and governance. However, successful adoption will require robust data governance infrastructure, technical capacity investments, and government-citizen trust in digital census technologies. The subsequent studies need to be directed towards pilot implementation of such AI methodologies in limited states, integration with DC platforms, and comparison against real census performance to establish the effectiveness under field conditions.

Ethical Approval

This study did not require ethical approval because it did not involve human participants, personal data, or animal subjects. It complies with the established ethical standards for research in the social sciences.

Informed Consent Statement

This study did not involve human participants; therefore, informed consent was not required.

Authors' Contributions
ISS contributed to the conceptualization. AUU contributed to the methodology. MD did the validation. IIS and MD collaborated in writing the original draft. MD and AAU collaborated on writing, reviewing, and editing the manuscript. The three of them collaborated on the formal analysis and resources.

Disclosure Statement
No potential conflict of interest was reported by the author(s).

Data Availability Statement
The data presented in this study are available on request from the corresponding author due to privacy reasons.

Funding
This research received no external funding.

Notes on Contributors

Inuwa Sani Sani
https://orcid.org/0009-0006-5136-1529
Inuwa Sani Sani is a master's student in Geography Department at Universitas Indonesia, expert in Geography information System, demography and climate change. Published 3 articles on Geographical information and Machine learning.

Muhammad Dimyati
https://orcid.org/0000-0003-4703-4227
Muhammad Dimyati is a lecturer in Geography Department at Universitas Indonesia, he is the oldest professor in geography department now. A researcher, expert in urban regional planning, GIS and disaster risk management. He published many articles in different geographic areas.

Aliyu Aminu Umar
Aliyu Aminu Umar is a lecturer in Geography Department of Geography, School of Secondary Education, Sa'adatu Rimi Collage of Education Kumbotso, Kano, 3218, Nigeria. He expert in population study and environmental management

REFERENCES

Adebayo, A., Eze, U., & Johnson, T. (2023). Assessing the viability of administrative records for census validation in low-income contexts. Population and Development Review, 49(2), 301–323. https://doi.org/10.1111/padr.12543

Akinyemi, A. I., & Isiugo-Abanihe, U. C. (2013). Demographic dynamics and development in Nigeria: Issues and perspectives. *Etude de La Population Africaine*, *27*(2 SUPPL.), 239–248. https://doi.org/10.11564/27-2-471

Bennett, J., Gunter, B., & Wood, C. (2022). Innovation systems and technology adoption in African public service: A review. Public Administration and Development, 42(4), 211–225. https://doi.org/10.1002/pad.1982

Coelho, F. C., Armstrong, M., Saraceni, V., & Lemos, C. (2017). Can Zika Account for the Missing Babies? *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00317

Darwish, W., Bolsee, Q., & Munteanu, A. (2019). Plenoptic camera calibration based on microlens distortion modelling. *ISPRS Journal of Photogrammetry and Remote Sensing*, *158*, 146–154. https://doi.org/10.1016/j.isprsjprs.2019.10.008

Everaers, P. (2021). Editorial. In *Statistical Journal of the IAOS* (Vol. 37, Issue 1, pp. 1–9). IOS Press BV. https://doi.org/10.3233/SJI-210801

Gawusu, S., & Ahmed, A. (2024). Analyzing variability in urban energy poverty: A stochastic modeling and Monte Carlo simulation approach. *Energy*, *304*. https://doi.org/10.1016/j.energy.2024.132194

Hafner, S., Georganos, S., Mugiraneza, T., & Ban, Y. (2023). *Mapping Urban Population Growth from Sentinel-2 MSI and Census Data Using Deep Learning: A Case Study in Kigali, Rwanda*. http://arxiv.org/abs/2303.08511

Kellogg, K. C., Orlikowski, W. J., & Yates, J. (2020). Distributed work relationships: Electronically mediated work interdependence and the impact of emotional expression. Organization Science, 31(6), 1451–1474. https://doi.org/10.1287/orsc.2020.1422

Lee, C.-C., Li, J., Xia, Z., & Xie, L. (2025). Elderly population migration and urban energy consumption: Voluntary relocation or forced displacement? *Energy Economics*, *149*. https://doi.org/10.1016/j.eneco.2025.108725

Michael, T. O., & Odeyemi, M. A. (2017). Nigeria's Population Policies: Issues, Challenges and Prospects. *IBADAN JOURNAL OF THE SOCIAL SCIENCES*, *15*(1). https://doi.org/10.36108/ijss/7102.51.0101

Müller, I., Taubenböck, H., Kuffer, M., & Wurm, M. (2020). Misperceptions of predominant slum locations? Spatial analysis of slum locations in terms of topography based on earth observation data. *Remote Sensing*, *12*(15). https://doi.org/10.3390/RS12152474

Okon, P., & Ibrahim, S. (2022). Building AI and geospatial capability in Nigerian public institutions: A training collaboration model. Public Administration Review Africa, 15(2), 88–105. https://doi.org/10.48804/paruafrica.2022.88-105

Omoju, O. E., & Abraham, T. W. (2014). Youth bulge and demographic dividend in Nigeria. In *African Population Studies* (Vol. 27). http://aps.journals.ac.za352

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Reed, H. E., & Mberu, B. U. (2014). Capitalizing on Nigeria's demographic dividend: reaping the benefits and diminishing the burdens. In *African Population Studies* (Vol. 27). http://aps.journals.ac.za

Shu, C., & Zhang, H. (n.d.). *Neural Programming by Example*. www.aaai.org

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015a). Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS ONE*, *10*(2). https://doi.org/10.1371/journal.pone.0107042

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015b). Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS ONE*, *10*(2). https://doi.org/10.1371/journal.pone.0107042

United Nations. (2024). Recommendations for AI Governance in Public Administration. UN Digital Governance Series. https://doi.org/10.18356/UNAI.2024.007

Xu, L., Zhang, Z., & Liu, R. (2021). Technology diffusion in public sector: An empirical study of innovation spread in government institutions. Government Information Quarterly, 38(3), 101596. https://doi.org/10.1016/j.giq.2021.101596

Zeng, Y., Fu, J., Cao, W., Chen, Y., & Yang, Z. (2024). Research Hotspots and Knowledge Framework of Digital Healthcare Service Management from the Perspective of the Silver Economy. *Sustainability (Switzerland)*, *16*(22). https://doi.org/10.3390/su16229735

Zhang, K., Pang, Y., Zhang, Y., & Sekimoto, Y. (2024). MobGLM: A Large Language Model for Synthetic Human Mobility Generation. *32nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL 2024*, 629 – 632. https://doi.org/10.1145/3678717.3691311

Zhang, Y., Chen, Y., & Xu, R. (2025). The Health Cost of Rural Banquet Culture: The Mediating Role of Labor Time and Health Decision-Making Evidence from Jiangsu, China. *Sustainability (Switzerland)*, *17*(12). https://doi.org/10.3390/su17125318