Global Epidemiology 11 (2026) 100241

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology

Modelling seizure-related predictors of epilepsy diagnostic gap in two urban informal settlements of Nairobi using machine learning

Daniel Mwanga [a,b,c,*], Frederick Murunga Wekesah [a,d], Frank Ouma [a], Symon M. Kariuki [a,c,e], Joan Kinuthia [a], Peter Otieno [a,f,g], Thomas Kwasa [h], Quincy Mongare [h], Abigael Machuka [h], Steve Cygu [a], Samuel Iddi [a], Gabriel Davis Jones [c,i], Arjune Sen [c,j], Charles R. Newton [c,e,k], Gershim Asiki [a,l], Damazo T. Kadengye [a], for the EPInA Study Group[1]

[a] *Research Division, African Population and Health Research Center, Nairobi, Kenya*
[b] *Department of Mathematics, University of Nairobi, Nairobi, Kenya*
[c] *Center for Global Epilepsy, University of Oxford, Oxford, United Kingdom*
[d] *Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, the Netherlands*
[e] *Neurosciences Department, Kenya Medical Research Institute Wellcome Trust Research Programme, Kilifi, Kenya*
[f] *Department of Global Health and Population, Lown Scholars Program, Harvard T.H. Chan School of Public Health, Boston, MA, United States of America*
[g] *Department of Public & Occupational Health, Amsterdam UMC Locatie AMC, Amsterdam, the Netherlands*
[h] *Department of Clinical Medicine and Therapeutics, University of Nairobi, Nairobi, Kenya*
[i] *Oxford Digital Health Labs, Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, United Kingdom*
[j] *Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom*
[k] *Department of Psychiatry, University of Oxford, Oxford, United Kingdom*
[l] *Department of Women's and Children's Health, Karolinska Institute, Stockholm, Sweden*

ARTICLE INFO

Keywords:
Convulsive epilepsy
Non-convulsive epilepsy
Epilepsy diagnostic gap
Machine learning
Nairobi urban health and demographic surveillance system (NUHDSS)
Seizures
Treatment gap

ABSTRACT

Background: There is a wide gap in epilepsy diagnosis, particularly in low- and middle-income countries. We used machine learning models to identify seizure-related factors associated with the epilepsy diagnostic gap within the Nairobi Urban Health and Demographic Surveillance System (NUHDSS), Kenya, to inform effective community-level interventions.

Methods: Data were drawn from a two-stage, population-based census. In Stage-I, 56,425 residents of NUHDSS were screened for possible convulsive and non-convulsive epilepsy using a standardized questionnaire. In Stage-II, individuals who screened positive were invited for clinical assessment and diagnostic confirmation by neurologists. We used latent class analysis to classify symptom patterns. Seven machine learning models were trained, with extreme gradient boost and random forest models achieving the highest area under the receiver operating characteristic curve (98 %).

Results: A total of 528 individuals were diagnosed with epilepsy, among whom 80 % (n = 420) had not been previously diagnosed. The epilepsy diagnostic gap was 100 % (n = 160/160) in persons with non-convulsive epilepsy, meaning that none of them had been diagnosed before the survey. Among those with convulsive epilepsy, the diagnostic gap was 71 % (n = 260/368). Experiencing fewer types of seizure symptoms, non-convulsive seizures, or seizures with subtle features, such as those involving only one body part and those whose first experience of a seizure was recent, were associated with a wider epilepsy diagnostic gap.

Conclusion: There is critically huge diagnostic gap for epilepsy in Nairobi's informal settlements. People with subtle, fewer or less obvious seizure types are more likely to be undiagnosed. These findings highlight the importance of seizure symptom characteristics in understanding patterns of underdiagnosis. Thus, approaches to

* Corresponding author at: P.O. Box 10787-00100, Nairobi, Kenya.
E-mail address: dmwanga@aphrc.org (D. Mwanga).

[1] EPInA Study Group Members

Albert Akpalu, Antony Godi, Arjune Sen, Bruno Mmbando, Charles R. Newton,Cynthia Sottie, Damazo T. Kadengye, Dan Bhwana, Daniel Mtai Mwanga, Daniel Nana Yaw Abankwah, David McDaid, Dorcas Muli, Emmanuel K. Darkwa, Frederick M Wekesah, Gergana Manolova, Gershim Asiki, Guillaume Pages, Hellen Cross, Henrika Kimambo, Isolide Massawe, Joan W. Kinuthia, Josemir W. Sander, Mary Bitta, Mercy Atieno, Neerja Chowdhary, Patrick Adjei, Peter O. Otieno, Richard Walker, Sabina Asiamah, Samuel Iddi, Simone Grassi, Sloan Mahone, Sonia Vallentin, Stella Waruinge, Symon M. Kariuki, Tarun Dua, Thomas Kwasa, Timothy Denison, Vivian Mushi, William Matuja

https://doi.org/10.1016/j.gloepi.2025.100241
Received 25 September 2025; Received in revised form 24 December 2025; Accepted 24 December 2025
Available online 29 December 2025

reducing the diagnostic gap should take into consideration subtle and non-convulsive seizure presentations, such as training on symptom recognition and timely care-seeking.

Background

The World Health Organization (WHO) estimates that approximately over 50 million people live with epilepsy globally, of whom >80 % reside in the low- and middle-income countries (LMICs) [1] where diagnostic and care services are often limited [2]. Approximately 70 % of the people living with epilepsy (PWE) can live seizure-free with appropriate and cheap anti-seizure medications (ASMs) [1], yet LMICs have high epilepsy treatment gaps (ETG), defined as the proportion of people with active epilepsy who do not receive appropriate ASM [3]. In LMICs, the ETG ranges between 60 and 90 % [3,4]. Yet, many of these estimates overlook the Epilepsy Diagnostic Gap (EDG), which represents the proportion of individuals with epilepsy who have not received a formal accurate diagnosis of epilepsy [5]. A crucial first step towards narrowing the ETG and ensuring appropriate and timely treatment is the accurate diagnosis of epilepsy. The EDG contributes to delayed access to care, disability and poor quality of life [6].

In Kenya, like is the case with other LMICs, and particularly in urban informal settlements, access to healthcare is hindered by several factors. This includes health system weaknesses, limited awareness in the community about epilepsy and its symptoms resulting from low literacy levels, lack of trained experts, insufficient diagnostic capacity particularly for non-convulsive epilepsy, stigma, and lack of prioritization of epilepsy at governance and policy levels [7]. While population-based surveys have previously estimated the factors associated with the burden of epilepsy [8–10] and the ETG [3,11], there has been limited focus on identifying seizure-related factors that predict the EDG in the community.

