Vokasi Unesa Bulletin of Engineering, Technology and Applied Science (VUBETA)
https://journal.unesa.ac.id/index.php/vubeta
Vol. 3, No. 3, 2026, pp. 461~480 DOI: 10.26740/vubeta.v3i3.47643
ISSN: 3064-0768

.

Data-Driven Modeling and Predictive Analysis of Internal Migration and Displacement Trends in Nigeria for Enhanced Humanitarian Planning

Ajayi Ore-Ofe[1*], Muhammad Bashir Yahya[2], Shehu Mohammed Yusuf[1], Abubakar Umar[1], Hassan Maharazu[3], Ibrahim Ibrahim[1], Murtala Abduljalil Ahmad[4]
[1]Department of Computer Engineering, Faculty of Engineering, Ahmadu Bello University, Zaria, Nigeria
[2]Nigerian Immigration Service Headquarters Sauka, Abuja.
[3]Department of Electrical and Electronics Engineering, Federal University of Transportation, Daura, Katsina State Nigeria
[4]Electrical Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

Article Info

Article history:

Received November 25, 2026
Revised April 07, 2026
Accepted May 15, 2026

Keywords:

Communal Clashes
Insurgency
Internal Displacement
Internal Migration
Predictive Analysis

ABSTRACT

Internal displacement in Nigeria remains a major humanitarian challenge driven by conflict, insurgency, communal clashes, and climate-related disasters. Existing responses are often reactive, highlighting the need for data-driven approaches to support proactive humanitarian planning. This study, titled Data-Driven Modeling and Predictive Analysis of Internal Migration and Displacement Trends in Nigeria for Enhanced Humanitarian Planning, develops a machine learning framework to analyze and predict displacement trends using data from the International Organization for Migration (IOM) Displacement Tracking Matrix, particularly the Baseline Assessment and Needs Monitoring surveys. The dataset comprised over 114,000 records and 19 variables, including demographic characteristics, site accessibility, household size, and priority needs. Data preprocessing involved missing-value handling, categorical standardization, one-hot encoding, and numerical normalization. The highly imbalanced distribution of displacement causes was also considered, with insurgency accounting for 95% of cases, communal clashes 4%, and natural disasters 1%. Exploratory analysis identified Borno, Adamawa, and Yobe as the most affected states, with insurgency consistently dominating and September 2022 recording the highest displacement peak. Logistic Regression, Random Forest, and XGBoost were evaluated. XGBoost achieved the best performance, obtaining a macro F1-score of 0.96 and balanced accuracy of 0.968. It demonstrated strong classification performance for communal clashes (F1 = 0.99) and natural disasters (F1 = 0.90), while correctly identifying nearly all insurgency cases. These findings demonstrate that displacement patterns can be effectively modeled using socio-economic and geographic indicators. The proposed framework can assist humanitarian agencies in anticipating displacement, optimizing resource allocation, and strengthening early-warning systems to improve crisis preparedness and resilience..

1. INTRODUCTION

Internal migration and displacement are longstanding features of human societies, but their expression in Africa has been especially pronounced due to recurrent conflict, political violence, and structural vulnerabilities. Sub-Saharan Africa hosts some of the most protracted displacement situations globally, with the combined effects of emerging conflicts and climate-related hazards driving millions from their homes in 2020—yet many such crises remain under-recognized and under-funded. Nigeria exemplifies these patterns:

*Corresponding Author
Email: ajayi.oreofe17@gmail.com

from the late-colonial era through the 1967–1970 civil war to the present, waves of forced movement have recurred, propelled by ethnoreligious and communal violence alongside other human-induced and natural shocks [1].

By the close of 2018, an estimated 41.3 million people had been forced to leave their homes due to conflict and violence within their own countries, while natural disasters were responsible for displacing an additional 20 to 25 million individuals annually. Projections from the Global Displacement Forecast, which assessed 26 countries, indicated that displacement would continue to rise, with an expected increase of about 2.9 million people in 2022 and a further 3.9 million in 2023 [2]. Nigeria has been plagued by diverse crises cutting across its regions, including Boko Haram insurgency in the North (as illustrated in Figure 1), separatist agitations in the East, widespread kidnappings and militancy in the oil-rich Niger Delta, as well as recurrent flooding and ethnic or boundary clashes in parts of the South and West. These issues are often interconnected, transcending regional boundaries and leaving no part of the country untouched. In such contexts, the protection of minority rights becomes increasingly fragile and threatened [1].

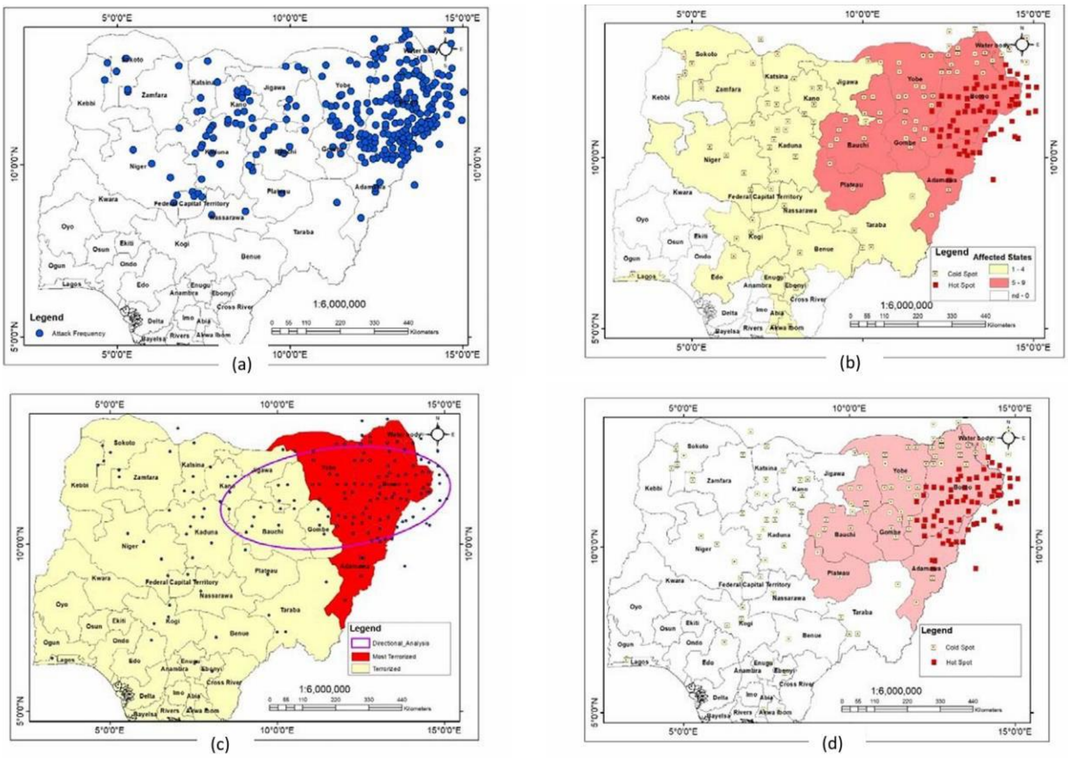


Figure 1. (a) Map of states in Nigeria experiencing frequent Boko Haram attacks, (b) Analysis of hotspots showing Boko Haram attacks in the worst-affected states, (c) Directional analysis of states in Nigeria experiencing frequent insurgency events, (d) Hotspot states with high numbers of internally displaced persons [3]

In West Africa, internally displaced persons (IDPs) face complex challenges largely stemming from armed conflicts, widespread violence, and persistent human rights abuses, with women, children, and the elderly often bearing the heaviest burdens. Among pastoralist groups, forced migration frequently intensifies competition for rangeland, triggering disputes over grazing rights and further straining host communities. Efforts to evaluate the vulnerability of displaced persons are often constrained, as many integrate discreetly into households of relatives and friends, making them less visible to aid agencies and complicating accurate needs assessments. According to the Internal Displacement Monitoring Center, Nigeria alone accounts for nearly 1.5 million displaced individuals, a figure that underscores the scale of the crisis. The dynamics of internal displacement in Nigeria are further aggravated by communal, political, and ethno-religious tensions, whose impacts are both spatial—spanning different regions and issues—and temporal—linked to recurrent events such as elections, resource disputes, and tribal or religious hostilities [4]. Across the African continent, recurring conflicts have continually driven significant population movements, both within and across national borders. According to reports from the African Union, more than 35 million individuals have been forcibly displaced as a result of civil unrest, acts of terrorism, and ethnic violence. Nations such as Sudan, South Sudan, the Central African Republic, and Nigeria remain among those most severely impacted by these crises [5].

Nigeria's experience with internal migration stretches far back into its pre-colonial era, where population movements played a pivotal role in shaping states, sustaining agriculture, advancing trade, and strengthening market networks. During this period, villages, towns, and empires acted as both origins and destinations for migrants, depending on their social or economic pursuits. For example, Hausa traders from the northern region frequently relocated to communities under the Oyo Empire as part of expanding commercial exchanges. Following the empire's collapse in 1837, large-scale population movements and resettlements further reshaped the region. Later, under colonial administration, migration intensified due to urban growth, the establishment of administrative hubs, expansion of cash crop production, and the development of transportation systems such as railways and roads, all of which collectively transformed migration patterns within Nigeria [6]. There remains a pressing need for comprehensive databases within relevant agencies to effectively manage migration-related information. However, the absence of a structured framework for data exchange among the various Ministries, Departments, and Agencies (MDAs) involved continues to pose significant challenges [7]. Nigeria has in recent years grappled with persistent security crises that have deeply influenced patterns of migration across the country. Insurgent activities in the Northeast, particularly by Boko Haram, along with recurrent clashes between farmers and herders in the Middle Belt and other violent conflicts, have led to widespread displacement. Millions have been forced to leave their homes, resulting in increased internal migration toward safer urban areas and a growing trend of emigration as citizens seek security and stability abroad [8].

