International Journal of Disaster Risk Reduction 127 (2025) 105693

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

Identifying social vulnerability profiles for coastal flood using supervised and unsupervised machine learning: A case study of Lekki Peninsula, Lagos, Nigeria

Adesola Akindejoye [a,*], Christophe Viavattene [a], Sally Priest [a], David Windridge [b]

[a] Flood Hazard Research Centre, Middlesex University, London, UK
[b] Department of Computer Science, Middlesex University, London, UK

A R T I C L E I N F O

Keywords:
Vulnerability indicators
Support vector regression
Expectation maximization algorithm
Permutation feature importance

A B S T R A C T

Coastal flooding disproportionately impacts households based on pre-existing vulnerability characteristics. Identifying these vulnerabilities is critical for effective flood risk reduction. Despite its significance, there is a paucity of techniques for identifying suitable Social Vulnerability Indicators at a local scale. This study investigates an evidence-based indicator approach to rank factors contributing to social vulnerability to coastal flooding using a purposive sample of 1334 flood-affected households in Lekki Peninsula, Nigeria. By integrating the Expectation Maximization Algorithm with Support Vector Regression (EM-SVR), and employing permutation feature importance, we identified distinct social vulnerability clusters and their associated indicator profiles. The findings reveal that a substantial (over 60 %) of the case study had moderate level of vulnerability, with clusters of similar rankings exhibiting variations in indicator profiles. Also, significant differences within the wards were observed across all areas, especially in Ajiran/Osapa and Maroko/Okun Alfa. The EM-SVR models were evaluated using various metrics, which revealed that the EM-SVR achieved a high R-squared accuracy across the seven clusters, ranging from 88.8 % to 95.7 % for the training set and 90.2 %–96.1 % for the testing set. Furthermore, the models demonstrated a low Mean Absolute Error, ranging from 0.051 to 0.075 for training and 0.051 to 0.077 for testing. Financial instability, poor social networks, lack of insurance, and pre-existing health conditions consistently emerged as the most influential indicators across clusters. These findings offer actionable insight for decision-makers by providing a well-structured and targeted approach to identifying vulnerable households and enhancing mitigation strategies.

1. Introduction

Coastal floods in marginalised communities have gained significant attention in recent times due to their increased impact and frequency[1–3]. The Intergovernmental Panel on Climate Change (IPCC) projected that small islands, low-lying coastal lands, and river basins are likely to experience more severe impacts[4]with communities sited in flood-prone areas being more at greater risk depending on their circumstances. These conditions play a key role in influencing those who could be at risk than the others. Intrinsic characteristics such as socio-economic status, and access to infrastructure influence how diverse groups experience the impacts of coastal floods. Recognizing that these characteristics can heighten vulnerabilities and risks associated with the impact of climate

* Corresponding author. Middlesex University, The Burroughs, Hendon, London, NW4 4BT, UK.
E-mail address: aa5288@live.mdx.ac.uk (A. Akindejoye).

https://doi.org/10.1016/j.ijdrr.2025.105693
Received 1 November 2024; Received in revised form 14 May 2025; Accepted 6 July 2025
Available online 8 July 2025
2212-4209/

change on coastlines, global advocacy efforts have intensified. In response to these growing concerns and advocacy, the Nigerian coastal states including Lagos, have implemented various strategies to mitigate coastal flooding. The Lagos State Climate Action Plan, for instance, was designed to reduce the impact of coastal flooding on its residents and infrastructure. However, despite these proactive approaches, coastal areas like the Lekki Peninsula remain highly vulnerable to frequent and intense flooding [5,6].

Disasters are a function of both hazards and vulnerabilities[7,8]. To understand the dynamic of disaster risk we must consider the various dimensions of vulnerability that may influence communities during hazardous events. Recognizing why different groups may be unequally affected is important to enhance preparedness and designing effective mitigation strategies that build adaptive capacities [3,9]. Social vulnerability involves the inherent conditions that may reduce or increase the susceptibility of groups, communities, or assets to disaster impact [3,10,11]. These conditions include socioeconomic components (financial stability, insurance, education), adaptive capacity (social network, post-flood support availability, length of residences), dwelling characteristics, language and health status. Studies have shown that certain neighbourhoods and social groups may be more vulnerable to the impact of a flood than others, such as low-income households, children, the elderly, and individuals with prolonged illnesses[12–14]. These impacts are noticeably more severe in coastal urban built environments, particularly in developing countries[15], where rapidly growing populations heavily rely on critical infrastructure for the smooth functioning of settlements and socio-economic activities[16]. Their geographic location exposes them to a higher frequency of flood events, threatening the performance and functionality of these critical infrastructures and resulting in disruption. These disruptions will significantly impact the quality of life, and safety of the people, particularly the vulnerable population within the community.

Vulnerability is not a static condition; it is a dynamic process that is constantly evolving [17]. Households and communities can improve their living conditions and reduce their vulnerability to hazards through various factors such as economic growth[18], improved housing conditions[19] and infrastructural development[20]. On the other hand, changes in socioeconomic conditions, health conditions or policy decisions can increase vulnerability. For example, a coastal household might reduce its vulnerability through income growth and floodproofing its dwelling. However, a decline in income or the loss of social support networks could lead to increased vulnerability [21]. Understanding vulnerability at various spatial scales, especially the neighbourhood level, is critical for effective disaster risk reduction[22]. Among these, neighbourhood-level assessments are valuable as they capture shared resources, risks, and dependencies that may not be apparent when examining vulnerability on a larger scale[23]. They also provide insights into the spatial clustering of vulnerabilities, such as areas with concentrated, informal housing, or inadequate access to public services[24]. Despite their significance, neighbourhood-level assessments remain a challenging task, particularly in developing countries where comprehensive and up-to-date data are often limited[25]. Nonetheless, understanding local-scale vulnerability can provide valuable insights into the specific needs and capacities of individuals and households, enabling more targeted and effective interventions. By integrating vulnerability assessments across different scales, policymakers, practitioners, and researchers can develop more comprehensive and resilient strategies to address the impacts of disasters.

1.1. Vulnerability trends

Vulnerability in natural disasters is conceptualized as a multidisciplinary approach. The term vulnerability came into use in the context of disaster in the 1970s [26]when vulnerability was conceded as the actual cause of disasters [27]. However, this narrative soon changed when vulnerability was related to other components affecting the social system (i.e. social, economic, ecological, environmental, and political) in the late 1980s and the 1990s [28,29]. These ideologies have strived through the millennium, with different researchers sharing their views of vulnerability. Hence, the choice of the definition adopted in various studies was determined by the researcher's interpretation and suitability of the context of the hazards of discourse[30]. For instance, Cardona et al. [31] define Vulnerability as "the propensity of exposed elements such as people, livelihoods, and assets to suffer adverse effects from a climate disaster". This definition highlights that vulnerability is not inherent but is actively induced by exposure to extreme climate events. In contrast to this[32], viewed vulnerability as the "pre-existing, inherent characteristics or qualities of social systems that create the potential for harm." This emphasizes that the determinant of vulnerability (preexisting condition) moulds the state of vulnerability, and it is not just a consequence of exposure to hazards but rather a product of inherent characteristics or qualities that are present before such events.

In recent times, the trend of this theoretical debate on vulnerability has transformed when the UNDRR [33] defined vulnerability in terms of risk analysis as "the conditions determined by physical, social, economic and environmental factors or processes that increase the susceptibility of an individual, a community, assets or systems to the impact of hazards". They argued the multifaceted and complex nature of vulnerability. The use of the term "processes" denotes that vulnerability is not just a fixed characteristic of individuals, communities, assets, or systems, but can also arise from ongoing or dynamic processes that shape their susceptibility to hazards. These definitions of vulnerability reveal its intricate nature, as perceived by different authors. It suggests that vulnerability is not solely a socioeconomic problem or a matter of social well-being but rather a multifaceted component that affects the entire system's functionality(Downing, 1991; Cardona, 2013). Furthermore, despite the existence of various definitions of vulnerability (Adger [34] Alexander & Alexander [35] Birkmann et al.[36,37] Fatemi et al.[38] Gallopín[39]and UNDRR [33] different conceptualizations of it, the lack of agreement on their practical applications, make it challenging to arrive at a precise inference [27,40]. This difficulty is partly due to the contradictory and ambiguous nature of vulnerability [27,41,42]. Nevertheless, the [33] definition of vulnerability is adopted for the present study, as it aligns well with this research perspective on vulnerability as a multidimensional and dynamic process.