Thus, identifying the factors that hinder timely diagnosis is essential to ensure PWE receive the necessary care. Efforts to bridge the EDG and enhance access to diagnosis and care have included the development of diagnostic tools and apps [7,12] and training primary healthcare workers to identify and diagnose epilepsy. The training and equipping of primary healthcare workers gives them the capacity to diagnose and manage epilepsy cases [13], or refer complicated cases to specialists. Identifying factors associated with EDG can inform the development of these task-sharing strategies and mobile-based epilepsy diagnosis companions so that they are tailored for the right users or populations. The decision to seek care is often related to the perceived severity of the condition, but there is a lack of research on how different seizure presentations and epilepsy factors impact care-seeking behaviors and, consequently, the likelihood of receiving a diagnosis.

We employed machine learning (ML) methods to analyze seizure characteristics and epilepsy factors to identify features most predictive of previously undiagnosed epilepsy [14]. Our objectives were twofold: gain a deeper understanding of the patterns of seizure and epilepsy characteristics that contribute to missed diagnoses, considering the different types of epilepsy – convulsive epilepsy (CE) and non-convulsive epilepsy (NCE); and to utilize these insights to design models that will inform more targeted strategies for improving seizure and epilepsy detection within communities, enhance community engagement and epilepsy awareness initiatives, and improve capacity-building interventions at the primary care level in the identification and management of epilepsy. Further, epilepsy studies have historically focused on rural populations, but changes in population dynamics mandate a better understanding of epilepsy in urban communities.

Materials and methods

Study setting

This study was conducted in the two urban informal settlements, Viwandani and Korogocho, which constitute the Nairobi Urban Health and Demographic Surveillance System (NUHDSS). These settlements, typical of many Nairobi's urban poor areas, face challenges such as limited infrastructure, poor sanitation, overcrowding, high unemployment, poverty, and insufficient healthcare facilities. Within the NUHDSS, Viwandani has a more mobile population, largely comprised of workers from nearby industrial companies. Korogocho has a more stable population, with many residents having lived there since birth. Further details about the NUHDSS have been published elsewhere [15–17]. These two distinct settings offer a valuable context for investigating the epilepsy diagnostic gap within informal urban settlements in a sub-Saharan African country.

Study design

This analysis uses data from the Epilepsy Pathway Innovation in Africa (EPInA) project (Protocol reference: NIHR200134) [18]. The EPInA project is led by the University of Oxford in collaboration with a consortium of partners who include the African Population and Health Research Center (APHRC), which runs the NUHDSS site, Kenya; the Kenya Medical Research Institute-Wellcome Trust Research Programme (KEMRI-WTRP), which runs the Kilifi Health and Demographic Surveillance System (KHDSS) site on the coast of Kenya; the University of Ghana, which runs the Accra site in Ghana; and the National Institute of Medical Research (NIMR), which runs the Mahenge site in Tanzania. The project aimed to promote epilepsy prevention, diagnosis, treatment, and awareness programs in four sites in Kenya, Tanzania and Ghana.

Data were collected through a two-stage population-based (census) epilepsy cross-sectional prevalence survey conducted within the NUHDSS. In stage I, trained field interviewers administered a standardized and validated 14-item screening questionnaire [10,19] to the head of each household or another adult representative if the head of the household was absent at the time of the interview. The questionnaire (see supplementary material, Table S1) was adapted from previously validated tools (with inputs from epilepsy experts) and designed to capture symptoms of both CE and non-convulsive epilepsy NCE, see Mwanga et al. (2024) [10]. The questionnaire was validated within the NUHDSS context as previously reported [10]. Field interviewers included a mix of staff with healthcare training or community service, all with at least a first degree. Training was done for one week and covered topics on study overview, question-by-question review of the questionnaire to ensure they understand each question adequately, research ethics and consent forms. The training also included a day for pilot testing the tools within a community, different from where the study would be conducted. This was to ensure the interviewers are adequately prepared for the data collection. Throughout the data collection, routine data quality assurance was conducted by running data quality check (QC) scripts to identify any possible errors in the data such as duplicated records, inconsistent entries, and prompt corrections done where needed. Supervision visits were conducted regularly and weekly check-in meetings conducted with all interviewers.

Socio-demographic information, including age, sex, education level, employment status, and marital status were also gathered for all household members. All residents of the NUHDSS were screened using this standardized epilepsy screening questionnaire [19]. Individuals who indicated a potential seizure history were classified as possible

cases of epilepsy.

In stage II of the survey, participants identified as possible cases during the initial screening were invited for a neurological assessment by neurologists. The criteria for the clinical assessment and final diagnosis of epilepsy has been published elsewhere [10]. Briefly, a diagnosis of epilepsy was made if there was a history of two unprovoked seizures occurring 24 h apart. In cases where one had both CE and NCE, the final diagnosis was considered to be CE. The assessments for diagnosis of epilepsy were scheduled, and participants who did not attend their appointments were followed up in person using the confidential contact and residential information they had provided in the first stage. Follow-up was done up to three times after which if they did not attend their appointment, they would then be classified as lost to follow-up. The stage I screening phase took place between September 17,2021 and December 23, 2021, while the assessments and confirmation of diagnosis by neurologists (stage II) were done between April 12 and August 6, 2022.

The analytical dataset

In this study, we analyzed data consisting of participants who were confirmed to have epilepsy. The primary outcome was the epilepsy diagnostic gap. Variables considered in this study included seizure symptoms used to screen participants at the first stage [10,19]. Specifically, the questionnaire captured a range of seizure-related symptoms, including episodes of unexplained falls, loss of consciousness, tongue biting, bladder incontinence, localized or generalized shaking, sensory disturbances (such as, abnormal smells), seizure duration and associations with fever or specific body part involvement.

Statistical analysis

Descriptive statistics, including frequency distributions and proportions, were used to summarize the characteristics of the study population and to explore the distribution of key variables across relevant subgroups. For continuous variables, we reported the median and interquartile range (IQR) to account for potential skewness in the data. Categorical variables were compared across groups using the Pearson's Chi-squared test. Latent class analysis (LCA) was used to categorize participants into several clinically interpretable classes of symptoms. We determined the number of classes iteratively by comparing multiple models based on the Bayesian information criterion (BIC) (Fig. 1) combined with intuitively examining the models with clearly distinct and clinically interpretable classes. Data management, statistical analysis, and visualization were conducted using R software (version 4.2.3) [20]. Specific key R functions and packages used included *dplyr* for data management, *poLCA* for latent class analysis, and *ggplot2* for visualization. Machine learning modelling was done using *caret* package.