Nigeria faces significant exposure to disasters arising from both natural and human activities, a situation worsened by insufficient environmental safeguards, limited disaster readiness, and fragile infrastructure. These events—whether occurring abruptly or developing over time—frequently lead to severe outcomes such as loss of life, damage to property, and deterioration of the environment [9]. In the northeastern region of Nigeria, internally displaced persons (IDPs) endure severe forms of mistreatment, deprivation, and exploitation. According to reports from the United Nations World Food Program, prolonged insecurity has left many displaced households facing acute food shortages, forcing large numbers of people to flee their homes. Life within the displacement camps remains extremely unstable, especially for children between the ages of 10 and 15, many of whom display signs of aggression linked to traumatic experiences in their former communities. Women and girls are particularly vulnerable, often exposed to gender-based violence, abduction, and coercive practices such as forced marriages or transactional exchanges. Consequently, many IDPs live in constant fear—hesitant to engage in any sustainable livelihood activities within the camps or to return to their original settlements [10].

Humanitarian personnel serve as indispensable actors in assisting internally displaced persons (IDPs) within regions affected by conflict. They operate in highly volatile and demanding circumstances, striving to deliver essential relief, uphold protection measures, and improve the general welfare of those uprooted from their homes. The significance of their contribution has become even more apparent amid the rising intensity and recurrence of conflicts across the globe, which continue to increase the population of displaced people. Today's global humanitarian emergencies—driven by violent conflicts or natural disasters—remain a major source of human suffering and loss of life. Armed hostilities and organized violence, often instigated by aggressive groups, lie at the core of these crises. In Nigeria, the lived experiences of humanitarian aid workers vividly capture the intricate realities and difficulties associated with supporting internally displaced communities [11]. Although humanitarian interventions play a vital role during disasters, equal attention must be given to minimizing community vulnerability and enhancing preparedness and response capacities. Neglecting these proactive measures can significantly undermine development planning and hinder the pursuit of sustainable growth in Nigeria [12].

Unlike many previous studies that primarily describe displacement trends in Nigeria, this study develops and evaluates a predictive machine learning framework for identifying the likely causes of displacement using humanitarian assessment data. The study contributes in four major ways. First, it integrates two large-scale IOM-DTM datasets into a unified analytical dataset containing over 114,000 observations across 19 variables. Second, it compares multiple machine learning approaches, including Logistic Regression, Random Forest, and XGBoost, under identical preprocessing conditions. Third, it demonstrates that XGBoost can achieve strong predictive performance despite severe class imbalance, obtaining a macro F1-score of 0.96 and balanced accuracy of 0.968. Finally, the study provides a practical foundation for integrating predictive analytics into humanitarian planning, resource allocation, and early-warning systems in Nigeria.

2. COMPREHENSIVE THEORETICAL BASIS

Internally Displaced Persons (IDPs) are generally understood as individuals or groups compelled to relocate within their own country due to circumstances beyond their control. According to the United Nations Guiding Principles, IDPs are those who have been forced to abandon their homes or regular places of residence as a result of armed conflict, widespread violence, human rights violations, or natural and human-induced disasters, but who remain within their nation's borders. The defining characteristic of internal displacement, therefore, lies in the involuntary nature of the movement and the fact that it occurs without crossing into another sovereign state [13]. Internal displacement has become a major global concern, affecting an estimated fifty million people, most of whom reside in Africa and Asia. In many cases, it stems from ongoing armed conflicts and recurring communal violence within nations [14]. Worldwide, an estimated 48 million people are internally displaced as a result of conflict, violence, disasters, and climate-related challenges, often facing serious human rights concerns. Despite the magnitude of this population and their vulnerability, the health conditions of IDPs remain less documented compared to other groups such as refugees and migrants, creating a significant knowledge gap considering their disadvantaged status in affected regions [15]. The Russian invasion of Ukraine, which started in February 2022, has triggered Europe's most significant refugee crisis since World War II and the most severe displacement challenge on the continent since the Yugoslav conflicts of the 1990s [16]. The Nigerian mass media, particularly traditional outlets, have often been faulted for their limited coverage of human rights violations affecting internally displaced persons (IDPs). While such issues frequently surface on social media platforms, mainstream news channels tend to downplay or overlook them, thereby limiting both local and international awareness needed for effective intervention [17]. In 2020, Nigeria experienced about 169,000 new cases of conflict-related displacement, adding to an already large population of roughly two million internally displaced persons. Within this group, older adults remain one of the least recognized and supported categories, as humanitarian interventions and academic studies often prioritize women and children whose vulnerabilities are more visible [18].

Forced migration is broadly understood as the involuntary movement of people caused by factors such as conflict, natural disasters, famine, or large-scale development activities. Unlike voluntary migration, where individuals relocate in search of better opportunities, forced migration reflects situations where people have little or no choice but to leave their homes. Within this context, distinctions exist between categories such as refugees—those who cross international borders to seek protection—and internally displaced persons (IDPs) (see Table 1), who remain within their country but are uprooted due to violence, human rights abuses, or disasters. While various terms like "displaced person" or "environmental refugee" are sometimes used, each carries limitations and does not fully capture the complex realities of forced displacement [19]. In this context, the term forced migrants encompasses refugees, asylum seekers, and internally displaced persons who are compelled to leave their homes against their will. Such displacement is often triggered by armed conflict, political unrest, poverty, environmental challenges, or violations of human rights, leaving individuals to live under conditions of involuntary migration [20]. Migration in the Lake Chad region, particularly northeast Nigeria, has been largely influenced by insecurity arising from the Boko Haram insurgency. The group's violent campaigns, rooted in opposition to Western ideals and a distorted interpretation of Islam, have resulted in widespread displacement and loss of life. According to UNHCR, the combined effects of insurgent attacks and counter-insurgency operations have forced over 1.8 million people in the region to flee their homes in search of safety [20]. Beyond displacement caused by insurgency, migration patterns in Nigeria are also shaped by human trafficking, with Edo State identified as the country's major hub. Over the past two decades, trafficking has expanded internally, moving vulnerable populations from rural areas to urban centers such as Lagos, Ibadan, Kano, Kaduna, Calabar, and Port Harcourt. Women and girls are the most affected, often subjected to domestic servitude and sexual exploitation, while reports suggest complicity of some government and security officials in perpetuating these abuses [21].

Internal migration in Nigeria refers to the movement of people within the country for either temporary or permanent settlement, often shaped by a mix of social, economic, and environmental factors. Since independence, migration patterns have included rural–urban, urban–urban, and urban–rural movements, largely driven by employment opportunities, uneven infrastructure development, and urbanization. Many Nigerians have relocated to cities such as Lagos and Ibadan in search of better livelihoods, while others move to rural areas to engage in agricultural activities. In recent years, insecurity—particularly the Boko Haram insurgency in the northern region—has intensified displacement, contributing to the rising number of internally displaced persons and altering the country's migration dynamics [6]. In line with these trends, the International Displacement Monitoring Center (IDMC) has estimated that over 1.5 million people in Nigeria have been displaced due to recurrent conflicts. These displacements, shown in Figure 2, are often linked to communal, political, and ethno-religious tensions that not only challenge national unity but also fuel the growing population of internally displaced persons. Such conflicts are both spatial—spanning diverse regions and issues—and temporal, arising around sensitive periods or events such as elections, tribal disputes, religious tensions, and competition over resources [4]. Conflicts such as the Boko Haram insurgency in the Northeast and recurring clashes between farmers and Fulani herders in the Middle Belt have been major drivers of displacement in Nigeria. These violent confrontations, along with other security challenges, continue to influence migration flows across regions and beyond national borders [8]. Internal displacement in Nigeria poses critical challenges to sustainable development, as it undermines economic stability, social cohesion, and governance capacity. Displaced persons often lose access to jobs, businesses, and other means of livelihood, resulting in heightened poverty, aid dependency, and vulnerability. Since many IDPs originate from rural farming communities, their displacement disrupts agricultural activities, reduces food production, and contributes to shortages and rising prices. These combined effects intensify food insecurity and place additional pressure on national resources, highlighting the importance of effectively addressing displacement in order to safeguard long-term development [14].

Table 1. United Nations High Commissioner for Refugees (UNHCR) Selected Terms and Definitions [19]

Terms	Definitions
Stateless person	Someone who is not legally recognized as a citizen by any country.
Refugee	A person who, due to a real fear of persecution based on race, religion, nationality, social group, or political opinion, has fled their country and cannot or is unwilling to return for safety reasons.
Internally displaced persons (IDPs)	People forced to leave their homes because of war, violence, human rights abuses, or natural/man-made disasters, but who remain within their country's borders.

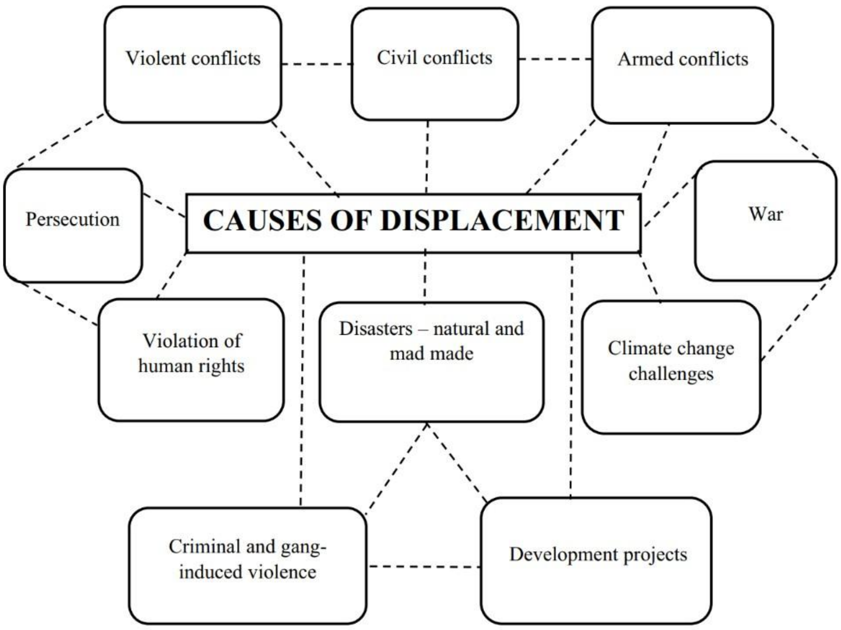


Figure 2. Summary of the Causes of Displacement [22]

Humanitarian interventions trace their roots to the Age of Enlightenment, a period when new ideas about empathy and responsibility began to challenge traditional limits of care. Today, such interventions are understood as organized efforts to provide relief aimed at saving lives, reducing suffering, and upholding human dignity during crises or disasters, with humanitarians serving as frontline caregivers in these situations. Conflicts between Tiv farmers and herders in Benue State (shown in Figure 3) have been marked by violent, coordinated attacks across several local government areas, leading to loss of lives, destruction of property, and mass displacement. These recurring incidents have forced many rural dwellers into Internally Displaced Persons (IDP) camps, where they rely heavily on humanitarian interventions provided by state agencies and international organizations such as National Emergency Management Agency (NEMA), State Emergency Management Agency (SEMA), United Nations High Commission on Refugee (UNHCR), United Nations International Children's Emergency Fund (UNICEF), International Committee of the Red Cross (ICRC), and International Organizations for Migration (IOM) [23]. Humanitarian organizations are institutions dedicated to supporting people during crises by upholding principles such as humanity, neutrality, impartiality, and independence. Their staff, often drawn from NGOs, international agencies, or government bodies, focus on reducing human suffering and meeting urgent as well as long-term needs caused by disasters, conflicts, and health emergencies [11].