1.2. Social vulnerability indicators

Assessing the vulnerability of a social system requires a straightforward approach that focuses on examining relevant indicators. Indicators have been widely used across various domains including health sciences, policy, education, and social development due to their ability to enhance the understanding of complex issues [43]. To maintain clarity in the context of this study, it's important to differentiate between indicators and indexes. Birkmann[44] define indicators as quantifiable variables that represent specific attributes or qualities of a system, offering information or insight into its vulnerability to natural hazards. An index, on the other hand, is a single numerical value derived from a weighted combination of two or more indicators [45]. A variable refers to a measurable characteristic of a population used to construct a vulnerability indicator. In this paper social vulnerability profile (SVP) is referred to as a group of households that share a similar pattern of vulnerabilities.

The indicator-based approach is a commonly adopted method for assessing vulnerability (Barankin et al., 2021). It is commonly accepted and implemented by policymakers and stakeholders as a core component of adaptation policy initiatives (Shi, 2019). The indicator-based approach allows a quantitative estimation of vulnerability, taking into consideration the characteristics of the hazard and the type of receptors within the geographical location(Viavattene and Priest, 2014). It also offers a useful mechanism for assessing social vulnerability and comprehensively depicting its multifaceted nature in relation to hazards[46]. Scholars have developed vulnerability indicators that attempt to address various dimensions, including demographic components[47], socioeconomic characteristics[48], Adaptive capacity[49] and socio-ecological systems[50].

Despite the efficacy of vulnerability indicators, challenges still exist in determining the ideal set of indicators that accurately depict the vulnerability of a given geographic area to hazards[51]. This is because vulnerability is complex and cannot be captured by a single indicator[52]. Although various physical, social, biophysical, and economic factors influence vulnerability, their importance can vary depending on the specific location[53]. Furthermore, there is a lack of consensus regarding the selection of indicators to construct context-specific variables, statistical procedures, and assumptions[46,54]. As such, this discourse extends to questioning the validity of commonly used vulnerability models[55]. Considering that some datasets used to construct these indicators may not have similar properties. Each community possesses distinctive characteristics that differentiate it from others, thereby making it challenging. Moreover, there is uncertainty regarding the aggregation technique used in assessing the indicator-based model, and the reason for the selection of this technique is usually unspecified[56]. Nevertheless, it is evident from this discourse that vulnerabilities across communities are unevenly distributed, and the indicators used to assess community vulnerability vary widely.

The quantification of social vulnerability has evolved in empirical literature through three main perspectives on how to approach the relative importance of vulnerability factors. The first approach adopts a factor weighting technique where experts use their knowledge to assign specific weights to each factor based on its significance to vulnerability [57,58]. While this method offers a refined perspective, its major drawback of this weighting system is the subjectivity and arbitrariness introduced by the techniques[59]. The second approach is the equal weight, where all factors contribute equally to the overall vulnerability score[28,48,60]. The third approach is a combination of both scaling and weighing of vulnerability factors[61].

Existing techniques for evaluating social vulnerability have contributed to the understanding of vulnerability indicators. Several studies have adopted statistical methods such as factor analysis[62], Principal Component Analysis[63–65], regression models [66–68], integrating GIS approaches such as geographical weighting regression[65,69]. These techniques have provided valuable insights into the multifaceted nature of social vulnerability indicators, effectively simplifying complex datasets and revealing broad patterns. However, the inherent limitation within these traditional methods lies in developing granular social vulnerability profiles, particularly in determining indicators relevant to groups that may not be easily discernible at a broader analytical scale[70,71]. In addition to these drawbacks, traditional linear models often struggle with capturing the complex relationships between variables and may encounter difficulties representing complex interdependencies and non-linear dynamics within the data[72,73].

These limitations underscore the need for advanced techniques, such as machine learning, which can identify hidden patterns, handle high-dimensional data, and adapt to the complexity of social vulnerability without being constrained by rigid statistical assumptions. Machine learning provides a flexible and data-driven framework[74] that can enhance the accuracy and complement existing social vulnerability techniques. Recent advancements in machine learning have shown significant potential for predicting flood susceptibility[75,76] and mapping flood risk[77–79]. However, the application of machine learning and artificial intelligence to assess social vulnerability to flooding remains relatively under-explored. While a few researchers have attempted to apply machine learning to assess vulnerability in other hazard contexts, such as earthquakes[80], typhoons [81] there is limited research on how machine learning can be employed to analyse flood vulnerability. More importantly, there is a pressing need for an in-depth understanding of how this approach can be used to effectively identify social vulnerability indicators, profile and quantify their relevance and influence across groups. This can lead to more accurate and objective assessments of social vulnerability, particularly at the neighbourhood scale. By examining how vulnerability indicators manifest within these subgroups, we can gain a broader perspective of the conditions and characteristics that make certain people disproportionately more vulnerable than others. Thus, serving as a guide for a more targeted approach within flood-prone communities. Drawing on the insights gained from this literature review, this study addresses the identified gaps by developing a machine-learning approach for ranking the social vulnerability profile of groups within the Lekki peninsula.

We anticipate that our approach will contribute to the field by identifying the most significant drivers of social vulnerability at a granular level. To achieve this, we employed a hybrid machine-learning approach to evaluate the indicator. We ranked using permutation feature importance (PI) of eighteen indicators derived from household experiences to develop a set of social vulnerability indicators. Our proposed approach to social vulnerability indicators offers a novel method for identifying the potency and relevance of indicators within any given context using advanced analytical techniques, a concept that has not yet been fully explored in social vulnerability literature. Beyond its empirical contributions, this study is significant as it represents one of the few attempts to assess social vulnerability to coastal flooding using a hybrid machine-learning approach. This framework is designed to be scalable and adaptable for use in diverse spatial contexts, providing a robust and efficient method for vulnerability assessment.

2. Case study

The Lekki peninsula sub-region lies in the wave-dominated bright of Benin, on the southeastern plank of the Lagos metropolis. It is situated between latitude 6°27' and 6°27′30''N and longitude 3°27′ and 3°30′E in the South- Western Nigeria [82]. Stretching between 70 and 80 km in length with an average width of 10 km[83]. It is bordered to the North by the Lagos mainland and the Lagos Lagoon while to the south by the Atlantic Ocean[83]. Despite being referred to as a peninsula', it functions as an island forming parts of the barrier lagoon complex[5,84]. The Lekki peninsula is a rapidly developing area featuring various developments including private gated estates such as the Victoria Garden City, Lekki Gardens Estate and Lekki Palm Estate, other land use includes commercial institutions, healthcare facilities recreational resorts, educational institutions, and areas designated for Free Trade Zones (FTZ). These developments are distributed across various wards within the peninsula, where a ward is defined as a political and administrative subdivision of a Local Government Area. The study area map is shown in Fig. 1.

The influx of residents has led to increased uncontrolled reclamation of coastal lands, growing urban slum settlements, and sand filling of the Lagos lagoon shoreline[85,86]. These activities contribute to a rise in impervious surfaces, which in turn results in increased coastal flooding ([87]; J. N. [5]). The Lekki shoreline experiences year-round oblique long swells originating from the high latitude of the South Atlantic Ocean, rolling into the Lagos coast with a significant wave height of approximately 1.36 m. These relentless waves combine with micro-tidal actions, ranging between 1.5 and 1.9 m[88–90]. These short waves transmit from the tropical bands of 6°N to 15°S [90,91]. As April–May and August–September approach, the inhabitants of Lekki coastal communities face rising apprehension due to the historically high ocean surges that occur during these months [92,93]. Contrary to the mainland, the study area has been repeatedly exposed to coastal flooding events, incurring millions of dollars in damages, some occurring within a very short timeframe[94]. These frequent occurrences, coupled with the ongoing erosion and migration of the unstable shoreline due to coastal development, further exacerbate the problem by promoting episodic sand erosion and depositions[95–97].