Training and evaluation of machine learning models

Seven machine learning models were trained and used to determine the most important predictors of EDG. These included logistic regression, decision tree, naïve Bayes, random forest, gradient boosting machine, extreme gradient boost, and support vector machine. Detailed mathematical descriptions of these models are provided elsewhere [21]. These were selected because they are common and four of them have previously been used in other epilepsy studies with satisfactory or acceptable performance [7,12].

The data were randomly partitioned into training and testing subsets in a 7:3 ratio. To minimize confounding, both datasets were balanced by recruitment site, sex, age, education level, employment status and marital status. Models were trained using classification generated by the LCA and the sociodemographic characteristics. Categorical variables were encoded using one-hot encoding [22,23]. To address class imbalance in the outcome variable, the synthetic minority over-sampling technique (SMOTE) was applied [24], generating synthetic examples of the minority class based on nearest neighbors interpolation. Ten-fold cross-validation was used to identify the optimal set of hyperparameters for applicable algorithms.

To evaluate the predictive performance of the models, we performed bootstrapping with 500 replications as recommended by Efron B and Tibshirani RJ (1994) [25]. In each replication, test data were sampled with replacement to generate the bootstrap samples. Predictive performance metrics, including accuracy, area under the receiver operating characteristic curve (AU-ROC), and F1 score (harmonic mean of precision and recall), were computed for each bootstrap sample. The AU-ROC ranges from 0 to 1, where an AU-ROC of 0.5 represents a model with no discriminative ability (equivalent to random guessing), and higher AU-ROC values indicate better model performance. An AU-ROC of greater than 0.7 was considered acceptable [26]. F1 score ranges from 0 to 1. Higher F1 indicates a better-performing model, mostly recommended to be greater than 0.7 [27]. The cut-off points that maximized sensitivity and specificity of the models were chosen based on Youden's J statistic [28]. The metric results were summarized using descriptive statistics, including quantiles. The 95 % confidence intervals (95 % CI) were computed for each metric across the 500 bootstrap samples. Accuracy of the prediction ranges from 0 to 1, where higher values mean greater accuracy. The receiver operating characteristic (ROC) curve plots sensitivity (true positive rate) against 1-specificity (false positive rate) at

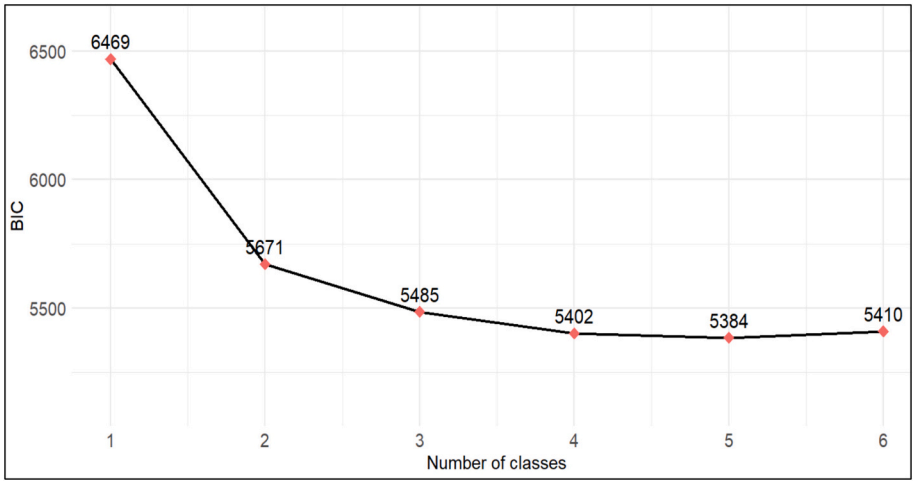

Fig. 1. BIC values of six LCA models evaluated.

various threshold values. The best-performing models were selected using AU-ROC.

Results

Sociodemographic characteristics of study participants with epilepsy diagnosis

Table 1 shows the sociodemographic characteristics and the proportions of individuals with convulsive and non-convulsive epilepsy diagnoses. Of the 528 participants who were diagnosed with epilepsy, 368 (70 %) presented with CE and 160 (30 %) had NCE. The median age was 26 years (IQR = 14–35 years), and there were approximately equal numbers of males and females. Most participants had primary or lower education (58 %), and over half were never married (55.3 %) and majority were unemployed (60 %) or engaged in informal/self-employment (32 %). Non-convulsive epilepsy was higher in Viwandani site (43 %) compared to Korogocho (16 %).

Table 1
Socio-demographic characteristics among participants classified as convulsive epilepsy and those classified as non-convulsive epilepsy.

	All § n (%)	Convulsive epilepsy ɣ n (%)	Non-convulsive epilepsy ɜ n (%)
Overall	*n* = 528	368 (69.7)	160 (30.3)
Age Categories			
0–5 years	38 (7.2)	28 (7.6)	10 (6.3)
6–12 years	77 (14.6)	56 (15.2)	21 (13.1)
13–18 years	64 (12.1)	44 (12.0)	20 (12.5)
19–28 years	135 (25.6)	95 (25.8)	40 (25.0)
29–49 years	183 (34.7)	121 (32.9)	62 (38.8)
50 years or older	31 (5.9)	24 (6.5)	7 (4.4)
Site in the NUHDSS			
Viwandani	289 (54.7)	166 (45.1)	123 (76.9)
Korogocho	239 (45.3)	202 (54.9)	37 (23.1)
Sex			
Female	267 (50.6)	184 (50.0)	83 (51.9)
Male	261 (49.4)	184 (50.0)	77 (48.1)
Highest education level			
<primary/no formal	207 (39.2)	154 (41.8)	53 (33.1)
Primary	205 (18.8)	141 (38.3)	64 (40.0)
Secondary	99 (18.8)	60 (16.3)	39 (24.4)
Post secondary	17 (3.2)	13 (3.5)	4 (2.5)
Marital status			
Never married	292 (55.3)	212 (57.6)	80 (50.0)
Married/living with a partner	168 (31.8)	109 (29.6)	59 (36.9)
Separated	68 (12.9)	47 (12.8)	21 (13.1)
Employment			
Not employed	317 (60.0)	227 (61.7)	90 (53.2)
Full-time or Part-time Employed	41 (7.8)	30 (8.2)	11 (6.9)
Self employed	73 (13.8)	41 (11.1)	32 (20.0)
Informal	97 (18.4)	70 (19.0)	27 (16.9)

Notes: § means the denominator is 528 (all participants confirmed to have epilepsy); ɣ means the denominator is the number diagnosed to have CE; and ɜ means the denominator is the number diagnosed to have NCE.