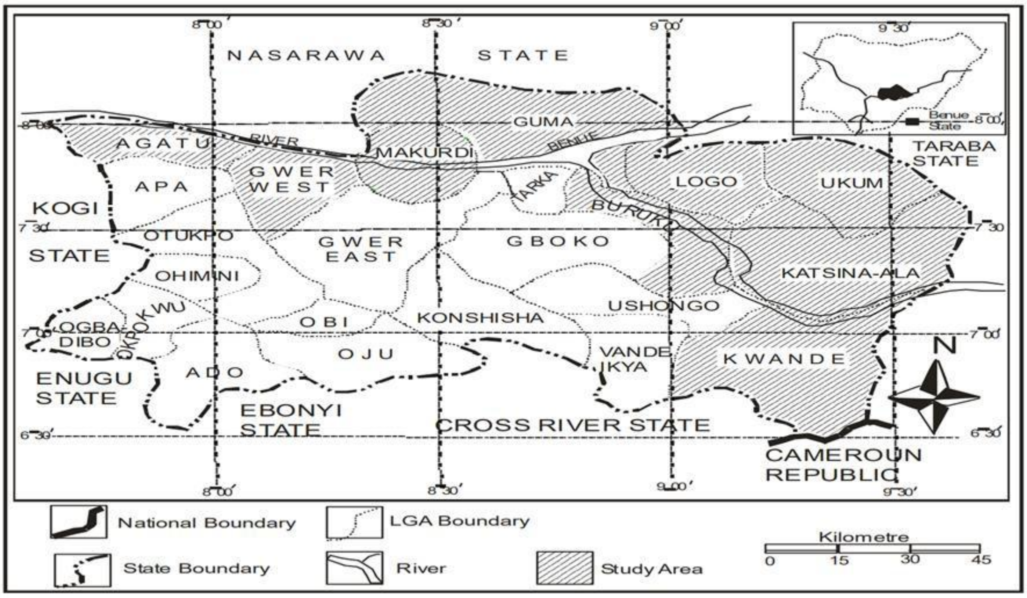


Figure 3. Areas Affected by Farmers and Herdsmen Conflict in Benue State [24]

Sub-Saharan Africa remains a major hotspot for conflict-induced displacement, recording over 16 million displaced persons by 2019, with both the region and the Middle East/North Africa ranking highest for new cases in 2020 [25]. Vulnerability, in the context of displaced populations, is generally understood as a condition where individuals or groups lack adequate power, resources, or protection to safeguard their well- being. It has been described both clinically, as heightened exposure to risks such as illness or adverse events, and ethically, as the inability to defend against exploitation or neglect. Populations often identified as vulnerable include the economically disadvantaged, ethnic minorities, the elderly, the homeless, persons with chronic illnesses, and those facing psychological or social insecurities. These categories highlight that vulnerability is not limited to physical health but also encompasses social and emotional dimensions, making internally displaced persons particularly susceptible to multiple layers of insecurity [26]. Internally displaced persons (IDPs) face pressing needs such as shelter, food, water, security, education, and basic welfare in order to survive and integrate within host communities. However, inadequate protection frameworks, restrictions on movement, threats of forced return, and limited institutional support leave them vulnerable. Beyond humanitarian concerns, displacement also fuels insecurity, disrupts economic productivity, damages infrastructure, and discourages investment, thereby compounding the social and developmental challenges of affected nations [27]. Building on such displacement realities, the UNDP's 1994 report broadened the idea of human security, outlining it across seven areas—economic, food, health, environmental, personal, community, and political security. In Nigeria, these dimensions highlight stark realities: while the South enjoys economic activity, the North remains poor and unstable; food and health insecurity persist as many lack consistent access to nutrition and affordable care; environmental risks like desertification and flooding threaten livelihoods; personal security is undermined by terrorism, banditry, and communal clashes; community security is fragile as religious and ethnic divisions fuel conflict; and weak governance continues to erode political security, leaving citizens vulnerable and grievances unaddressed [28].

Internally displaced persons (IDPs) often face greater vulnerability than refugees, particularly regarding access to basic needs, healthcare, and protection. Despite having formal rights to social support, many encounter barriers such as weak institutions, limited medical facilities, and discrimination in host communities. These challenges heighten their exposure to trauma, psychosocial distress, and preventable health risks, with children being especially affected by malnutrition and infectious diseases [29]. Women and children, who make up the majority of displaced populations, often face severe health challenges. Studies reveal that women in IDP camps are at heightened risk of sexual violence, unsafe pregnancies, maternal complications, and long-term physical and psychological harm. Displaced children equally struggle with poor nutrition and growth problems, while both groups frequently experience trauma, depression, and anxiety. Such vulnerabilities not only worsen immediate health outcomes but also contribute to a growing burden of chronic diseases among displaced communities [26].

Most cases of internal displacement in Nigeria stem from violent conflicts driven by ethnic, religious, or political tensions, though both human-induced and natural causes also contribute [30][1]. Nigeria's national security remains a major concern due to recurring threats from insurgency, communal clashes, and criminal activities. Boko Haram, in particular, has destabilized the northern region with its agenda of enforcing Sharia law, while issues such as herder–farmer conflicts, Niger-Delta militancy, and widespread kidnappings further compound insecurity [27]. In Nigeria, violent conflicts such as the Boko Haram insurgency, communal clashes, and herder–farmer disputes have been major drivers of displacement, particularly in the northeast. Borno State alone accounted for over half of the country's 2.6 million internally displaced persons in 2019, while religious and ethnic tensions, as well as the introduction of Sharia law in 2000, further fueled unrest. These crises, alongside recurrent farmer–herder violence, have continued to generate tens of thousands of new displacements each year [1]. Nigeria's secular structure has encouraged religious plurality, yet the fusion of religion and politics has fueled recurring ethno-religious conflicts. These tensions, especially in the North, have led to violence, insurgencies, and mass displacement, with millions uprooted from their homes. The North-East remains the most affected, largely due to the Boko Haram insurgency, which accounts for the highest concentration of internally displaced persons [31]. Banditry in Nigeria, which initially stemmed from farmer–herder clashes around 2011, has since escalated into widespread criminal activities such as cattle theft, abductions for ransom, sexual violence, and frequent killings [32]. Although government forces have weakened Boko Haram's operations by eliminating key leaders and cells, the group continues to exploit porous borders, and peace efforts through dialogue have yielded little progress [27].

Nigeria faces significant environmental risks, with desertification and recurrent floods driving internal displacement. For instance, in 2019, heavy rains and dam releases in the Niger River basin led to the destruction of thousands of homes in Niger State, displacing over 150,000 people [1]. Ecological shifts, environmental stress, and dwindling natural resources have intensified competition among communities, while poverty, unemployment, and weak governance have pushed many into crimes for survival. The presence of vast forests has further enabled armed groups to establish hideouts beyond the reach of under-resourced security forces. These conditions have disrupted socio-economic life, created widespread insecurity, and fueled national debates on displacement in the country [32]. Communal conflicts remain a persistent challenge within Nigeria and across Africa, undermining security and obstructing development efforts. Such clashes often result in loss of life, destruction of property, and lingering social scars, while placing heavy demands on government resources for reconciliation and reconstruction [33]. Although violence is a widely discussed phenomenon, arriving at a universally accepted definition remains difficult, as its meaning often shifts with social and cultural contexts. The term generally conveys the use of force or coercion to dominate or harm others, usually in unlawful or oppressive ways, but its interpretation varies depending on disciplinary and societal perspectives [34].

Researchers emphasize that the humanitarian field needs to embrace data-driven strategies, shifting toward a coordinated and innovative use of information to improve efficiency, effectiveness, and overall impact [35]. Since the rise of Industry 4.0, humanitarian agencies have increasingly adopted digital technologies to improve crisis response. Tools such as mobile data applications, GIS mapping, and social media have enabled faster assessments, efficient distribution tracking, and real-time communication during emergencies. Organizations like UNHCR, WFP, and the Red Cross have leveraged these tools to support refugees, monitor food supplies, and coordinate disaster relief, while institutions such as WHO employ big data and predictive analytics to detect and forecast disease outbreaks. More recently, innovations like blockchain have been piloted to strengthen aid delivery and enhance supply chain resilience [36]. Humanitarian mapping guided by data plays a crucial role in shaping technologies and policies, integrating computational methods with legal, ethical, and governance perspectives [37].