Lekki is largely populated by the Yoruba and Igbo ethnic groups, with a significant minority of other ethnic groups and immigrants. It is marked by socioeconomic disparities among its residents[98]. Post-flood support in Lekki are mainly provided by private individuals such as community members, friends, and relatives, though in some instances, affected individuals may receive support from the government, particularly the Lagos State Emergency Management Agency(LASEMA) [99]. LASEMA offers immediate relief through emergency funds, temporary shelters, drainage cleaning, canal dredging, and public health services, collaborating with local

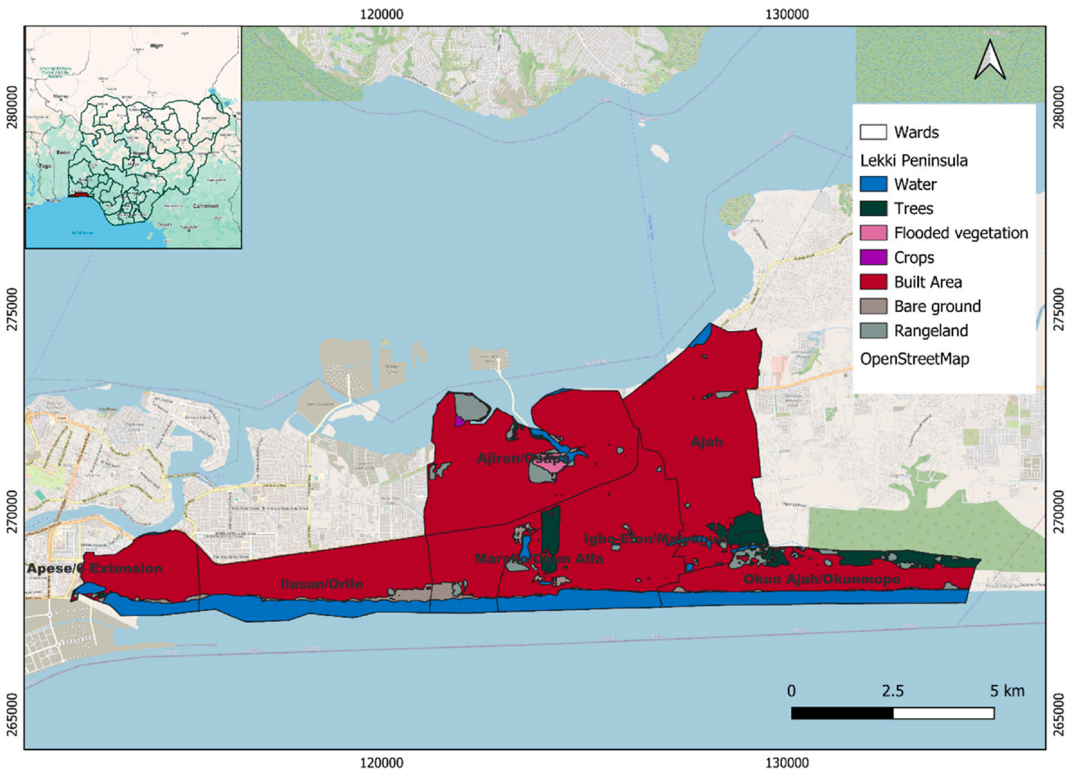

Fig. 1. The study area - Lekki peninsula.

governments for long-term reconstruction efforts. Non-governmental organizations (NGOs) may also provide support, such as food supply and psychosocial support during and after disasters, but their involvement is often inconsistent [99]. Although flood insurance is available in the state through private institutions to mitigate flood risk, it remains underutilized [100] due to factors such as affordability, inadequate awareness of its value, and mistrust of insurance organization [101].

3. Material and methods

To investigate factors influencing social vulnerability in the case study, the methodological approach presented in Fig. 2 was adopted. The initial step involved identifying and selecting the indicators employed in vulnerability literature based on a comprehensive review of existing studies. These preliminary indicators were then contextualized through consultations with seven local stakeholders, including community chiefs, local government representatives, and a landlord association representative. Given the challenges associated with acquiring local-scale secondary data that fulfils the required resolution of the data, a questionnaire survey was designed and administered. This survey was tailored to address the contextualized relevant indicators and their associated variables.

Before data collection, ethical approval was granted by the Middlesex University Ethics Committee with application reference number 21855. All participants were informed of the aim of the study and provided informed consent before completing questionnaires. They were also informed of their right to withdraw at any time. During the machine learning model training, ethical considerations related to artificial intelligence, including fairness, bias mitigation, and transparency, were addressed. These were undertaken by examining the results of the survey, checking the distribution and representativeness of key demographic variables to identify potential biases present in the dataset. The distribution of these variables was confirmed to be relatively balanced, thus negating the need for explicit balancing techniques. Also, a sensitivity analysis was conducted to ascertain whether the removal of the most influential indicator, 'financial stability,' would alter the ordering of the profiles. However, even after eliminating this predominant indicator, the relative ranking of indicators across the clusters remained consistent. Efforts were made to ensure the model's outputs were free from discriminatory patterns through the monitoring of the predictive performance metrics MAE, MSE, and RMSE across different subgroups. This process allowed the identification of disparities in prediction accuracy across these subgroups.

3.1. Population survey

The survey was administered in March–April 2024 across seven wards in Lekki Peninsula: Ilasan/Orile, Igbo-Efon/Maiyegun, Apese, Maroko/Okun-Alfa, Ajiran/Osapa, Ajah, and Okun-Ajah/Okun-Mopo. These wards were selected based on their consistent identification as high-risk areas for coastal flooding by both the Climate Central flood risk map and the Global Flood Awareness System (GloFAS) hazard map. The questionnaires were distributed using purposive sampling techniques with specific priority for households

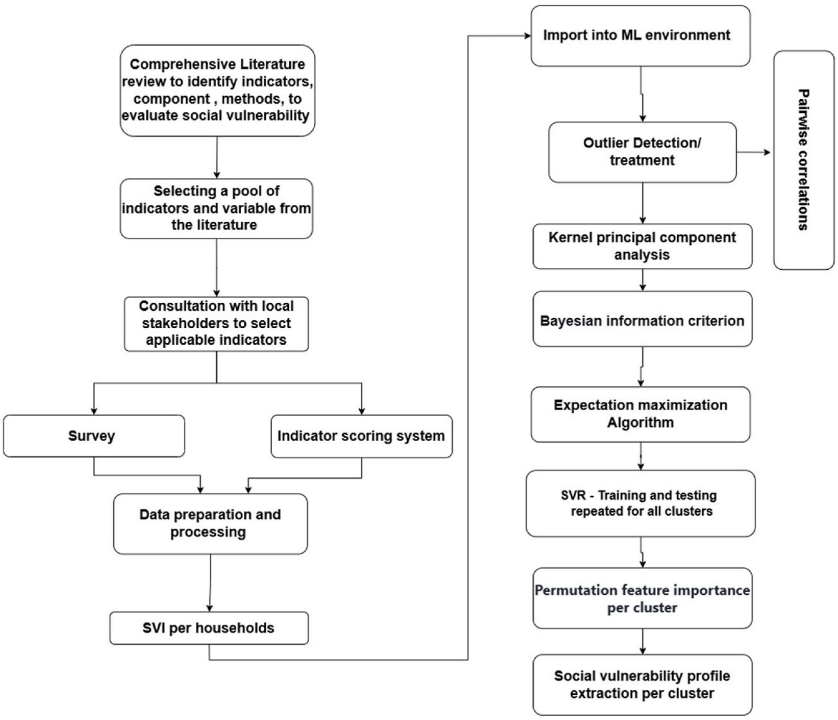

Fig. 2. Methodology flow chart.

who had been affected by the flood, as their insights were deemed valuable. The instrument was administered to heads of households above the age of 18. A total of 1334 questionnaires were completed, representing 50.43 % of the total expected sample households in the study area. While this response rate is reasonable, it may introduce some limitations regarding the representativeness of our sample. To assess the likely impact of this response rate, we compared the demographic characteristics of our sample to census data for the study area. The results of this comparison revealed that the key demographic characteristics of our sample were mostly consistent with the census data, and the sample was a reasonable representation of the entire population in the case study.