The epilepsy diagnostic gap

A total of 420 individuals did not have prior knowledge of their diagnosis. Thus, the overall diagnostic gap was 80 %, higher for NCE than for CE. Specifically, all (n = 160) the participants with non-convulsive epilepsy had not been diagnosed previously (100 % diagnostic gap) compared to 71 % (n = 260/368) diagnostic gap among the convulsive epilepsy cases. Table 2 presents the diagnostic gap by the socio-demographic characteristics. Results in Table 2 focus on the convulsive epilepsy diagnostic gap because the diagnostic gap for non-convulsive epilepsy was 100 % across all levels.

The overall diagnostic gap for CE was 75 % among males and 67 % among females. Between the two sites of the NUHDSS, the gap was 76 % in Viwandani and 66 % in the Korogocho. The EDG ranged from 59 % among those aged 13 to 18 years old to 86 % among children under 5 years old. The EDG was higher among those who were married or living together with a partner at (81 %) compared to those who had never married (67 %) and those who had separated (66 %).

Latent class analysis results

Fig. 1 presents BIC values of six LCA models with different numbers of classes. Model 1 comprised a single class, while models 2 through 6 included two to six classes, respectively.

The model with five classes had the lowest BIC (5384), but the generated classes 2 and 3 in the model did not have a clear distinction of the EDG (Fig. S2, Supporting information). The next model with the second lowest BIC was the one with four classes, but similarly, some classes were not distinct. Thus, the model that gave sufficient balance between model parsimony, the number of observations per class and the EDG (the outcome) per class was the model with three classes. It provided a clearer and more distinct classification of the symptoms and association with EDG. Based on this model (with three classes), 119 out of 368 (32 %) were classified into the first class, 29 % into the second

Table 2
Diagnostic gap by socio-demographic characteristics.

	Convulsive epilepsy diagnostic gap (N = 260)	*p*-value
Age Categories		
0–5 years	24/28 (85.7)	0.07
6–12 years	41/56 (73.2)	
13–18 years	26/44 (59.1)	
19–28 years	72/95 (75.8)	
29–49 years	78/121 (64.5)	
50 years or older	19/24 (79.2)	
Site in the NUHDSS		
Viwandani	126/166 (75.9)	**0.05**
Korogocho	134/202 (66.3)	
Sex		
Female	123/184 (66.9)	0.1
Male	137/184 (74.5)	
Highest education level		
<primary/no formal	102/154 (66.2)	0.4
Primary	106/141 (75.2)	
Secondary	43/60 (71.7)	
Post secondary	9/13 (69.2)	
Marital status		
Never married	141/212 (66.5)	**0.02**
Married/living with a partner	88/109 (80.7)	
Separated	31/47 (66.0)	
Employment		
Not employed	155/227 (68.3)	0.4
Full-time or part time employed	24/30 (80.0)	
Self employed	32/41 (78.1)	
Informal employment	49/70 (70.0)	

class, and 39 % into the third class (Fig. S1, Supporting information).

Symptom probabilities in each class are presented in Fig. S3 (Supporting information). However, for clearly distinguishing the most defining symptoms for each class, we conveniently isolated the top five symptoms in each of the three classes, and the results are displayed in Fig. 2.

The first class consisted of patients with subtle self-limiting seizures, which may not easily be noticed or recognized by patients, caregivers, or healthcare providers as indicative of epilepsy. These are symptoms localized to one body part, such as one arm or leg or in the face, without major motor involvement or manifestations. Loss of consciousness was present but not prominent in this class. The second class consisted of more dramatic and prolonged seizures including status epilepticus. The class represents individuals experiencing mixed focal and generalized seizures, with prolonged seizure durations. This included those who had lost consciousness and seizures that lasted more than 30 min. The third class consisted of patients prominently indicative of generalized seizures. The class is characterized by symptoms such as loss of consciousness, and twitching, biting tongue, loss of bladder control and falling. These are more classic and recognizable seizures.

We compared the results of latent class analysis against the seizure classification as determined by the neurologists during assessments at the health facility (stage II), see supporting information Fig. S4. Participants with focal onset seizures were more likely to be classified in the first latent class (40 %) compared to the second (24 %) and third (30 %). Participants with generalized onset seizures were more likely to be classified in the second (66 %) or third class (65 %) compared to the first class (44 %).

Machine learning modelling results

We included the classes together with sociodemographic covariates to train machine learning models to predict the risk of not being diagnosed. Fig. 3 visualizes the AU-ROC curves, and the ranking of the features used in the models. Other metrics used in the evaluation of machine learning models are presented in the supporting information Table S2.

All the models had good performance based on all the performance metrics (AU-ROC>70 %) but extreme gradient boost and random forest were the best among the models trained (AU-ROC = 98 %). The number of seizures experienced, the time duration since the first experience of a seizure, the type of symptoms one experienced as represented by the latent class analysis, site and age were ranked in the top five as the most important features by at least two of the three best-performing models. Logistic regression model, which is often preferred due to its interpretability, also had satisfactorily good performance and Table 3 thus presents both the crude and adjusted odds ratio (OR) on factors associated with the EDG, alongside the counts used to estimate crude OR.

Association of diagnostic gap and most important EDG predictors

Fig. 4 visualizes the diagnostic gap disaggregated by the main predictors of EDG based on the findings from the tests for association using Chi-square test, machine learning and latent class analysis.

The EDG was highest among participants with subtle symptoms (class 1, at 96 %), followed by those with mixed symptoms with prolonged duration of seizures (class 2, at 64 %) and lastly those with more dramatic, generalized seizures (class 3, at 54 %). The EDG decreased with duration since the first experience of a seizure, with those whose first experience was under 1 year having the highest diagnostic gap (87 %) compared to those whose first experience of a seizure was over 16 years ago. The number of symptoms was also an important predictor with those who experienced fewer symptoms (<3) having higher EDG (96 %) compared to 47 % among those who experienced 7 or more symptoms.

Discussion

In this study, we estimated the EDG and examined its predictors within the NUHDSS using LCA and machine learning. Our findings indicate an overall EDG of 80 %. This gap reached 100 % for individuals experiencing NCE and 71 % for those with convulsive epilepsy. We also observed geographic variations in the EDG between the Viwandani and Korogocho settlements, and by duration since an individual experienced their first seizure. Latent class analysis revealed three distinct symptom-based subgroups, each exhibiting different diagnostic gaps. The classes included subtle self-limiting seizures (class 1), those with mixed focal and generalized seizures with prolonged seizure duration lasting more than 30 min (class 2) and those with more dramatic generalized seizures (class 3). Most of the socio-demographic profiles of individuals with different epilepsy types were generally similar, except for age and residential location (site). Machine learning models, particularly extreme gradient boosting and random forest demonstrated the strongest ability to predict the lack of a formal diagnosis among people living with CE.