Recent progress in artificial intelligence has shifted from basic data handling to building advanced models that can uncover migration trends hidden within large and complex datasets. By combining diverse sources such as economic indicators, satellite data, social media, and demographic records, AI enables more holistic migration flow analysis. Beyond forecasting, these tools are increasingly applied to areas like resource planning, skills matching, and policy evaluation, offering governments and organizations a proactive, data- driven way to address migration challenges [38]. Humanitarian agencies responsible for refugees and internally displaced persons often rely on forecasting to allocate shelter, food, and healthcare resources effectively. Traditionally, such projections guide field offices in planning budgets and supplies, but recent attention has shifted toward the use of computational approaches—including machine learning, artificial intelligence, and statistical modeling—to enhance the accuracy of predicting displacement trends [39]. Machine learning involves creating algorithms (illustrated in Figure 4) that enable computers to learn from data and make predictions. Its applications span diverse fields such as business, engineering, medicine, and statistics, offering insights that support decision-making. In migration studies, these techniques are particularly valuable for governments, as they help anticipate demographic shifts, labor market dynamics, disease spread, and broader economic impacts [40]. Recent studies increasingly explore how computational methods such as machine learning, artificial intelligence, and simulations can assist field operations by forecasting migration flows. These efforts apply diverse modeling approaches across different displacement contexts, aiming to predict arrivals at varying scales of time and location [39].

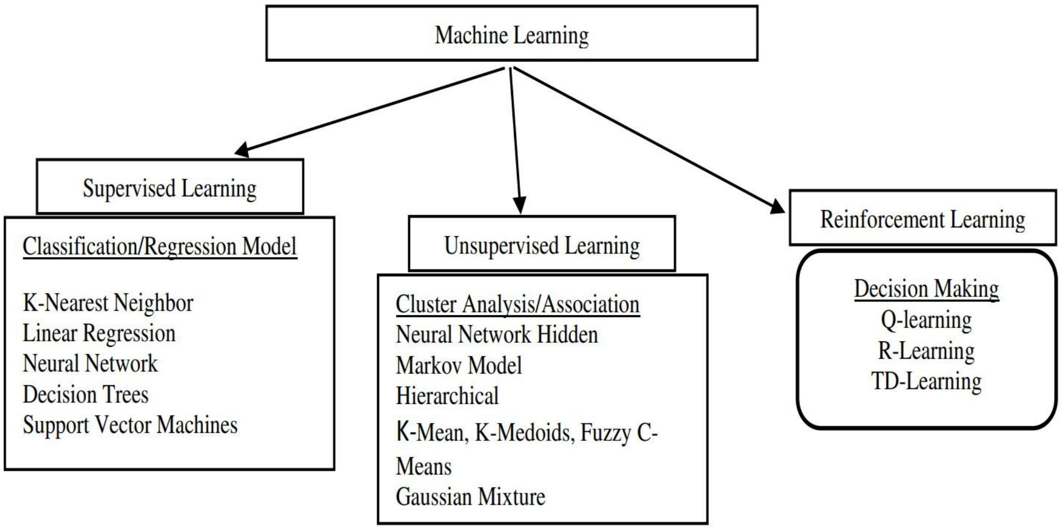


Figure 4. Machine Learning Algorithms [40]

Although several studies have examined the causes and consequences of displacement in Nigeria, relatively few have applied predictive analytics to humanitarian data. Existing studies in countries such as Sudan, Somalia, and the Democratic Republic of Congo have increasingly explored machine learning methods for forecasting conflict-induced migration and refugee movements. However, many of these studies rely on small datasets, focus on a single conflict driver, or use only traditional statistical methods. In contrast, the present study combines large-scale humanitarian datasets with machine learning algorithms to model multiple displacement drivers simultaneously. This approach provides stronger predictive capability, greater interpretability, and improved relevance for operational humanitarian planning.

3. METHOD

The datasets used in this research were obtained from the International Organization for Migration's (IOM) Displacement Tracking Matrix (DTM), accessed via the Humanitarian Data Exchange (HDX) platform. The HDX platform is a global, open-access portal where humanitarian organizations publish verified datasets for use by researchers, policymakers, and practitioners. This ensures that the data used is both credible and reliable.

The DTM is a specialized system developed by IOM to monitor population movements, displacement trends, and the evolving needs of crisis-affected communities. In Nigeria, the DTM is particularly important as it captures large-scale data on internally displaced persons (IDPs), returnees, and host communities affected by conflict, disasters, and other forms of crisis. It provides both baseline demographic data and in-depth assessments of humanitarian needs, making it a rich source for both descriptive and predictive analysis. For this research, two main datasets were selected: Baseline Assessment and Needs Monitoring. The combination of these two datasets provided a comprehensive foundation, with the Baseline Assessment offering demographic and displacement patterns, while the Needs Monitoring dataset supplied deeper insights into the challenges and vulnerabilities faced by displaced populations. The summary of both is given below:

a. Baseline Assessment Dataset: A relatively concise dataset with 38 variables, focusing on household demographics, reasons for displacement, and settlement characteristics.
b. Needs Monitoring Dataset: A larger and more complex dataset containing 544 variables, covering a wide range of humanitarian indicators including access to food, water, healthcare, shelter, security, and barriers to receiving assistance.

The process of data acquisition was designed to ensure that the datasets were downloaded, stored, and prepared in a way that preserved their accuracy and authenticity. The first step involved navigating to the HDX platform and searching specifically for Nigeria's IOM–DTM datasets. From the search results, the most recent and relevant versions of the Baseline Assessment and Needs Monitoring datasets were selected. Both datasets were provided in compressed CSV (Comma-Separated Values) formats, which are widely compatible with Python and other data analysis tools. Once downloaded, the datasets were uploaded into the analysis environment, which in this case was Google Colab. Google Colab was chosen because it offers cloud-based storage, preinstalled machine learning libraries, and flexibility for running Python code without requiring complex local configurations. The CSV files were uploaded into Colab's session storage, after which the

Python pandas library was used to read them into dataframes (df for the Baseline dataset and df2 for the Needs Monitoring dataset). To maintain data integrity, the datasets were obtained directly from the HDX platform without modification from third-party sources. This avoided the risk of working with outdated or altered versions. Table 2 provides a summary of the raw structure of the datasets before further cleaning and processing.

Table 2. Overview of Raw Dataset Structure

Dataset Name	No. of Columns	Main Categories of Information
Baseline Assessment (df)	38	Demographics, displacement reasons, household details
Needs Monitoring (df2)	544	Living conditions, needs, priorities, barriers, accessibility

Raw datasets downloaded from humanitarian repositories often include metadata rows or additional headers as shown in Figure 3.2. These rows typically contain survey notes, administrative details, or repeated headers, which cannot be used in analysis. The first cleaning step therefore involved scanning and removing any non-analytical rows and redundant headers from both datasets, shown in Figure 3.3. This ensured that only valid observations remained in the working dataframes. In addition, initial filtering was conducted to remove irrelevant or placeholder entries. Some rows in large humanitarian datasets may represent test entries, incomplete surveys, or records without valid location identifiers. Since such entries provide no analytical value and can distort statistical outcomes, they were carefully removed. By conducting this filtering early, the dataset was reduced to a clean and relevant form, which is essential for ensuring the reliability of subsequent steps such as exploratory analysis and model training. The Baseline dataset, with its 38 columns, was relatively concise, and most variables were retained for analysis as they aligned with the research objectives. The Needs Monitoring dataset, however, contained 544 variables, many of which were highly detailed or administrative in nature. An important step was therefore to carefully identify and retain only the most relevant columns. The selection process was guided by the study's objectives, which focused on demographics, displacement drivers, household needs, and barriers to accessing assistance. For example, variables describing location identifiers, household size, gender composition, and reported top needs were kept, while administrative codes, survey logistics fields, and redundant metadata were excluded. This reduction was necessary not only to reduce computational complexity but also to ensure that the final dataset focused squarely on variables that could meaningfully contribute to the predictive modeling process. After cleaning the two source files and standardizing text fields, the Baseline Assessment and Needs Monitoring subsets were merged on administrative geography only. Specifically, the join keys were State and LGA, which both datasets shared after text normalization (.str.strip().str.title()). The merge used an inner join so that only locations present in both datasets were retained. This choice ensured that each record in the final dataset combined displacement counts and demographics from the Baseline file with site accessibility status and self-reported priorities/barriers from the Needs Monitoring file, but only for LGAs with information in both sources.

No household-level identifiers were used, and no additional deduplication rules were applied beyond the inner join. Because multiple entries can exist per LGA in either source, the join intentionally preserved all valid many-to-many matches. This reflects the reality that several sites/assessments may exist within the same LGA and period. Post-merge checks confirmed that the resulting frame contained 19 columns and 114,563 rows, with no missing values in the joined columns used for analysis. This unified dataset served as the single source for all subsequent EDA and modeling steps. Table 3 summarizes the merge specifications and results.

The Exploratory Data Analysis (EDA) phase was carried out to provide a deeper understanding of the datasets before developing predictive models. The goal of EDA was to identify patterns, detect anomalies, summarize key characteristics, and visualize relationships within the data. By examining the data systematically, EDA helped in formulating hypotheses, validating assumptions, and ensuring that the dataset was suitable for predictive modeling. In this research, EDA was conducted using both descriptive statistics and visualization techniques. Numerical summaries such as mean, median, minimum, and maximum values were computed for quantitative fields, while frequency counts and percentages were calculated for categorical variables. Visualization tools such as bar charts, histograms, line plots, and heatmaps were employed to make patterns and trends more interpretable. The EDA process was structured into multiple stages: univariate analysis (to examine each variable in isolation), bivariate analysis (to explore the relationship between two variables), time series analysis (to examine trends across time), and thematic analysis of priority needs and barriers (to highlight pressing humanitarian challenges). By following this structured approach, a comprehensive understanding of the data was achieved before moving into modeling. The univariate analysis focused on examining each variable individually to understand its distribution and main characteristics. For categorical variables such as gender, displacement reason, or type of settlement, frequency counts were computed and visualized using bar charts. This made it possible to identify dominant categories—for example, conflict and insurgency emerged as the leading causes of displacement in the dataset, as shown in Figure 5.