3.2. Development of indicators

A 5-point scoring system was adopted to score the eighteen indicators based on their relationship to flood vulnerability, as shown in Table 1. (1 = least vulnerable, 5 = most vulnerable). Equation (1) was used to calculate a household's overall social vulnerability index (dependent variable) by summing the scores of all indicators per household. This approach minimizes information loss and allows for a more precise understanding of social vulnerability on a household basis.

$$SVI_{hh} = \frac{1}{n}\sum_{i=1}^{n} Ind_i \tag{1}$$

Where:

n is the total number of indicators.

Ind_i is the score for the $i - th$ indicator.

SVI_{hh} is the social vulnerability index per household

The SVI dataset was imported into the machine learning environment, where a correlation matrix was created to assess multi-collinearity among the independent variables (indicators). This involved pairwise comparisons between all independent indicator pairs. Pairwise correlations were examined, with a threshold of ±0.7 used to identify collinearity. The evaluation revealed that all indicator coefficients were below this threshold (maximum absolute correlation: 0.42 between income and insurance), which suggests that multicollinearity was not a significant concern in the dataset.

3.3. Analysing social vulnerability indicators

To ensure the equal contribution of each indicator to the analysis, the initial 5-point scale ratings for each indicator were standardized using scikit-learn's StandardScaler. This normalization prevents indicators with wider value ranges from disproportionately influencing the results. While StandardScaler is a well-established technique in data preprocessing, its application in this context offers a robust and transparent method for handling multi-dimensional social vulnerability data. Kernel Principal Component Analysis (KPCA) was then employed to project the data into a higher-dimensional feature space using a sigmoid function. KPCA is a non-linear dimensionality reduction method that uses kernel functions to project data into a higher-dimensional space, enabling the extraction of non-linear principal components[102]. The KPCA eigenvalues and explained variance ratios were used to create a scree plot (Fig. 3) and the elbow method is used to determine the optimal component. This method relies on identifying the sharp descent in the plot [103], i.e. where the rate of decrease in the metric levels off, that is, other additional components provide minimal improvement. The elbow point was identified at 4, thus determining the optimal number of principal components.

These optimal numbers of principal components were then used to estimate the Bayesian Information Criterion (BIC) score, whi6ch in turn guided the selection of the optimal number of clusters. The BIC is a tool that assesses the trade-off between the goodness of fit of a model and its complexity, taking into account the number of observations and the model's likelihood[104]. It balances the log-likelihood of the model with a penalty for the number of parameters used in the model[105]. A lower BIC value indicates a more suitable model, denoting a better balance between fit and complexity[106]. As shown in Fig. 4, the lowest BIC value was observed when using seven clusters. Hence, a cluster size of seven was adopted for the subsequent analysis. The EM algorithm was used within the Gaussian Mixture Model (GMM) method to estimate the parameters (means, covariances, and mixing coefficients), enabling the clustering of the entire dataset. The EM cluster numbering was integrated back into the initial dataset to group the household-level data for SVR modelling. The average household vulnerability (computed in equation (1))index served as the prediction label, and the eighteen indicators acted as the independent variables in the SVR. SVR were run iteratively on each of the seven EM created clusters. For each cluster, we performed the following: (a) split the data into training and testing sets, (b) used GridSearchCV to determine the optimal SVR hyperparameters, (c) trained an SVR model with those optimized hyperparameters, and (d) used the trained model to predict the Household Vulnerability Index. Permutation feature importance was used to rank the importance of each indicator in the EM-SVR model per cluster.

The hybrid machine learning approach used in this study focuses on developing local social vulnerability profiles, where indicators are ranked based on their influence within specific groups. The clusters generated using the EM algorithm were formed based on distinctive patterns identified during the unsupervised learning, drifting from the geographical boundaries. The EM algorithm identifies groups with similar vulnerability profiles, irrespective of spatial proximity, thereby reducing the bias inherent in approaches that assume homogeneousness within administrative boundaries. Also, in contrast to other existing methods for assessing vulnerability indicators, PI is applied to EM-SVR models. By systematically shuffling each feature and quantifying the resulting impact on the SVR model's predictive accuracy within each identified cluster, we were able to objectively rank the indicators according to their contribution to identifying vulnerable groups and the indicators that make them most vulnerable. This methodology provides critical

Table 1
Components and indicators for measuring social vulnerability.

Social vulnerability components	Vulnerability indicators and variables	Indicator rationale to vulnerability score	SVI scoring system	Supporting references
Demographic component	• Age Number of children per household (<14 Yrs.) Number of aged parents per HH (>65 Yrs)	Households with more children between 0 and 14 and the elderly above the age of 65 are more vulnerable due to dependence and potential difficulties in self-evacuation than other age groups.	0–14 years: 5 15–19 years: 3 20–59 years: 1 60–64 years: 3 65 years above 5	[62,102–104]
	• Gender Female-led HH	Female-led households demonstrate increased vulnerability, compounded by challenges related to resource access and caregiving responsibilities that can increase the impact of flood events compared to their male counterpart.	Male: 1 Female: 5	[59]
	• Household composition HH who are lone parents with children. HH who are single parents with children and grandparents. Friends living together. Single-person HH.	Single parents without support are vulnerable to floods as they are left alone to care for their children, making it difficult to cope with the hazard without external help.	Single person: 3 Couple with other relatives: 2 Couple with children: 2 Lone parent family: 5 Friends living together:3 Single parent: 5	[105,106]
Socio-economic component	• Employment status HH heads who are unemployed. HH heads who are underemployed. HH heads who are retired. Student Households	Households with unemployed and underemployed households are more vulnerable compared to those who are fully employed and well-paid. Unemployment leads to inadequate income, severely limiting a household's ability to prepare and respond, recover following flood event.	Employed:1 Under-employed: 4 Unemployed:5 Retired:4 Student:4 Part-time:3 Self-employed:2	[4,68,107]
	• Financial stability. Financially unstable HH –	Financially unstable households are more vulnerable compared to those who are not.	Unstable: 5 Stable: 1	[43,102]
	• Income level HH earning less than 90,000 naira	Lower-income households are more vulnerable to flooding compared to high-income households	less than 90,000 Naira: 5 90,001–120,000 Naira: 3 120,001–180,000 Naira: 2 Above 180,001 Naira: 1	
Minority component	• Non-native language speakers. HH who are Non-native speaker • Immigrant HH who are immigrant	Minority groups who are non-native speakers of the dominant language tend to experience challenges communicating. There are often misunderstood evacuation instructions exposing them to greater flood risk.	English:1 Igbo: 2 Hausa:5 Yoruba:1 Others: 5 Immigrant:4	[18,108,109]
Dwelling characteristics	• Housing Characteristics HH whose dwelling unit are bungalow. HH living in a mobile or temporary structure.	Households residing in multistorey buildings are less vulnerable than those in bungalows due to easier evacuation. Bungalows don't have vertical evacuation route, making residents more susceptible to rising floodwaters, making safe evacuation challenging Households that live in temporary dwelling units made from wood, polythene and aluminium are more vulnerable to flooding compared to those who live in concrete buildings.	Bungalow:5 Multistorey:1 Two storey:1 Concrete:1 Wooden frame and polythene:5 Clay/mud: 1 Aluminum:5 Caravan:5	[110,111]
Health component	• Health conditions HH with disabled individuals. HH with individuals with prolonged illness.	Households with individuals who have long-term sickness, disabled, are more vulnerable. This is because they would require special assistance, which is limited during flooding.	Long-term illness: 5 Disability:5	[18,112,113]
Adaptive capacity	• Length of residence. HH with a short length of residences	Households with shorter residences and without prior experience of flooding tend to have limited knowledge of how to respond to floods in their new home.	Short term:5 Long term: 1	[68,114]
	• Insurances HH without any form of insurance	Households without insurance to support themselves from the losses sustained from the flood are more vulnerable.	No: 5 Yes: 1	[18,105]
	• Post flood support HH without a history of post-flood assistance	Households that have never received flood support are likely to have unmet needs, especially if they are deprived. This also indicates a systematic exclusion of	Friends:1 Families:1 Government:1 Landlord	[115,116]