The substantial diagnostic gap, particularly for NCE (up to 100 %), highlights ongoing difficulties in recognizing epilepsy in resource-limited environments. There are also few studies on NCE, and so these seizures are often missed on national treatment guidelines. Non-convulsive seizures are often misattributed to behavioral or psychological issues [29] or misdiagnosed as vertebrobasilar insufficiency [30], leading to missed opportunities for timely diagnosis and appropriate care. The observed higher diagnostic gap in Viwandani compared to Korogocho may be linked to Viwandani's larger migrant population. Situated near Nairobi's industrial area, Viwandani hosts many short-term work migrants, contrasting with the longer-term residents of Korogocho. It is also possible that there were differences in cultural perceptions about epilepsy and healthcare systems or access between the two study settings. The elevated diagnostic gap among individuals with

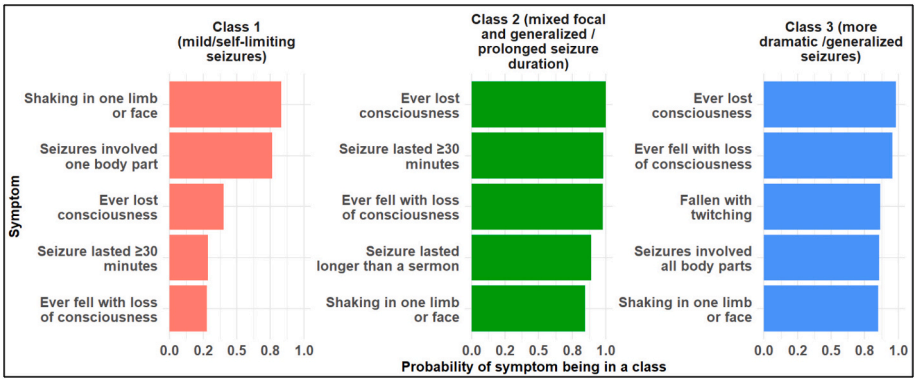

Fig. 2. Top five symptoms in each class.

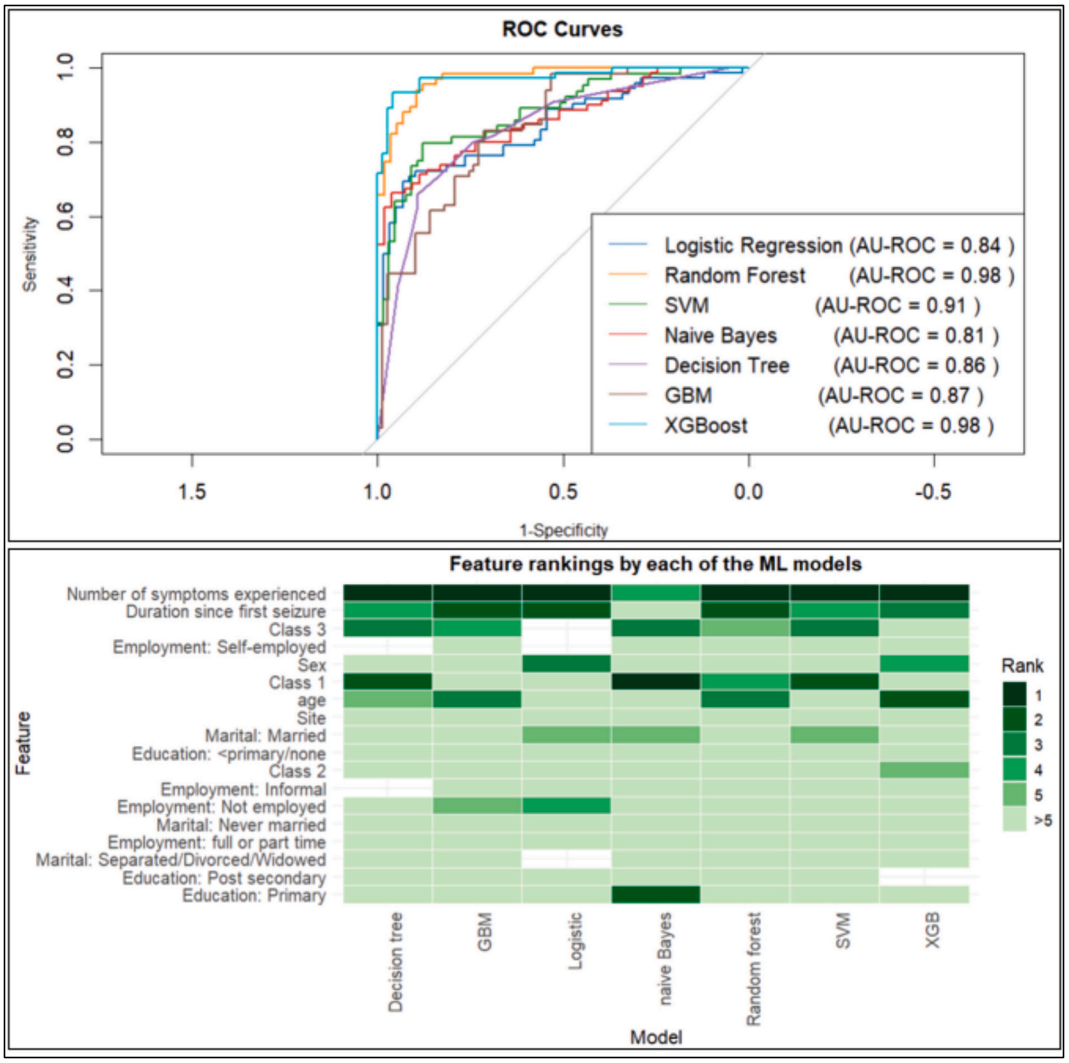

Fig. 3. AU-ROC and feature ranking for the seven models trained to predict the epilepsy diagnostic gap.

fewer or subtle and less visible features could suggest that care-seeking behaviors and decisions to seek medical attention are often triggered only when symptoms become more overt or dramatic. Seizure symptoms vary in their presentation and severity, with individuals experiencing seemingly overt convulsive seizures potentially more inclined to seek immediate medical attention compared to those with subtle or non-convulsive seizures. This can also be influenced by other psychosocial factors, including stigma and culture, necessitating more studies to delineate these associations more clearly. The public health implications of EDG is missed opportunities for treatment resulting in poor outcomes, including premature death and reduced quality of life.