Table 3. Merge Specification and Result

Item	Details
Join keys	State, LGA
Join type	Inner join (how='inner')
Text standardization	.str.strip().str.title() on State, LGA in both dataframes
Date handling before merge	Date of Survey (baseline) and survey_date (needs) parsed with pd.to_datetime(..., errors='coerce')
Rows after merge	114,563
Columns after merge	19

```
[ ] df_merged['Reason of displacement of majority of IDPs'].value_counts()
```

Reason of displacement of majority of IDPs	count
Insurgency	108835
Communal Clash	5040
Natural Disaster	576
Farmers /Herders	64
Flood	36
Banditary	12

dtype: int64

Figure 5. Code Snippet for the Plot for Most Common Displacement Reasons

Bivariate analysis was conducted to investigate relationships between two variables. This step was essential for uncovering associations that could later serve as predictive features in the modeling stage. Bivariate analysis examined how the reason for displacement relates to key categorical covariates that are available in the merged dataset. Three relationships were explored with count plots: population type × reason for displacement; site accessibility × reason for displacement; and state × reason for displacement (ordered by state frequency). These plots reveal where specific displacement drivers are more prevalent and how site conditions (e.g., whether a site is accessible) vary across displacement reasons. In addition, a correlation heatmap was produced for numeric variables (Households, Individuals, Total male, Total female). This numeric view helped confirm expected internal consistency (e.g., strong positive relationships between Individuals and sex-specific totals). A summary of the bivariate analysis is given in Table 4.

Before building the models, the dataset had to be prepared so that all features could be properly understood by machine learning algorithms. The raw data came from the International Organization for Migration (IOM) Displacement Tracking Matrix assessments, and it included a wide range of variables — state names, reasons for displacement, site accessibility, household counts, gender distribution, and more. Since these variables were of different types (categorical, numerical, and textual), leaving them in their raw form would have made them difficult for the models to interpret. Some features had large differences in scale, others contained missing values or inconsistencies, and the target variable itself was imbalanced. To solve these issues, a series of transformations were applied to clean, standardize, and properly represent the data. These steps made the dataset consistent, comparable, and suitable for predictive modeling. The transformations summary is presented in Table 5.

Table 4. Bivariate Analyses Produced

Relationship	Method	Rationale
Population type × Reason for displacement	Count plot (sns.countplot)	Shows how displacement drivers differ for IDPs in camps vs host communities vs returnees
Site accessibility × Reason for displacement	Count plot	Indicates whether accessibility patterns vary by displacement driver
State × Reason for displacement	Count plot (states ordered by count)	Highlights geographic concentration of specific drivers
Numeric feature inter- relationships	Correlation heatmap for households, individuals, total_male, total_female	Validates internal consistency in population counts

Table 5. Summary of Feature Transformation

Feature Type	Example Variables	Transformation Applied
Categorical Variables	state, lga, population_type, population_types, site_accessibility, site_safe_area, main_priorities_first/second/third, main_barrier_first/second/third	One-Hot Encoding (OHE) → Converted categories into binary indicator variables without imposing order.
Numerical Variables	households, individuals, total_male, total_female, avg_hh_size, male_share	Standardization using StandardScaler → Rescaled to zero mean and unit variance.
Target Variable	reason_for_displacement (mapped to Insurgency, Communal Clash, Natural Disaster)	Label Encoding → Converted textual labels into numeric class indices.
Class Weights	Distribution of reason_for_displacement	Balanced using computed class weights to address class imbalance during training.
Missing/Inconsistent Data	Categorical: e.g., "Do not know/nanswer", "Access to cooking fuel" Numerical: missing household/individual counts	Categorical → Standardized values, replaced missing with "missing"; Numerical → Filled missing with 0.0.
Pipeline Integration	All categorical, numerical, and target fields	Combined within ColumnTransformer and Pipeline to ensure consistent preprocessing during training/testing.

The classification problem was defined as: predicting the reason for displacement from available features. The raw target column was Reason of displacement of majority of IDPs, standardized to lowercase/underscore format in code. To reduce sparsity and align with humanitarian interpretation, the target was mapped to three classes using keyword rules:

a. Insurgency (e.g., “insurgency”, “boko”)
b. Communal Clash (e.g., “communal”, “clash/es”, “bandit, “farmer/herder”, general “conflict/attack”)
c. Natural Disaster (e.g., “flood”, “natural disaster”, “flooding”)

Only rows mapping to these three classes were kept for modeling. The resulting distribution is highly imbalanced, with Insurgency dominating, as illustrated in Table 6.

Table 6. Target Mapping and Class Distribution

Raw label examples → Mapped class	Examples of matching keywords	Count (after mapping)	Share
Insurgency	“insurg…”, “boko…”	108,835	~95%
Communal Clash	“communal”, “clash”, “bandit…”, “farmer/herder”, “conflict/attack”	5,116	~4%
Natural Disaster	“flood”, “natural disaster”, “flooding”	612	~1%

Given the extreme skew toward Insurgency, the models were trained with class weights only. Class weights were computed from the training split via compute_class_weight('balanced', ...), then passed to the estimators. For tree-based models that accept per-class weights (Random Forest), the computed weights were provided directly. For XGBoost, the per-row sample weights derived from the class weights were supplied during fit. This approach penalizes mistakes on minority classes more heavily and helps the classifier avoid collapsing to the majority class. All models (configuration summarized in Table 7) were embedded in the same preprocessing pipeline so that comparisons reflect model capability rather than preprocessing differences. Three algorithms were trained and compared:

a. Logistic Regression (multinomial, lbfgs): a transparent baseline; benefits from standardized numeric variables and one-hot encoded categoricals.
b. Random Forest (200 trees, class_weight applied): robust to non-linearities and mixed feature types; provides feature importance for interpretation.
c. XGBoost (if available): gradient-boosted trees optimized for tabular data; trained with per-row sample weights reflecting class weights.

Table 7. Model Configurations

Model	Key settings	Notes
Logistic Regression	multi_class='multinomial', solver='lbfgs', max_iter=1000, class weights	Baseline linear classifier
Random Forest	n_estimators=200, class_weight applied, random_state=42	Non-linear, importance available
XGBoost (optional)	objective='multi:softprob', eval_metric='mlogloss', n_estimators=300	Sample weights used for imbalance

Table 8. Definition of Variables Used in the Final Model

Variables	Definitions
state	Nigerian state where displacement occurred
lga	Local Government Area of displacement
households	Number of affected households
individuals	Number of affected individuals
total_male	Total number of male individuals
total_female	Total number of female individuals
avg_hh_size	Average household size
male_share	Ratio of males to total population
population_type	Category of displaced group
population_types	Camp, host community, or returnee grouping
site_accessibility	Whether the site is accessible
site_safe_area	Whether the location is considered safe
main_priorities_first/second/third	Top reported humanitarian needs
main_barrier_first/second/third	Main barriers to assistance
reason_for_displacement	Target variable representing insurgency, communal clash, or natural disaster

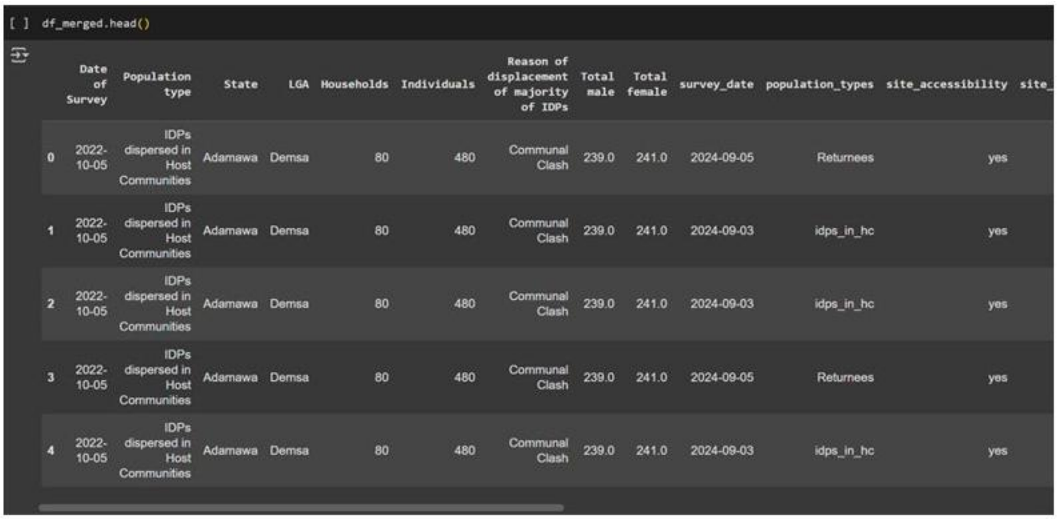
.

A unified Pipeline was created with a ColumnTransformer to apply One-Hot Encoding to categorical variables and scaling to numeric features. Categorical features included: state, lga, population_type, population_types, site_accessibility, site_safe_area, and the three priority and three barrier fields (main_priorities_*, main_barrier_*). Numeric features included: households, individuals, total_male, total_female, and two engineered variables: avg_hh_size = individuals/households (safe-guarded for zero denominators) and male_share = total_male / (total_male + total_female). Missing categorical values were filled with 'missing', and numeric NaNs were coerced to 0.0. Data was split 80/20 with stratification. Models were trained within the pipeline to prevent leakage and ensure identical preprocessing at train and test time. Performance was assessed using F1-macro (primary), balanced accuracy, and standard diagnostics (classification report, confusion matrix).

Alternative imbalance mitigation techniques such as SMOTE, ADASYN, and random oversampling were considered. However, these approaches may generate synthetic minority samples that do not accurately reflect the complex humanitarian realities of displacement data. Because the minority classes represented only a small proportion of observations, the study instead used class weights, which preserve the original data distribution while penalizing misclassification of minority classes more heavily during training. Table 8 shows the variables used in the final model.