(continued on next page)

Table 1 (*continued*)

Social vulnerability components	Vulnerability indicators and variables	Indicator rationale to vulnerability score	SVI scoring system	Supporting references
		certain groups of people, resulting in an unequal recovery within the community.	Association:1 NGOs: 1 Nobody: 5	
	• Social network involvement. HH without social network	Households with members affiliated with a group will have the opportunity to request support if they are affected by the flood. Households that aren't part of a group or association are more vulnerable.	No: 5 Yes: 1	[105,114]
	• Flood risk awareness HH without someone who participates in a flood risk awareness program	Households that are unaware of flooding risk in the community and have never participated in flood awareness programs are more vulnerable to floods.	Yes:1 No:5 Am not sure:3	[46,117]

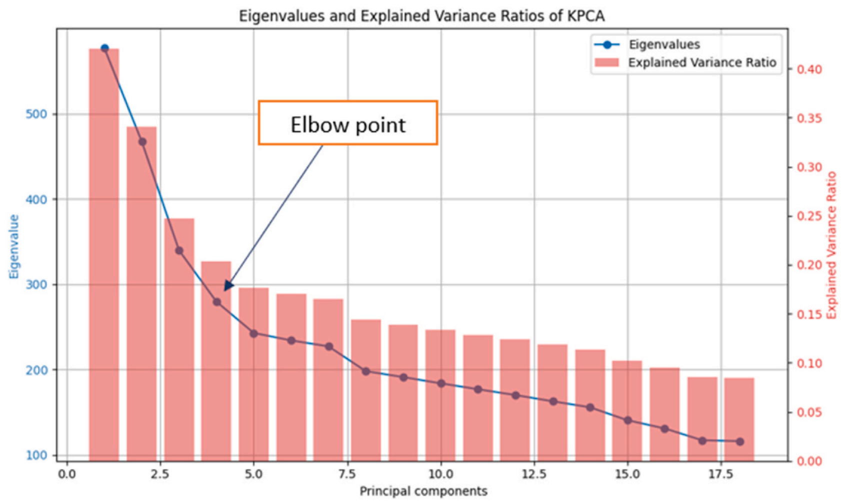

Fig. 3. Scree plot showing the Eigenvalue and Explained variance ratio of the KPCA.

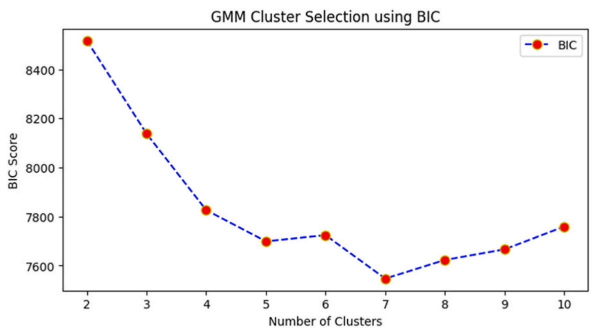

Fig. 4. Bayesian information criterion by cluster.

insight into the complex interrelationships between vulnerability factors, representing a significant advancement in our understanding of vulnerability dynamics.

3.4. Expectation-Maximization (EM) algorithm

The EM algorithm is an iterative technique used in estimating the maximum likelihood of an incomplete observation[107,108]. It is widely used for density estimation and unsupervised clustering, where data points can have partial memberships across multiple clusters[109]. The EM algorithm alternates between two steps: the Expectation (E) step and the Maximization (M) step. In the E-step, the expected likelihood is calculated by incorporating latent variables. The M-step then determines the maximum log likelihood based on the E-step's estimates, with the aim of maximizing the expected likelihood and updating the Gaussian mixture model parameters (means, covariances, and mixing coefficients). The parameters derived from the M-step are subsequently used to refine the E-step in the next iteration. This process continues until convergence[110]. Each cluster is modelled with a probabilistic distribution, reflecting the likelihood of a data point's association[108] often using Gaussian mixture models(GMM). The final cluster assignment is then determined by the highest probability.

In this study, the EM algorithm, coupled with a GMM, is employed to cluster the dataset into the optimal number of clusters, via minimization of the BIC (determined to be 7 for the current data set). GMMs require initial parameter estimates. We model the grey-level distribution of the dataset as a mixture of K Gaussian components, each representing a distinct grey-level probability density. The combined probability density is formed by the mixture of these k Gaussian components. The probability density of this mixture model, composed of k components, is expressed in equations (2)–(7).

$$f(x|\theta) = \sum_{i=1}^{k} \alpha_i f_i(x|\theta_i) \tag{2}$$

Where x is the characteristic of the vector, α_i is the weight of the mixture as $\sum_{i=1}^{k} \alpha_i = 1$, θ is the parameters $(\alpha_{1,}\ \alpha_{2,}\ldots,\ldots\ \alpha_{k,}\theta_1\ \theta_{2\ldots,}\theta_k)$ and f_i the density of the gaussian parameterized by θ_i that is $(\mu_i \sigma_i)$:

$$P(x|\theta_i) = \frac{1}{\sqrt{2\pi\sigma_i}} exp\left\{-\ \frac{(x-\mu_i)^2}{2\sigma_i^2}\right\} i = 1, 2, \ldots..N \tag{3}$$

Where $\theta_i = (\mu_i, \sigma_i)$ is the Gaussian mixture distribution parameter. On the assumption that density is derived from a gaussian mixture, We begin by fixing K a priori and then employ the Expectation-Maximization (EM) algorithm to find a local maximum of the likelihood function. The EM algorithm formula is expressed as follows:

The input is set to H = histogram, K is the gaussian number, ε = error

Output: model parameter $(p_1, p_2, \ldots.., p_k \ldots \alpha_1, \alpha_2 \ldots \alpha_k)$

Step 1. Evaluation Expectancy (E)

$$\varphi(i|x_{j,}\theta) = \frac{p_i f_i(x_j|\alpha_i)}{\sum_{k=1}^{K} p_k f_k(x_j|\alpha_k)} \tag{4}$$

Step 2. Maximization (M); GMM parameter update

$$p_i^{new} = \frac{1}{N} \sum_{j=1}^{N} \varphi(i|x_j, \theta^{old}) \tag{5}$$

$$\mu_i^{new} = \frac{1}{N} \sum_{j=1}^{N} \varphi(i|x_j, \theta^{old}) \tag{6}$$

$$\sum_i^{new} = \frac{\sum_{j=1}^{N} \varphi(i|x_j, \theta^{old})(x_j - \mu_i^{new})(x_j - \mu_i^{new})^T}{\sum_{j=1}^{N} \varphi(i|x_j, \theta^{old})} \tag{7}$$

The iteration stops when $||\ \theta^{new} - \theta^{old}|| \le \varepsilon$

The rationale for employing the EM algorithm for clustering our dataset lies in the complexities inherent to social vulnerability [111]. Since, vulnerability is not a directly observable characteristic but rather an latent factor, influenced by multiple underlying and often unobserved conditions [112]. These interconnected factors reflect hidden patterns of inequality that contribute to differential levels of vulnerability. The strength of the EM algorithm's lies in its ability to identify these latent variables by analysing observed data [113]. This capability is important for gaining an understanding of the multidimensional factors driving vulnerability. Also, this methodological approach prioritizes the categorization of households into distinct vulnerability levels. By adopting the EM algorithm in conjunction with GMM, household can be effectively grouped based on their shared vulnerability characteristics. Thereby revealing groups with varying degrees of vulnerability to flooding.