Our findings indicate that the number of seizures one experienced, time since experience of the first seizure and the subtlety of the seizures and associated symptoms are the strongest predictors of an epilepsy diagnosis, followed by age of the patient. These outweighed the other sociodemographic factors. This suggests potential shortcomings in identifying possible seizures within communities and in individual decisions to seek care. Specifically, people who experience more overt and higher numbers of seizures, such as those involving falls, tongue-biting, or loss of consciousness, are more likely to receive a diagnosis because epilepsy with generalized tonic-clonic seizures, for instance is the most recognized seizure disorder in the community. This implies that less apparent symptoms may go undiagnosed, and would miss treatment. Subtle seizures such as lip smacking or eye rolling are common in children than adults, hence the finding of an association of EDG with age. Since health workers are key motivators for health-seeking behaviors [31], the findings underscore the importance of training and sensitizing primary health workers [13]. These trainings should incorporate utility of epilepsy diagnostic apps [7] at all levels on the detection of NCE. By equipping these frontline healthcare providers with relevant knowledge, we can improve early detection of seizures and consequently narrow the EDG.

We also found that the EDG was higher among those whose first experience of a seizure was more recent compared to those whose first experience was several years ago. This pattern may reflect delays in help-seeking, awareness, or access to diagnostic services in the initial stages of the condition. It underscores the critical importance of early detection and timely diagnosis of epilepsy. Early identification allows for the prompt initiation of ASM, which would result in better outcomes such as reduced seizure frequency, preventing injury, and improving long-term health and psychosocial outcomes for PWE [1]. This could be done by enhancing the capacity of primary healthcare workers to diagnose all types of epilepsy, including the use of epilepsy diagnostic applications. Equipping community health promoters with skills and tools to effectively refer potential cases to primary care can also enhance timely diagnosis and care for PWE.

Our findings align with previous research in SSA, which consistently indicates substantial diagnostic and treatment gaps for epilepsy, with treatment gap ranging from around 50 % to over 90 % [4,5]. For example, research conducted in 2012 in Kilifi, Kenya, identified an

Table 3
Crude and adjusted odds ratios from the logistic regression model on factors associated with epilepsy diagnostic gap.

	Diagnosed n (%)	Not diagnosed n (%)	OR $_{crude}$ (95 % CI)	p-value	OR $_{adjusted}$ (95 % CI)	p-value
Seizure class based on LCA						
Class 3 (more dramatic/generalized seizures)	65 (45.8)	77 (54.2)	Ref.		Ref.	
Class 2 (mixed/seizures with prolonged durations)	38 (35.5)	69 (64.5)	1.53 (0.92, 2.57)	0.1	1.52 (0.83, 2.81)	0.2
Class 1 (Subtle seizures)	5 (4.2)	114 (95.8)	19.25 (7.41, 49.99)	<0.001	4.87 (1.22. 19.36)	0.03
Number of symptoms/seizures experienced						
7–10 symptoms	86 (53.1)	76 (46.9)	Ref.		Ref.	
4–6 symptoms	18 (15.3)	100 (84.8)	6.29 (3.49, 11.33)	<0.001	4.50 (2.26, 8.94)	<0.001
1–3 symptoms	4 (4.6)	84 (95.5)	23.76 (8.32, 67.87)	<0.001	5.59 (1.22, 25.57)	0.03
Duration since first seizure						
16+ years	31 (45.6)	37 (54.4)	Ref.		Ref.	
10–15 years	23 (41.8)	32 (58.2)	1.17 (0.57, 2.39)	0.7	1.66 (0.63, 4.39)	0.3
6–10 years	25 (32.9)	51 (67.1)	1.71 (0.87, 3.36)	0.1	1.58 (0.65, 3.85)	0.3
1–5 years	22 (18.8)	95 (81.2)	3.62 (1.86, 7.04)	<0.001	3.44 (1.43, 8.26)	0.006
<1 year	7 (13.5)	45 (86.5)	5.39 (2.13, 13.63)	<0.001	2.50 (0.77, 8.12)	0.1
Age						
0–5 years	4 (14.3)	24 (85.7)	Ref.		Ref.	
6–12 years	15 (26.8)	41 (73.2)	0.46 (0.14, 1.53)	0.2	0.65 (0.16, 2.67)	0.5
13–18 years	18 (40.9)	26 (59.1)	0.24 (0.07, 0.81)	0.02	0.33 (0.07, 1.45)	0.1
19–28 years	23 (24.2)	72 (75.8)	0.52 (0.16, 1.66)	0.3	0.55 (0.16, 2.64)	0.5
29–49 years	43 (35.5)	78 (64.5)	0.30 (0.10, 0.93)	0.04	0.20 (0.04, 1.02)	0.05
50 years or older	5 (20.8)	19 (79.2)	0.63 (0.15, 2.69)	0.5	0.33 (0.04, 2.47)	0.3
Site						
Viwandani	40 (24.1)	126 (75.9)	Ref.		Ref.	
Korogocho	68 (33.7)	134 (66.3)	0.63 (0.39, 0.99)	0.05	1.04 (0.58, 1.89)	0.9
Sex						
Female	61 (33.2)	123 (66.9)	Ref.		Ref.	
Male	47 (25.5)	137 (74.5)	1.45 (0.92, 2.27)	0.1	1.22 (0.68, 2.23)	0.5
Education level						
No formal education	52 (33.8)	102 (66.2)	Ref.		Ref.	
Primary	35 (24.8)	106 (75.2)	1.54 (0.93, 2.56)	0.09	1.73 (0.80, 3.76)	0.2
Secondary	17 (28.3)	43 (71.7)	1.29 (0.67, 2.48)	0.4	1.05 (0.37, 2.93)	0.9
Post secondary	4 (30.8)	9 (69.2)	1.15 (0.34, 3.90)	0.8	0.63 (0.11, 3.58)	0.6
Marital status						
Never married	71 (33.5)	141 (66.5)	Ref.		Ref.	
Married or living with a partner	16 (34.0)	31 (66.0)	0.98 (0.50, 1.90)	0.9	1.15 (0.38. 3.52)	0.8
Separated/Divorced/Widowed	21 (19.3)	88 (80.7)	2.11 (1.21, 3.68)	0.008	1.35 (0.50, 3.67)	0.5
Employment						
Not employed	72 (31.7)	155 (68.3)	Ref.		Ref.	
Full- or part-time employed	6 (20.0)	24 (80.0)	1.86 (0.73, 4.74)	0.2	1.90 (0.79, 4.60)	0.2
Informal employment	21 (30.0)	49 (70.0)	1.08 (0.61, 1.94)	0.8	2.48 (0.81, 7.57)	0.1
Self employed	9 (22.0)	32 (78.1)	1.65 (0.75, 3.64)	0.2	2.40 (0.68, 8.47)	0.2

Notes: OR = Odds ratio; CI=Confidence interval; Ref. = Reference category.