4. RESULT AND DISCUSSION

The data collected from the International Organization for Migration (IOM) Displacement Tracking Matrix (DTM) provided the foundation of this study. After integrating the raw files, the dataset was found to consist of diverse attributes including displacement locations (state and local government areas), demographic details (households, individuals, males, females), and contextual indicators such as site accessibility, safety, and priority needs. A unified dataset was successfully created by merging multiple assessment rounds. This integrated dataset provided a comprehensive representation of displacement patterns across different Nigerian states. The final dataset shown in Figure 6 after integration contained 114,563 rows and 19 columns. This ensured that sufficient data points were available for both training and testing machine learning models.



```
[ ] df_merged.head()
```

	Date of Survey	Population type	State	LGA	Households	Individuals	Reason of displacement of majority of IDPs	Total male	Total female	survey_date	population_types	site_accessibility	site_
0	2022-10-05	IDPs dispersed in Host Communities	Adamawa	Demsa	80	480	Communal Clash	239.0	241.0	2024-09-05	Returnees	yes	
1	2022-10-05	IDPs dispersed in Host Communities	Adamawa	Demsa	80	480	Communal Clash	239.0	241.0	2024-09-03	idps_in_hc	yes	
2	2022-10-05	IDPs dispersed in Host Communities	Adamawa	Demsa	80	480	Communal Clash	239.0	241.0	2024-09-03	idps_in_hc	yes	
3	2022-10-05	IDPs dispersed in Host Communities	Adamawa	Demsa	80	480	Communal Clash	239.0	241.0	2024-09-05	Returnees	yes	
4	2022-10-05	IDPs dispersed in Host Communities	Adamawa	Demsa	80	480	Communal Clash	239.0	241.0	2024-09-03	idps_in_hc	yes	

Figure 6. Final Dataset After Integration

Exploratory Data Analysis was carried out to understand the overall patterns, distributions, and relationships in the dataset before model development. The analysis highlighted the dominance of insurgency as the major driver of displacement in Nigeria, while also uncovering demographic patterns such as variations in household size, gender composition, and the geographical spread of internally displaced persons (IDPs). Furthermore, the analysis shed light on the key priorities and barriers faced by displaced populations, as reported through field assessments. The insights gained from EDA not only informed the preprocessing steps but also guided the interpretation of the model outcomes in later sections. The analysis revealed that insurgency accounted for the overwhelming majority of recorded displacements, representing nearly 95% of all cases, as illustrated in Figure 7. Communal clashes and natural disasters, although present, formed very small portions of the dataset. This heavy imbalance emphasized the need to address skewed class distributions during the modeling process. When examining the geographical distribution of displacement, it was observed that states such as Borno, Adamawa, and Yobe ranked among the top ten with the highest number of individuals affected (see Figure 8), with Borno being the highest. These states are located within the North-East region of Nigeria, which has been severely impacted by insurgency-related violence.

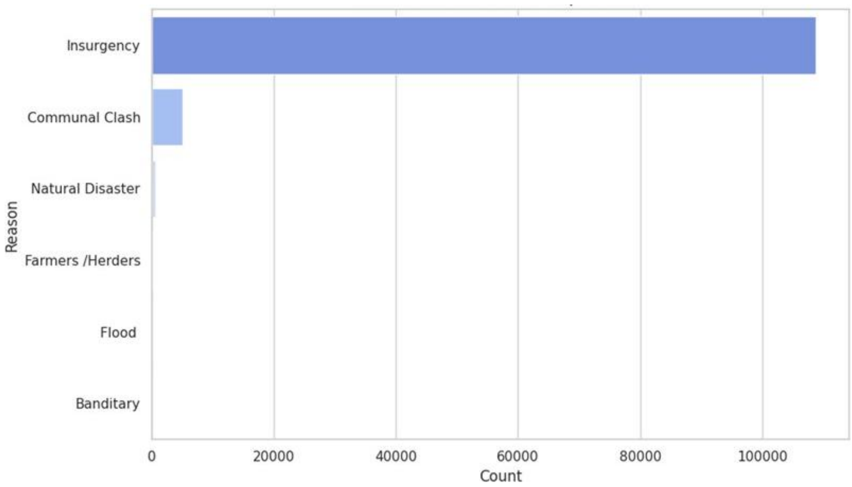


Figure 7. Distribution of Displacement for Reasons

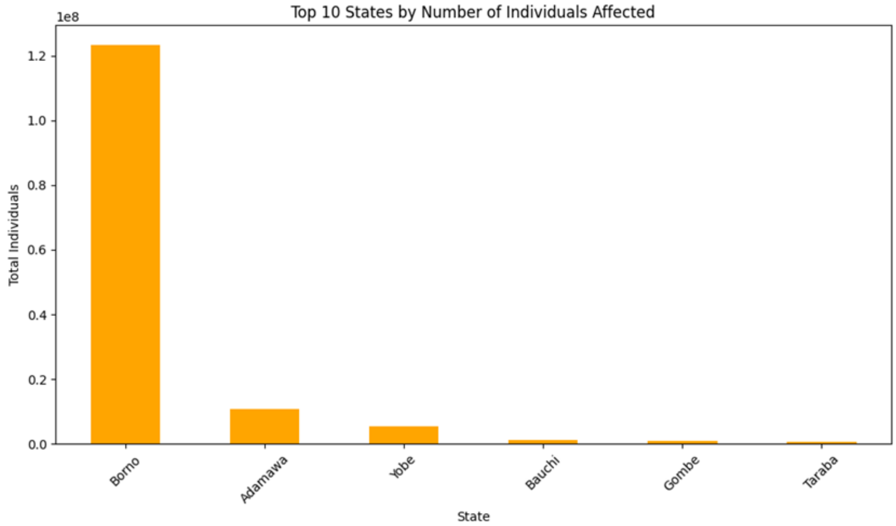


Figure 8. Top 10 States by Number of Individuals Affected

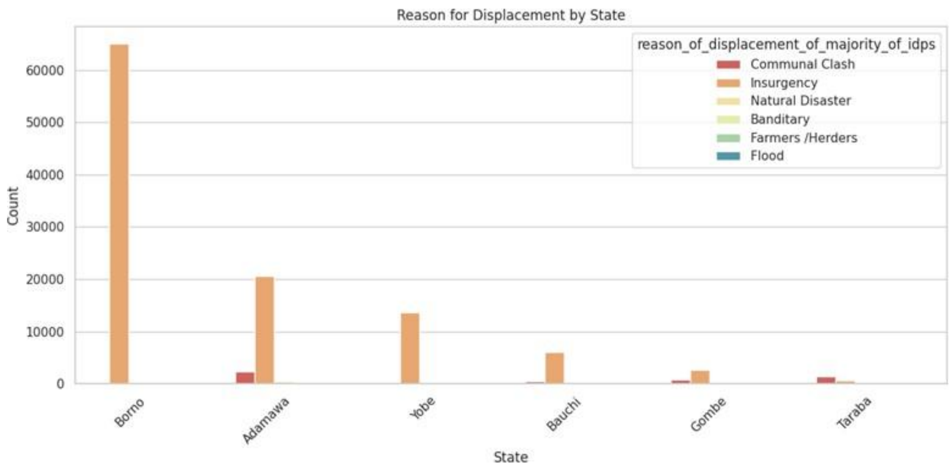


Figure 9. Reason for Displacement by State

To gain deeper insights, relationships between selected variables were explored. The cross-tabulation of displacement reasons across states revealed strong geographical clustering. For example, insurgency-related displacements were concentrated in the North-East, illustrated in Figure 9, while communal clashes were more frequent in parts of the Middle Belt. Natural disasters were reported in scattered areas but occurred at much lower frequencies.

The transformations applied during preprocessing were successfully verified. Categorical features were converted into binary indicators through One-Hot Encoding, numerical features such as household counts and gender ratios were standardized, and the target variable (reasons for displacement) was reduced to three main categories (Insurgency, Communal Clash, Natural Disaster) before label encoding, an example of which is summarized in Table 8. The effectiveness of these transformations was confirmed during model fitting, as all variables were now in a machine-readable form. Additionally, the application of class weights helped reduce the imbalance effect, ensuring that minority classes were not completely ignored by the classifiers.

Table 9. Example of Feature Transformation Results

Raw Variable	Raw Value	Transformed Representation
Reason	"Insurgency"	0
Reason	"Communal Clash"	1
Reason	"Natural Disaster"	2
Population Type	"Household"	Household=1, Individual=0
Household Size	15	Standardized = 0.84

Three models were trained and compared: Logistic Regression, Random Forest, and XGBoost. Using the same training/validation setup, the same preprocessing (one-hot encoding for categorical variables and scaling for numeric ones), and class weighting to address the strong class imbalance (Insurgency ≈ 95%, others ≈ 5%). The models were then judged by F1-macro (treats classes equally), Balanced Accuracy, F1-weighted, and a practical score that combines accuracy and speed. As shown in Figure 10, XGBoost achieved the best overall performance and the highest practical score. The confusion matrix depicted in Figure 11 shows how well the best model, XGBoost, separates the three classes after all transformations and class-weighting. It correctly classifies almost all Insurgency cases and performs strongly on the minority classes Communal Clash and Natural Disaster.

```
[34] results_df
```

	model	f1_macro	f1_weighted	balanced_acc	train_time_s	pred_time_total_s	pred_time_per_sample_s	inv_time_score
0	XGBoost	0.962717	0.998918	0.968012	11.826446	0.997081	0.000044	0.000000
1	RandomForest	0.937518	0.995875	0.931835	57.389520	0.519814	0.000023	0.551363
2	LogisticRegression	0.640928	0.954985	0.923373	25.300291	0.131469	0.000006	1.000000

Figure 10. Model Comparison

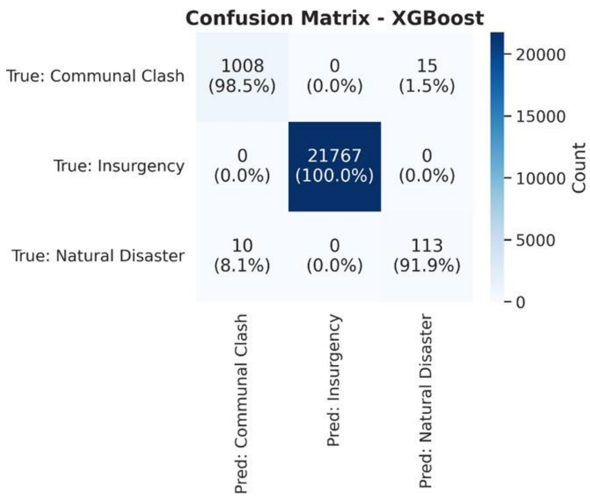


Figure 11. Confusion Matrix for XGBoost

To complement the confusion matrix, the per-class report summarized in Table 10 confirms the same pattern: very high scores for Insurgency and strong, balanced performance for the two minority classes given the severe skew. Tree-based models like XGBoost provide feature importance that highlights which inputs most influenced predictions. In this case, the most informative signals comeX from both demographics and location context summarized in Table 11. The feature-importance table for the consistent top 15 signals for the XGBoost model is summarized in Table 12.