3.5. Support Vector Machine for Regression

Support Vector Machine for Regression (SVR) is a supervised learning algorithm used in regression analysis. The SVR follow the principle of the support vector machine for classification (SVC) which classifies data through an optimal hyperplane in a high-dimensional space to separate data points belonging to different classes [114–116]. This hyperplane maximizes the margin between the closest data points of each class by obtaining the maximum geometric distances through function-based distances (Y. [116]).

In contrast to the SVC, the Support Vector Regression finds a function that approximates the continuous target variable while keeping prediction errors within a certain margin (ϵ)[114]. Both SVC & SVR can employ a kernel function for estimating nonlinear relationships in data, thus offering a more flexible method for regression [114]. This is carried out via 'the kernel trick' that implicitly maps the data point into a higher dimensional space (determined by the choice of kernel function) before proceeding with the ML task as a convex optimization problem. The commonly used kernel functions include Linear functions, polynomial functions, sigmoid functions and radial basis functions (RBF)[117]. The primal formulations are used in linear SVR as shown in equation 8.

For small data quantities, studies have shown that Support vector machines (SVMs) often outperform other classifiers such as Artificial Neural Networks (ANN), and Decision Trees (DT), in terms of generalization performances [118]. In our study, we prioritized feature selection accuracy and a model with reduced complexity, leading us to choose SVMs. Moreover, the SVR has a high estimating accuracy in real-world application[119], along with relative stability and performance in handling high dimensions and low sample sizes datasets. The ability to handle high dimensionality is also important as we used 18 indicators as input features. SVMs are also robust to outliers and can model non-linear relationships through the use of kernel functions[120]. While a key drawback of SVR is its high computational complexity, especially for large datasets, to resolve this limitation, studies have suggested techniques such as Least Square SVR [121] and parameter-insensitive kernel to reduce tunning complexity[122].

The sigmoid is one of the kernel functions used in the SVR[123] that draws on the principle of neural networks[124]. Sigmoid kernel functions behave as multi-layer perceptron and can capture distinctive behaviour that may not be effective with the RBF kernel. However, its less commonly used because of sensitivity to parameters and less stable results compared to the RBF kernel [123]. These two kernels offer complementary properties visible through their strength and limitations. The RBFs are effective where local density and proximity of data points are crucial as they assign greater weight to nearby points in the feature space and capture the local structure of data[125]. These properties make it suitable for image recognition, and clustering. While on the other hand, the sigmoid can capture the global relationships[126] The RBF introduces the local density element that is not present in the sigmoid function here by complementing this limitation. Furthermore, the sigmoid kernel through hyperparameter optimization techniques can be tuned to attain similar performances as the Gaussian RBF kernel (equation (10))[127]. Where there are high dimensional feature vectors or nonlinear decision boundaries, the sigmoid kernel can perform effectively[128]. This performance is largely subjected to the number of cross-validations undertaken during the tuning and dataset. This study contributes to existing knowledge by introducing the application of a sigmoid kernel function for modelling social vulnerability as shown in Equation (14). The sigmoid kernel function was selected because it offers more flexibility for hyperparameter tuning such as gamma and regularization parameters and is less sensitive to feature scaling compared to the RBF kernel.

$$f(x) = w.x + b \tag{8}$$

Where $w \in \mathbb{R}^n$ the weight vector and $b \in \mathbb{R}$ is the bias term; the dot product $w.x$ is the linear combination of the input feature. This equation approximates the target variable while maintaining simplicity by minimizing the weight vector $||w||$ norm with the primal optimization problem for linear SVR. Similar to the soft margin loss function adopted in SVM[129] the slack variable ξ_ι *and* ξ_ι^* are introduced to address infeasible constraints[130] in the optimization problem in Equation (9):

$$\min_{w, b, \xi, \xi^*} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{9}$$

Subject to:

$$\begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_\iota \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{10}$$

Where w is the weight vector, B is the bias term, ξ_i *and* ξ_ι^* are the slack variable indicating the deviation from the epsilon tube, C is the regularization parameter which balances the flatness of the function and the amount of deviation larger than ϵ tolerated within the width of the epsilon tube, for which no penalty is assigned. The L2 regularization term is applied to prevent overfitting by penalizing large weight thus reducing the complexity of the model.

$$\left(\frac{1}{2}||w||^2\right) \tag{11}$$

By introducing the Lagrange multiplier α_i*and* α_i^* a dual formulation is derived thus eliminating the primal variable the dual formula is expressed in equation (12).

$$max_{\alpha,\alpha^*} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)\left(\alpha_{j-}\alpha_j^*\right)K(x_i, x_j) - \epsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n}y_i(\alpha_i - \alpha_i^*) \tag{12}$$

Subject to:

$$0 \leq \alpha_i - \alpha_i^* \leq C$$
$$\sum_{i=1}^{n} \left(\alpha_i - \alpha_i^*\right) = 0 \tag{13}$$

Such that $\alpha_i - \alpha_i^*$ are the Lagrange multipliers for each data point; and $K(x_i, x_j)$ is the sigmoid kernel function. The hyperbolic tangent sigmoid kernel is expressed as

$$K(x_i, x_j) = \tanh\left(\gamma\langle x_i, x_j\rangle + r\right) \tag{14}$$

Where x_i and x_j are two input vectors; γ *(gamma) is* the kernel coefficient learnt during the *training;* $\langle x_i, x_j\rangle$ *is the dot product of the vectors* x_i and x_j and r is an additional bias term. The RBF kernel is given by

$$K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{15}$$

Where $K(x_i, x_j)$ is the RBF kernel function that measures the similarity between two input vectors x_i and x_j; $\|x_i - x_j\|^2$ is the Euclidean distance between the two data points in a feature space; $\exp(-)$ is the exponential function that maintains the RBF kernel value to be positive and within the range of 0–1; σ is the parameter that controls the gaussian function that influences the kernel's locality. While both the sigmoid and RBF kernels are often employed in Support Vector Regression (SVR) to handle non-linear data, they may be used as alternatives. By offering distinct approaches to addressing non-linearity, both kernels contribute to the flexibility of SVR, allowing it to adapt to various data patterns.

3.6. Hyperparameter optimization for EM-SVR

The survey data within each of the seven cluster was split into training data (80 %) and the test dataset (20 %). The SVI- SVR model optimization was conducted using grid search. The grid search is a selection procedure, that involves the testing of multiple models and hyperparameters iteratively and validating each combination[131], with the aim to determine the best-performing combination which can be used in the model prediction. The 10-fold cross-validation was used to ascertain and evaluate the model's performance and generalization ability on unseen data. Across the seven clusters, the regularization parameter C and epsilon-insensitive tube were held constant at 0.1. The gamma *(γ)* parameter was tuned to "auto" for most clusters, except for clusters 3 and 6 that tuned to "scale".

The Gamma played a significant role in determining the performance of the EM-SVR model. The hyperparameters (C, epsilon, gamma) are mutually dependent, and their tuning was significant for optimal outcomes. The C controls the trade-off between complexity and error tolerances[132]. Thus, a greater value of C will penalise the training error more heavily, leading to poor performance on unseen data. Whereas, lower C values allow for more flexibility, tentatively leading to underfitting. The parameter ϵ reveals the width of the epsilon-insensitive tube around the predicted values. Deviations incurred within the tube have no penalty. A smaller ϵ value can lead to a more complex model that tries to fit closer to more data points, here by increasing the risk of complexity of the function. In contrast, a larger ϵ value creates a wider "tolerance zone" for deviations, resulting in a smoother model but possibly missing important data variations[132,133]. The parameter "Gamma" γ is inherent to the sigmoid kernel, it controls the influences within each data point and have an influence on model accuracy[134]. It is also responsible for shaping the model decision boundaries, such that a small gamma value will likely yield a smooth boundary[135], whereas a large gamma value may result in a complex decision boundary that tightly fits the training data, increasing the risk of an overfitted model[136].