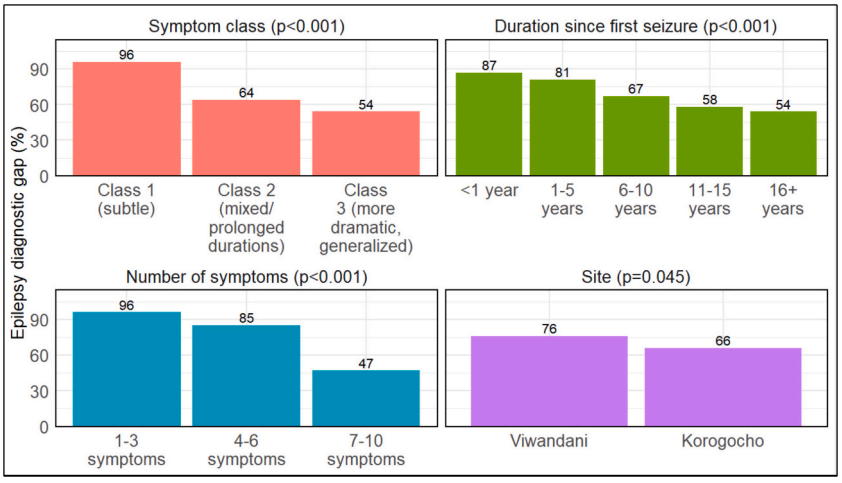

Fig. 4. The epilepsy diagnostic gap disaggregated by the most important predictors.

association between focal seizures and a lack of seeking biomedical treatment; however, their study focused on the treatment gap among individuals already diagnosed, and as our research demonstrates, the factors influencing diagnosis can differ [32]. The finding from our study and Mbuba et al.'s study [32], that focal seizures was associated with both diagnostic and treatment gap, indicates that seizure subtlety is crucial for both diagnosis and treatment. In the recent ILAE report [5], EDG is considered a subset of ETG, which is often overlooked in previous

ETG studies.

Our study offers several strengths. We utilized a large, population-based sample from a demographic surveillance site with robust follow-up, enhancing the reliability of our findings. We uniquely integrated unsupervised (latent class analysis) and supervised (machine learning) modelling techniques. This layered approach allowed us to explore both the natural clustering of symptoms and individual-level predictions of missed diagnoses, using latent class analysis to probabilistically define symptom classes, followed by machine learning and traditional statistics to confirm predictors of missed diagnoses. Another strength is the multi-stage approach with careful screening of seizure symptoms at the community (stage I) and careful phenotyping of seizures by the neurologists at stage II who confirmed the diagnosis. This enhanced diagnostic precision and minimized misclassification of epilepsy cases.

Despite these strengths, we acknowledge certain limitations. Our sample was drawn from the NUHDSS, which may limit the generalizability of our findings to rural or peri-urban areas in Kenya or other African contexts with varying diagnostic resources and epilepsy awareness. Distance to health facilities can influence care seeking and hence the diagnostic gap, but we did not have data to examine this. One of the defining features of the second latent class was prolonged seizure duration, which may be suggestive of psychogenic non-epileptic seizures (PNES); however, PNES was not explicitly modelled in this analysis. While other semiological features were also present within this class, and all participants had been classified as having epilepsy based on neurologist assessment, the possibility of PNES or mixed seizure presentations within the class cannot be fully excluded. Our sample also yielded imbalanced outcome categories, which could bias the ML models if not taken into account. We addressed this by generating synthetic data for the training dataset using the SMOTE technique to balance the two categories of the outcome.

Conclusion

Our study highlights a critically huge diagnostic gap for epilepsy in Nairobi's informal settlements, particularly for individuals with less obvious or unusual seizure presentations. These results suggest that interventions aimed at reducing this gap should prioritize community education on both convulsive and nonconvulsive epilepsy, including the recognition of diverse symptoms, causes and the importance of seeking timely care for any, even the first, seizure-like event. Future research should focus on evaluating the practicality and effectiveness of implementing simplified screening tools, potentially supported by digital platforms, to determine if community health workers can accurately identify subtle seizure patterns.

Data sharing

We welcome collaborations. The data used in this study are part of the Epilepsy Pathway Innovation in Africa (EPInA) Project, which was conducted in Kenya, Tanzania and Ghana from 2020 to 2024. The metadata including data collection tools and protocol have been documented and stored in an online repository accessible through the link https://microdataportal.aphrc.org/index.php/catalog/ under the Epilepsy Pathway Innovation in Africa project collection. Full access will be granted after the embargo period (2025–2026) or upon request approved by the EPInA lead Principal Investigator, Prof. Charles R Newton, and the APHRC-run Nairobi site's principal investigator, Dr. Gershim Asiki.

CRediT authorship contribution statement

Daniel Mwanga: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Frederick Murunga Wekesah:** Writing – review & editing, Investigation, Conceptualization, Project administration. **Frank Ouma:** Writing – review & editing, Formal analysis, Data curation. **Symon M. Kariuki:** Writing – review & editing, Investigation. **Joan Kinuthia:** Writing – review & editing, Project administration, Investigation. **Peter Otieno:** Writing – review & editing, Project administration, Methodology, Investigation. **Thomas Kwasa:** Writing – review & editing, Investigation. **Quincy Mongare:** Writing – review & editing, Investigation. **Abigael Machuka:** Writing – review & editing, Investigation. **Steve Cygu:** Writing – review & editing, Methodology, Formal analysis. **Samuel Iddi:** Writing – review & editing, Methodology. **Gabriel Davis Jones:** Writing – review & editing, Methodology. **Arjune Sen:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Charles R. Newton:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Gershim Asiki:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Damazo T. Kadengye:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Ethical considerations

The study was approved by the Scientific Ethics Review Unit (SERU) at the Kenya Medical Research Institute (KEMRI) (Reference Number: KEMRI/RES/7/3/1). All participants provided written informed consent.

Funding

We acknowledge the funding support from the National Institute for Health Research (grant number NIHR200134) using Official Development Assistance (ODA) funding for the EPInA Programme. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care.