Table 10. Classification Report for XGBoost

Class	Precision	Recall	F1-Score	Support
Communal Clash	0.99	0.99	0.99	1,023
Insurgency	1.00	1.00	1.00	21,767
Natural Disaster	0.88	0.92	0.90	123
Accuracy	-	-	1.00	22,913
Macro Avg	0.96	0.97	0.96	22,913
Weighted Avg	1.00	1.00	1.00	22,913

Table 11. Summary of Key Features Highlighted by XGBoost

Feature / Group	Why it helps the model
male_share, avg_hh_size	Capture household composition differences across displacement types
households, individuals, total_male/female	Reflect scale and structure of affected groups
State indicators (e.g., Borno, Taraba, Adamawa, Yobe)	Encode regional exposure to insurgency vs. other drivers
LGA indicators (e.g., Guyuk, Numan, Maiduguri, Lamurde, Demsa)	Capture finer local patterns and hotspots
population_type / population_types	Distinguish IDPs in camps vs. host communities vs. returnees
site_accessibility, site_safe_area	Proxy for security/operational conditions tied to certain displacement causes
Priority/Barrier fields	Capture immediate needs and constraints that correlate with displacement drivers

Table 12. Summary of Top 15 Features Importance Scores by XGBoost

Rank	Feature	Importance (Gain)
1	male_share	0.276
2	avg_hh_size	0.193
3	households	0.142
4	individuals	0.112
5	total_male	0.089
6	total_female	0.074
7	population_type	0.052
8	site_accessibility	0.031
9	state_Borno	0.023
10	lga_Maiduguri	0.018
11	site_safe_area	0.015
12	state_Taraba	0.012
13	lga_Numán	0.010
14	priority_food	0.009
15	barrier_shelter	0.006

Overall, the results show that after careful preprocessing and class-weighting, XGBoost delivers the strongest, most balanced performance on this highly imbalanced task, with the best F1-macro and balanced accuracy while keeping prediction time low; it cleanly identifies the dominant Insurgency class and still performs well on the minority classes Communal Clash and Natural Disaster, as reflected by the near-diagonal confusion matrix and the per-class scores; the most influential signals combine demographic composition (e.g., male share, household size, population counts) and geography (state/LGA), with contextual fields such as population type, accessibility, and stated priorities/barriers adding helpful separation, all of which is consistent with known spatial and social patterns in displacement data and confirms that the feature transformations and modeling choices used here are appropriate for the problem.

The predictive framework developed in this study can be integrated into humanitarian early-warning systems used by organizations such as IOM, NEMA, SEMA, UNHCR, and NGOs. For example, if the model predicts a high likelihood of insurgency-driven displacement in selected LGAs in Borno or Adamawa State, agencies could pre-position food, shelter materials, healthcare personnel, and security support in nearby camps before large-scale arrivals occur. Similarly, if natural-disaster-related displacement is predicted during flood seasons in riverine communities, emergency agencies could allocate temporary shelters, water supplies, and evacuation transport in advance. Such predictive outputs would enable humanitarian organizations to move from reactive response toward anticipatory planning.

5. CONCLUSION AND LIMITATION

The findings of this study demonstrate the practical potential of machine learning in addressing one of Nigeria's most persistent humanitarian challenges—internal displacement. By employing a rigorous methodology that included comprehensive data preprocessing, exploratory analysis, and meticulous feature transformation, the study ensured that the models were trained on consistent and representative datasets. The comparative evaluation of algorithms revealed that while Random Forest performed strongly, the XGBoost classifier achieved the highest accuracy, making it a strong candidate as a scalable and operationally useful model for forecasting displacement trends. However, successful real-world deployment will depend on access to timely humanitarian data, technical capacity within response agencies, and careful consideration of ethical issues related to predictive analytics in vulnerable populations. This outcome highlights several important insights: predictive analytics can serve as a powerful early-warning mechanism, enabling humanitarian agencies and policymakers to forecast areas of heightened displacement risk and respond proactively; internal displacement follows discernible socio-economic, demographic, and environmental patterns rather than random occurrence; and the use of XGBoost effectively captures non-linear relationships in data, enhancing robustness and producing actionable outputs. Beyond technical performance, the study underscores the social realities of displacement in Nigeria, where women and children remain disproportionately affected, conflict continues to be the primary driver of forced migration, and regions such as the North-East bear a disproportionate burden. These findings emphasize the urgency of integrating predictive tools into humanitarian response systems, transforming aid delivery from reactive to anticipatory. Overall, this study successfully designed and evaluated a machine learning-based displacement prediction framework that is practical, scalable, and capable of integration into humanitarian early-warning systems. Although no model can entirely eliminate displacement risks, predictive analytics provides a strong foundation for enhancing preparedness, optimizing resource allocation, and strengthening resilience strategies for vulnerable populations in Nigeria. The study's major contributions include the development of a robust data- driven framework for predicting internal displacement trends in Nigeria, demonstration of the superior performance of the XGBoost model, application of advanced feature transformation techniques to improve accuracy and interpretability, empirical validation of machine learning's ability to anticipate displacement patterns, and provision of a technical foundation for future humanitarian decision-support systems. To build on these outcomes, future research should focus on integrating real-time data sources such as conflict incident reports, satellite imagery, and climate data to improve prediction accuracy; expanding the dataset to include recent years and emerging hotspots; developing user-friendly dashboards or mobile applications for policymakers and field teams; exploring deep learning models for temporal and spatial forecasting; promoting collaboration between government agencies, NGOs, and research institutions to operationalize predictive tools in humanitarian planning; and extending the framework to encompass displacement triggers beyond conflict, including pandemics and long-term climate change impacts. One limitation of this study is that the dataset was randomly split into training and testing subsets rather than divided chronologically. As a result, the model's ability to forecast future displacement trends across entirely unseen time periods was not directly evaluated. Future research should adopt temporal validation strategies, such as training on earlier years and testing on more recent years, to better assess the model's real-world forecasting capability.

REFERENCES

[1] A. T. Akujobi and U. V. Awhefeada, "Migration and displacement: Legal constraints of internally displaced persons in Nigeria," *International Journal of Law and Society*, vol. 4, no. 3, pp. 169–176, 2021. https://doi.org/10.11648/j.ijls.20210403.13

[2] M. Hamza, D. Amaratunga, R. Haigh, C. Malalgoda, C. Jayakody, and A. Senanayake, *Rebuilding Communities After Displacement: Sustainable and Resilience Approaches*. Springer, 2023. https://doi.org/10.1007/978-3-031-21414-1

[3] O. G. Omogunloye, N. S. Iyasele, O. A. Olunlade, O. E. Abiodun, T. J. Salami, and A. O. Alabi, "Mapping of human displacement by Boko Haram in Nigeria from 2009 to 2021," *South African Journal of Geomatics*, vol. 12, no. 1, pp. 73–85, 2023. https://doi.org/10.4314/sajg.v12i1.5

[4] A. H. Ibrahim, N. A. Yaro, M. Y. Ibrahim, and Y. Adama, "Evaluating the causes and consequences of internal displacement of persons in Kaduna State," *FUDMA Journal of Sciences*, vol. 6, no. 2, pp. 120–126, 2022. https://doi.org/10.33003/fjs-2022-0602-929

[5] K. W. Rwamparagi and N. P. K. John, "Modelling conflict-induced displacement of persons in the Democratic Republic of Congo," *International Journal*, vol. 13, no. 2, 2025.

[6] O. Yemisi, "Internal migration, exclusionary politics and crises of nation-building in Nigeria," *Àgídìgbo: ABUAD Journal of the Humanities*, vol. 13, no. 1, pp. 245–264, 2025. https://doi.org/10.53982/agidigbo.2025.1301.20-j

[7] F. Chikaji and E. Jegede, "An overview of the challenges of Nigeria Immigration Service in information management for effective documentation: The ways forward," *Lokoja Journal of Information Science Research*, vol. 2, no. 2, pp. 187–200, 2024.

[8] I. E. Ezeabasili, "Navigating insecurity and migration trends in Nigeria (2019–2023)," *International Journal of Advanced Multidisciplinary Research and Studies*, vol. 5, no. 1, pp. 159–165, 2025. https://doi.org/10.62225/2583049X.2025.5.1.3633

[9] O. S. Amoo *et al*., "Assessment of awareness and levels of preparedness for humanitarian disasters across some geopolitical zones in Nigeria," *Discover Geoscience*, vol. 3, no. 1, Art. no. 32, 2025. https://doi.org/10.1007/s44288-025-00139-0

[10] E. A. Alebiosu, "Displacement, humanitarian action and non-governmental organisations in northeast Nigeria," in *Identity, Social Inclusion and Sustainable Development in Nigeria*, 2024, pp. 3–35.

[11] O. S. Fatokun and M. C. Olise, "Analysing the challenges of humanitarian workers in provision of assistance to internally displaced persons (IDPs) in conflict zones: Study of North Eastern Nigeria," *Journal of Contemporary International Relations and Diplomacy*, vol. 5, no. 2, pp. 82–108, 2024.

[12] O. S. Amoo *et al*., "Assessment of awareness and level of preparedness for humanitarian disasters in Nigeria," 2024. https://doi.org/10.21203/rs.3.rs-4731196/v1

[13] A. S. Adesola and P. Akin, "Violent conflicts, forced migration and internal population displacement in Nigeria's Fourth Republic, 1999–2011," *Journal of Identity & Migration Studies*, vol. 18, no. 1, 2024.