3.7. Model evaluation criteria and validation

The EM-SVR model performance was evaluated using the Root Mean Square Error (RMSE)in Equation (16), Mean Absolute Error (MAE) in Equation (17), Mean Squared Deviation(MSD) (Equation (18)), and the coefficient of determination (R^2) in Equation (19). Also, the k-fold cross-validation method was used to ascertain and evaluate model performance and generalization abilities on unseen data. In the k-fold cross-validation, the data is randomly shuffled to avoid bias. It's then split into 10 equal-sized folds. Each fold is used for testing once, while the remaining (k-1) folds are combined for training. This process is iterated for all k folds. The stored results are then averaged to estimate the overall model performance.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_{ei} - X_{VI}\right)^2} \tag{16}$$

Where *n:* total number of observations; X_{ei} is the actual value for each observation; X_{VI} is the predicted value for each observation.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|Observed_i - Predicted_i\right| \tag{17}$$

Where N is the number of observations; $|Observed_i - Predicted_i|$ is the absolute difference between the observed value and the predicted value for each observation.

$$\mathrm{MSD} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{18}$$

Where: *n is* the number of observations; y_i is the actual observed value; $\hat{y}$ is the predicted value from the model.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{19}$$

Where SS_{res} *is* the residual sum of squares; SS_{tot} is the total sum of squares

To conclude our assessment, the indicator importance analysis of EM-SVR models for each cluster was conducted using permutation feature importance also known as randomized ablation feature importance[]. Examining the predictors' significance within each cluster revealed the key driver of social vulnerability among small groups, which otherwise might not be detectable at a broader scale. This granular analysis enhanced our understanding of the model's behaviour, leading to more reliable interpretations and facilitating data-driven decisions.

4. Result

4.1. Performance metrics and evaluation

This study used four performance metrics to assess the accuracy and goodness of fit between EM-SVR cluster models: The MAE, RMSE, MSD and R^2. The metrics provide insight into the model's performance[132,137]. The MAE reveals the average absolute deviation between the predicted and the observed social vulnerability index and highlights how close the estimates are to the actual data. The RSME on the other hand assesses the magnitude of errors, giving more weight to larger deviations. The R^2 revealed the variances of the observed SVI data that can be explained by the model, thus underscoring the model's goodness of fit. The combination of low MAE,MSD and RMSE values and a high R^2 value denotes a well-performing model[138–140].

The evaluation results shown in Table 2 reveal that MAE for the EM-SVR model across the seven clusters had consistently strong performance. The MAE was low, ranging from 0.051 to 0.075 for the training set and 0.051 to 0.077 for the test set. Similarly, the RMSE, a measure of predictive error, was also low, varying between 0.066 and 0.107 for training and 0.063 and 0.094 for testing across the clusters. The model revealed high predictive accuracy on both datasets, with R^2 values ranging from 0.888 to 0.957 for training and 0.902 to 0.961 for testing. The MSD was also low, ranging from 0.004 to 0.011 for training and 0.003 to 0.010 for testing. The consistently low RMSE, MSD, and MAE values, combined with the high R-squared values in all clusters, illustrated the reliability of the clustered EM-SVR model. This evaluation result shows that the model effectively captures the underlying patterns and characteristics of the dataset within each cluster while maintaining minimal prediction errors.

4.2. Profiling social vulnerability within clusters

Permutation feature importance (PI) was used to profile the relative importance of indicators within the clustered EM-SVR model. This technique involved shuffling the value of each feature and evaluating whether it increases the model error[]. Where a feature produces a significant error during permutation, then such a feature is more critical to the model's predictive capabilities[141]. Whereas, where a model error remains unchanged after permutation, such a feature is considered insignificant and can be ignored [142]. The PI was conducted using the test dataset to ensure an unbiased evaluation of the models. In this instance, we used the mean change in negative MSE as the performance metric in indicators assessment. A higher mean importance score signifies a stronger predictive influence of the corresponding indicator on model performance, hence its prominence within the clusters. Therefore, the results of the PI approach reveal which social indicators most significantly influence the vulnerability of each cluster.

As presented in Fig. 5, the indicators varied in their influence across the seven clusters. Financial stability was the most influential indicator in all clusters except for cluster 3. Other key indicators included social networks (significant in Clusters 1, 3, 5, and 6), length of residence (Clusters 1, 3, 4, 6, and 7), insurance coverage (Clusters 2, 3, 4, 6, and 7), language proficiency (Clusters 1, 4, and 5), and health conditions (Clusters 2, 4, and 6). Flood recovery plans and housing characteristics demonstrated moderate influence, while household composition and income had the least impact. Coping capacity and education also showed moderate influence.

An equal weighting interval scale (1–5, interval 0.8) was used to rank cluster vulnerability. The result of the assessment of SVI

Table 2
Performance metrics of EM-SVR models.

clusters	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R2	Test R2	Train MSD	Test MSD
1	0.066	0.063	0.051	0.051	0.956	0.961	0.004	0.003
2	0.08	0.076	0.065	0.061	0.945	0.943	0.007	0.003
3	0.082	0.094	0.066	0.077	0.906	0.902	0.005	0.004
4	0.095	0.082	0.074	0.066	0.917	0.928	0.009	0.009
5	0.09	0.093	0.071	0.075	0.919	0.946	0.008	0.006
6	0.107	0.063	0.075	0.053	0.888	0.917	0.011	0.01
7	0.068	0.068	0.053	0.054	0.957	0.96	0.007	0.005