Declaration of competing interest

None of the authors has any conflict of interest to disclose. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Acknowledgements

We acknowledge those who contributed to establishing the NUHDSS and the data collection teams at both stages of the study. Nairobi City County Health Department allowed the team to use the Nairobi public health facilities to conduct the assessments. We recognize Nelson Mbaya for supporting the programming of the study tools and managing the databases during data collection and the data collection teams at both stage-I and II of the study. All study participants are acknowledged for their participation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2025.100241.

References

[1] World Health Organization. New WHO brief sets out actions needed to improve lives of people with epilepsy. Epilepsy 2022. accessed 15 April 2023, https://www.who.int/news-room/fact-sheets/detail/epilepsy.
[2] Kissani N, Liqali L, Hakimi K, et al. Why does Africa have the lowest number of neurologists and how to cover the gap? J Neurol Sci 2022;434:120119.
[3] Mbuba CK, Ngugi AK, Newton CR, et al. The epilepsy treatment gap in developing countries: a systematic review of the magnitude, causes, and intervention strategies. Epilepsia 2008;49:1491–503.

[4] Meyer A-C, Dua T, Ma J, et al. Global disparities in the epilepsy treatment gap: a systematic review. Bull World Health Organ 2010;88:260–6.
[5] Kwon CS, Wagner RG, Carpio A, et al. The worldwide epilepsy treatment gap: a systematic review and recommendations for revised definitions - a report from the ILAE epidemiology commission. Epilepsia 2022;63:551–64.
[6] Dougherty ML. Addressing global disparities in epilepsy care. Pract Neurol 2019: 27–8.
[7] Jones GD, Kariuki S, Ngugi AK, et al. Development and validation of a diagnostic aid for convulsive epilepsy in sub-Saharan Africa: a retrospective case-control study. Lancet Digit Health 2023;5:e185–93.
[8] Ngugi AK, Bottomley C, Kleinschmidt I, et al. Prevalence of active convulsive epilepsy in sub-Saharan Africa and associated risk factors: cross-sectional and case-control studies. Lancet Neurol 2013;12:253–63.
[9] Kariuki SM, Ngugi AK, Kombe MZ, et al. Prevalence and mortality of epilepsies with convulsive and non-convulsive seizures in Kilifi, Kenya. Seizure 2021;89: 51–5.
[10] Mwanga DM, Kadengye DT, Otieno PO, et al. Prevalence of all epilepsies in urban informal settlements in Nairobi, Kenya: a two-stage population-based study. Lancet Glob Health 2024;12(8):e1323–30.
[11] Mambo R, Phiri AM, Trevisan C, et al. Gaps in treatment of epileptic seizures in a Zambian rural area. Neurol Sci 2024;45:4529–38.
[12] Duta I, Kariuki SM, Ngugi AK, et al. Evaluating the generalisability of region-naïve machine learning algorithms for the identification of epilepsy in low-resource settings. PLOS Digit Health 2025;4:e0000491.
[13] Bitta M, Kariuki SM, Mwanga D, et al. Reducing the diagnostic and treatment gap for priority mental, neurological and substance misuse disorders in primary care in Rural Kenya: results of a stepped wedge cluster randomized trial from the Epina study. Epub ahead of print. 2024. https://doi.org/10.2139/ssrn.4723079.
[14] Smolyansky ED, Hakeem H, Ge Z, et al. Machine learning models for decision support in epilepsy management: a critical review. Epilepsy Behav 2021;123: 108273.
[15] Beguy D, Elung'ata P, Mberu B, et al. Health & demographic surveillance system profile: the Nairobi urban health and demographic surveillance system (NUHDSS). Int J Epidemiol 2015;44:462–71.
[16] Emina J, Beguy D, Zulu EM, et al. Monitoring of health and demographic outcomes in poor urban settlements: evidence from the Nairobi urban health and demographic surveillance system. J Urban Health 2011;88(Suppl. 2):S200–18.
[17] Wamukoya M, Kadengye DT, Iddi S, et al. The Nairobi urban health and demographic surveillance of slum dwellers, 2002–2019: value, processes, and challenges. Glob Epidemiol 2020;2:100024.
[18] EPInA. National Institute of Health Research (NIHR) research and innovation for global health Epilepsy Pathway Innovation in Africa (EPInA):(Protocol reference: NIHR200134). https://epina.web.ox.ac.uk/; 2020.
[19] Placencia M, Sander JWA, Shorvon SD, et al. Validation of a screening questionnaire for the detection of epileptic seizures in epidemiological studies. Brain 1992;115:783–94.
[20] R Core Team. R: a language and environment for statistical computing. In: R foundation for statistical computing; 2022. https://www.R-project.org/.
[21] Mwanga DM, Kipchirchir IC, Muhua GO, et al. Modeling the determinants of attrition in a two-stage epilepsy prevalence survey in Nairobi using machine learning. Glob Epidemiol 2025;9:100183.
[22] Seger C. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. 2018. Retrieved from, https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-237426.
[23] Buwaider A, El-Hajj VG, MacDowall A, Gerdhem P, Staartjes VE, Edström E, Elmi-Terander A. Machine learning models for predicting dysphonia following anterior cervical discectomy and fusion: a Swedish Registry Study. Spine J 2025;25(3): 419–28. https://doi.org/10.1016/j.spinee.2024.10.010. Epub 2024 Nov 4. PMID: 39505010.
[24] Chawla NV, Bowyer KW, Hall LO, et al. Smote: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–57.
[25] Efron B, Tibshirani RJ. An introduction to the bootstrap. Chapman and Hall/CRC; 1994.
[26] Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5:1315–6.
[27] Manning CD. Introduction to information retrieval. Syngress Publishing; 2008.
[28] Ruopp MD, Perkins NJ, Whitcomb BW, et al. Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. Biom J 2008;50: 419–30.
[29] Beletsky V, Mirsattari SM. Epilepsy, mental health disorder, or both? Epilepsy Res Treat 2012;2012:1–13.
[30] Chen Y-S, Chen T-S, Huang C-W. Non-convulsive seizure clustering misdiagnosed as vertebrobasilar insufficiency. Heliyon 2020;6:e05376.
[31] Igwe WC, Umeadi EN, Echendu ST, et al. Sociodemographic factors influencing health care–seeking behavior for pediatric Epilepsy in Southeast Nigeria. J Neurosci Rural Pract 2022;13:448–52.
[32] Mbuba CK, Ngugi AK, Fegan G, et al. Risk factors associated with the epilepsy treatment gap in Kilifi, Kenya: a cross-sectional study. Lancet Neurol 2012;11: 688–96.