[14] E. M. Aondover, A. U. Daushe, O. Ogunbola, and P. O. Aondover, "Media coverage of internally displaced persons in two selected newspapers in Nigeria," *Journal of Migration and Health*, vol. 11, Art. no. 100301, 2025. https://doi.org/10.1016/j.jmh.2024.100301

[15] D. Cantor *et al*., "Understanding the health needs of internally displaced persons: A scoping review," *Journal of Migration and Health*, vol. 4, Art. no. 100071, 2021. https://doi.org/10.1016/j.jmh.2021.100071

[16] D. Rizzi, G. Ciuffo, M. Landoni, M. Mangiagalli, and C. Ionio, "Psychological and environmental factors influencing resilience among Ukrainian refugees and internally displaced persons: A systematic review of coping strategies and risk and protective factors," *Frontiers in Psychology*, vol. 14, Art. no. 1266125, 2023. https://doi.org/10.1016/j.jmh.2021.100071

[17] O. D. Apuke and E. A. Tunca, "Taking us to the vulnerability scene: Television news coverage and framing of internally displaced persons in Nigeria," *Journal of Refugee Studies*, vol. 34, no. 4, pp. 4406–4427, 2021. https://doi.org/10.1093/jrs/feaa075

[18] P. C. Ekoh, P. Okah, C. Onalu, S. Ebimgbo, J. Ene, and A. A. Agha, "Experiences of abuse and neglect of older adults in internally displaced persons (IDP) Camp, Lugbe, Abuja, Nigeria," *Journal of Population Ageing*, vol. 15, no. 4, pp. 943–957, 2022. ttps://doi.org/10.1007/s12062-022-09388-0

[19] S. Stankovic, J. Ecke, and E. Wirtz, "Forced migration," *Oxford Research Encyclopedia of Anthropology*, 2021. https://doi.org/10.1093/acrefore/9780190854584.013.468

[20] G. Palattiyil, D. Sidhva, A. Seraphia Derr, and M. Macgowan, "Global trends in forced migration: Policy, practice and research imperatives for social work," *International Social Work*, vol. 65, no. 6, pp. 1111–1129, 2022. https://doi.org/10.1177/00208728211022791

[21] L. Kiss, D. Fotheringhame, N. Kyegombe, A. McAlpine, L. Abilio, A. Kyamulabi, E. J. Walakira, K. Devries, and C. Tanton, "Violence, abuse and exploitation among trafficked women and girls: A mixed-methods study in Nigeria and Uganda," *BMC Public Health*, vol. 22, no. 1, Art. no. 794, 2022. https://doi.org/10.1186/s12889-022-13021-2

[22] L. N. Smart, D. Amaratunga, and R. Haigh, "Internal displacement in Nigeria: What are the preventive measures?" *Rebuilding Communities After Displacement: Sustainable and Resilience Approaches*. Springer, 2023, pp. 165–189. https://doi.org/10.1007/978-3-031-21414-1_8

[23] E. B. Inyang and E. N. Effiong, "Humanitarian intervention in Nigeria: A case of internally displaced people in Benue State, 2018–2021," *SSRN*, Paper 3998209, 2022.

[24] F. F. Daodu, T. Shabu, T. I. Kile, and J. Enefu, "Humanitarian crises in internally displaced persons (IDPs) camps in Benue State, Nigeria: A comprehensive analysis," *International Journal of Research and Innovation in Social Science*, vol. 8, no. 5, pp. 2282–2298, 2024. https://doi.org/10.47772/IJRISS.2024.805166

[25] E. Jafer, G. Imana, Z. Doda, and A. Lemessa, "Post conflict-induced displacement: Human security challenges of internally displaced persons in Oromia Special Zone Surrounding Finfinne, Ethiopia," *Cogent Social Sciences*, vol. 8, no. 1, Art. no. 2103252, 2022. ttps://doi.org/10.1080/23311886.2022.2103252

[26] E. Uzobo and R. Akhuetie, "Food security and health challenges among internally displaced persons in Nigeria," *Nigerian Journal of Sociology and Anthropology*, vol. 16, no. 1, pp. 47–71, 2018. https://doi.org/10.36108/NJSA/8102/61(0140)

[27] F. Babajide, M. Olaogun, L. Oluwadele, J. Ijeoma, and S. Idoko, "The state of internally displaced persons (IDPs) in Nigeria and implications to national security: Northwest and North Central in focus," *IOSR Journal of Humanities and Social Science*, vol. 29, no. 1, pp. 1–1, 2024. https://doi.org/10.9790/0837-2911020113

[28] J. Babajide and A. Victor, "Internally displaced persons, human security and challenges of state failures in Nigeria," *Journal of Peace Studies and Practice*, vol. 13, no. 3, pp. 121–132, 2024.

[29] T. S. Khai, "Vulnerability to health and well-being of internally displaced persons (IDPs) in Myanmar post-military coup and COVID-19," *Archives of Public Health*, vol. 81, no. 1, Art. no. 185, 2023. https://doi.org/10.1186/s13690-023-01204-1

[30] E. H. Emmanuel, C. K. Okoli, and W. G. Yilkang, "An analysis of the causes and impacts of internal displacement in Nigeria," *Journal of Refugee Law and International Criminal Justice*, vol. 3, no. 1, 2025.

[31] O. P. Oluwasanmi, "Ethno-religious crisis and the challenges of internally displaced persons in Nigeria," *Journal of Living Together*, vol. 9, no. 1, pp. 29–44, 2024. https://doi.org/10.2139/ssrn.4576797

[32] O. I. Brigid, O. Boniface, and A. E. Okonkwo, "Crisis of banditry and the internally displaced persons in Nigeria: A political economy approach," *Scholars Journal of Economics, Business and Management*, vol. 11, pp. 247–256, 2022. https://doi.org/10.36347/sjebm.2022.v09i11.001

[33] C. M. Anierobi *et al*., "Communal conflicts in Nigeria: Assessment of the impacts on internally displaced persons and settlements amidst COVID-19 pandemic," *Heliyon*, vol. 10, no. 9, 2024. https://doi.org/10.1016/j.heliyon.2024.e30200

[34] M. Bojaruniec, "Is this already violence? Defining violence and domestic violence and its impact on awareness of the problem," *Journal of Modern Science*, vol. 53, no. 4, pp. 291–307, 2023. https://doi.org/10.13166/jms/177203

[35] S. Scquizzato, "Data integration and data-driven evaluation for the non-profit sector: A study case: Response Innovation Lab," 2024.

[36] S. Bag, M. S. Rahman, G. Srivastava, M. Giannakis, and C. Foropon, "Data-driven digital transformation and the implications for antifragility in the humanitarian supply chain," *International Journal of Production Economics*, vol. 266, Art. no. 109059, 2023. https://doi.org/10.1016/j.ijpe.2023.109059

[37] S. Gaikwad, S. Iyer, D. Lunga, T. Yabe, X. Liang, B. Ananthabhotla, N. Behari, S. Guggilam, and G. Chi, "Data-driven humanitarian mapping and policymaking: Toward planetary-scale resilience, equity, and sustainability," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022. https://doi.org/10.1145/3534678.3542918

[38] Y. O. Adebayo, O. O. Adeusi, J.-P. Adjadeh, S. M. Obion, and R. O. Abdulsalam, "Artificial intelligence and predictive analytics to develop evidence-based migration policies for optimal integration and economic empowerment of migrants," *World Journal of Advanced Research and Reviews*, vol. 25, no. 1, pp. 2147–2155, 2025. https://doi.org/10.30574/wjarr.2025.25.1.0286

[39] K. Hoffmann Pham and M. Luengo-Oroz, "Predictive modelling of movements of refugees and internally displaced people: Towards a computational framework," *Journal of Ethnic and Migration Studies*, vol. 49, no. 2, pp. 408–444, 2023. ttps://doi.org/10.1080/1369183X.2022.2100546

[40] N. H. M. Hussain, A. Anuar, T. Mohd, S. Masrom, and N. A. Ahmad, "Machine learning of the reverse migration models for population prediction: A review," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 5, pp. 1830–1838, 2021. https://doi.org/10.17762/turcomat.v12i5.2197

BIOGRAPHIES OF AUTHORS

Ajayi Ore-Ofe is a lecturer at the Department of Computer Engineering, Ahmadu Bello University, Zaria, Nigeria. He received his MSc and Ph.D from Computer Engineering in Control Engineering, in 2017 and 2022 respectively. He received his MSc and Ph.D from the department of Computer Engineering in Ahmadu Bello University, Zaria, Nigeria. He is mainly research in control engineering. He can be contacted at email: ajayi.oreofe17@gmail.com.

Abduljalil Murtala Ahmad is an Electrical Engineering student at Ahmadu Bello University, Zaria, Nigeria, with a strong interest in high voltage, HV engineering, artificial intelligence, and intelligent systems. He holds a Professional Diploma in Applied Artificial Intelligence from Baze University, Abuja, and has completed several professional courses on Cisco Networking and IT Masters. His current project work on HV engineering plays a vital role on the stability of today's power systems. He continues to pursue advanced studies in AI and machine learning to further expand his expertise and apply these skills to practical and research-based innovations, particularly in robotics and intelligent automation. He can be reached via email at abduljalilmurtala4@gmail.com.

Abubakar Umar is a lecturer in the Department of Computer Engineering at Ahmadu Bello University, Zaria, Nigeria. He earned his BEng, MSc, and Ph.D. degrees from the same department, specializing in various aspects of computer engineering. His primary research focus is in Control Engineering, where he explores the development and optimization of control systems for different applications. He is dedicated to advancing his research and contributing to academic knowledge in this field. He can be contacted via email at abubakaru061010@gmail.com.

Ibrahim Ibrahim is a Computer Engineering graduate at Ahmadu Bello University, Zaria, Nigeria, with a strong interest in Data Science, Machine Learning, and Artificial Intelligence. He is passionate about using technology to solve real-world problems and is continuously expanding his expertise through self-learning and professional courses. In addition to his focus on AI, he is also developing skills in data analysis and backend development to build a solid foundation for intelligent system design. His goal is to become an innovative AI Engineer capable of applying data-driven solutions to both practical and research-based challenges, contributing meaningfully to technological advancement and social impact. He can be reached via email at eebrahiim04@gmail.com.

.