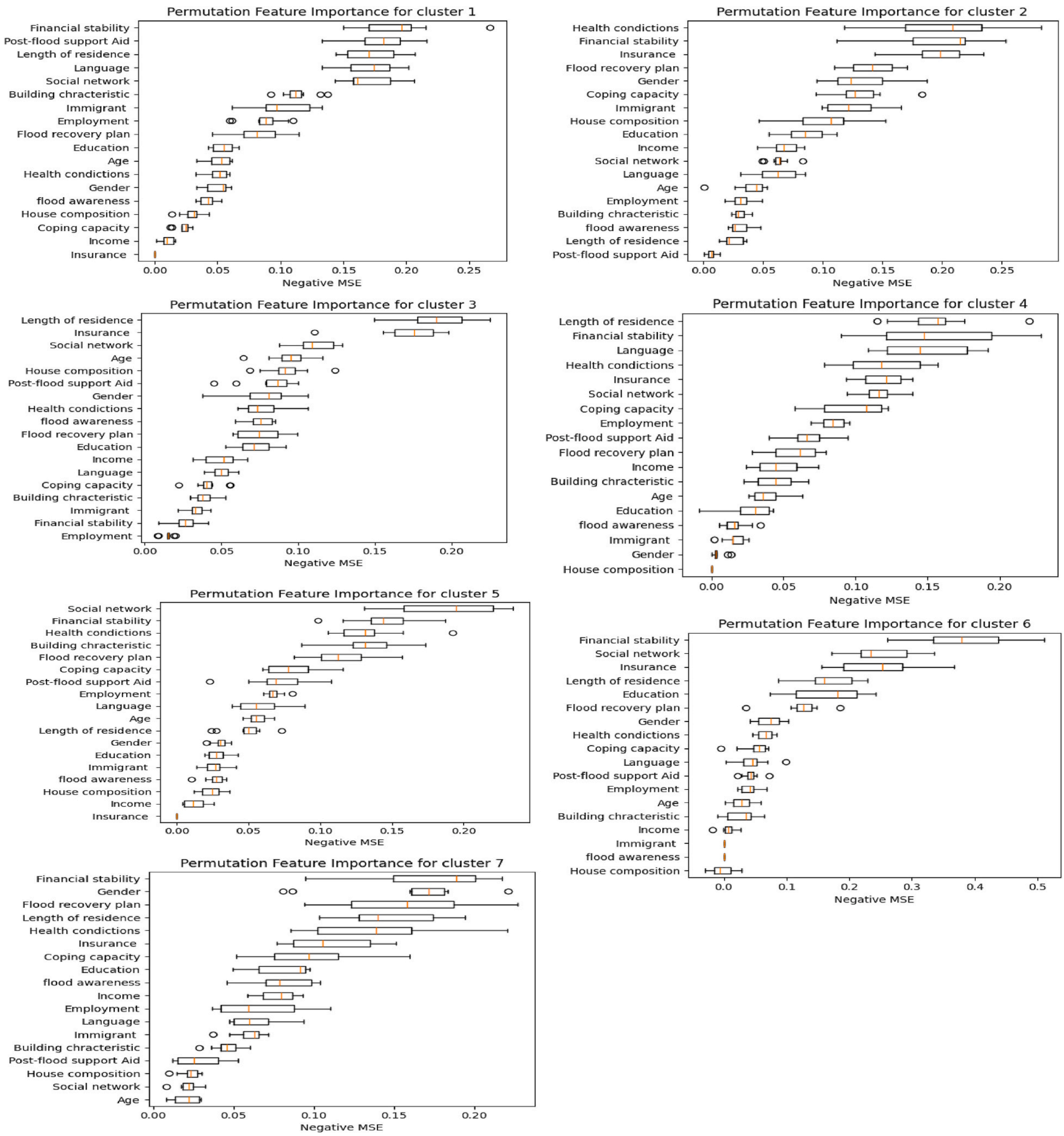

Fig. 5. The importance of each feature in the EM-SVR model.

prediction as shown in Table 3 reveal that, Clusters 4 (3.21), Cluster 5 (2.9), cluster 1 (2.84) and Cluster 7 (2.81) had moderated vulnerability. Clusters 2 (2.53), cluster 6 (2.36) and 3 (2.24) exhibited low vulnerability. There were no clusters that fell within the range of high vulnerability. The result was however expected as the observed SVI of households in the case study, had few households whose SVI exceeded 4. The standard deviation was calculated to measure the variability of the predicted vulnerability scores around the mean for each cluster. the result was somewhat comparable within a range from 0.20 to 0.26. Clusters 1, 5, and 7 had a slightly high degree of variability within the predicted vulnerability index this implies that the predicted SVI may have a wide range of vulnerability among its households.

Table 3
Social vulnerability profile and key indicators.

Vulnerability Profile	Key Indicators	SVI Mean Prediction	SVI Standard Deviation	Vulnerability Ranking
Cluster 3	A high proportion of long-term residency HH Moderate prevalence of uninsured HH HH with strong social networks A balanced mix of all age groups in HH Low presence of single-parent and lone-parent HH	2.24	0.26	Low Vulnerability
Cluster 6	Presences of financially stable HH HH with strong social networks High prevalence of insured HH A high proportion of long-term residency (HH) HH with high levels of education	2.36	0.21	Low Vulnerability
Cluster 2	Low prevalence of pre-existing health conditions Presences of financially stable HH A high proportion of insured HH Presence of flood recovery plans A low percentage of female-led HH	2.53	0.229	Low Vulnerability
Cluster 7	Prevalent financially unstable HH A high proportion of female-led HH Absence of flood recovery plans Predominantly long-term residency HH Presence of good social networks	2.81	0.264	Moderate Vulnerability
Cluster 1	High financial instability Absence of post-flood support and recovery aid High presence of long-term residency HH Presence of non-native speakers Presence of strong social networks	2.84	0.264	Moderate Vulnerability
Cluster 5	The presence of HH with strong social networks The presence of financially unstable HH Absence of pre-existing health conditions Poor housing conditions Absence of flood recovery plans	2.9	0.263	Moderate Vulnerability
Cluster 4	High proportion of long-term residency HH Presence of financially unstable HH Presences of non-native speakers HH Presence of pre-existing health conditions High prevalence of uninsured HH	3.21	0.248	Moderate Vulnerability

4.3. EM-SVR model prediction

This study investigates the primary factors contributing to social vulnerability in the study area. We used eighteen indicators to assess vulnerability. These include household age structure, gender, household composition, education, income, financial stability, employment, health conditions, housing characteristics, language, immigration status, length of stay, insurance, social network, flood awareness, post-flood support, and post-flood recovery. An analysis was conducted using the social vulnerability index per household. This served as the target variable for a Support Vector Regression (SVR) model.

The SVR model predicted household vulnerability indexes ranging from 1.6 to 4.1 across the seven identified clusters. These clusters showed a varied distribution across the wards within the case study, with all wards containing multiple social vulnerability profiles. An analysis was conducted to assess the proportionality of this distribution. Fig. 6 illustrates the vulnerability assessment of the wards. The entire Lekki exhibited a mix of low and moderate vulnerability, with Ajah showing a balanced distribution, with 50 % of

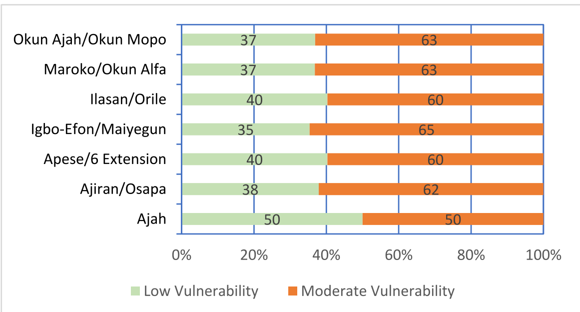

Fig. 6. Percentage of Social vulnerability prediction distribution within wards.

households in each category. Other wards demonstrated a significantly higher proportion (over 60 %) of households experiencing moderate vulnerability, particularly in Maroko/Okun Alfa, Ajiran/Osapa and Igbo-Efon/Maiyegun. However, it is important to note that the vulnerability ranking for Ajah, the ward with the balance distribution score, may be underestimated due to limited data availability from restricted access to the estates. The study reveals a heterogeneous distribution of social vulnerability within the wards, with neighbourhoods in close proximity exhibiting disparate levels of vulnerability. Even within relatively small geographic areas, there are significant variations in vulnerability levels.

5. Discussion

This study proposes a new method for assessing social vulnerability indicators and investigates the key contributing factors in the case study. We argue that geographical proximity does not necessarily equate to a shared vulnerability. Our findings reveal the multifaceted and spatial disparity of vulnerability that exists at a microscale through the profiling of vulnerability clusters. The indicators identified in the study were context-specific vulnerability variables that were developed in collaboration with local stakeholders. Contributing factors to social vulnerability, as identified in this case study, align with many established assessments [18,59,143–146] such as age (children and elderly), single-parent households, health status, educational attainment, social networks, insurance, income, language, unemployment, dwelling characteristics, and flood awareness, we also expand existing indicators by including flood recovery plans, financial stability, and length of residences. These indicators are relevant since they align with the Sendai Framework for Disaster Risk Reduction 2015.

The findings of the PI revealed that financial stability exerts the strongest influence among the social vulnerability indicators considered in the majority of the clusters. This result aligns with established literature that underscores the influence of economic inequalities, financial constraints and poverty can increase vulnerability to flood hazards[18,147,148–150]. While the case study is considered an affluent community, it is important to acknowledge the presence of pockets of impoverished neighbourhoods. Another economic indicator prominent within the model and previously highlighted in existing articles [151,152,153] is insurance. The outcome was however not surprising as the case study had a low coverage of flood insurance. Although higher-income residents may experience greater absolute losses due to the higher value of their assets, they also possess greater means to recover, often through insurance and other financial support mechanisms. In contrast, the burden of flood impacts falls disproportionately on those lacking insurance coverage and the financial resources for repairs and reconstruction. These households are more susceptible to long-term consequences such as debt accumulation and increased poverty, exacerbating their vulnerability. While it is often assumed that poverty and vulnerability to flood hazards are directly correlated, the relationship is more complex [18]. Also, other indicators distinctive in the EM-SVR models are social networks, enhanced information flow, communication, and resource sharing [59,154]. The structure and status of a community's social networks directly influence the effectiveness of knowledge dissemination, which, in turn, contributes to reducing the impact of floods, accelerating recovery processes, and supporting the restoration of livelihoods. Therefore, social networks are crucial for reducing vulnerability and well-being within flood-affected communities, particularly in mitigating social vulnerability among households.

The finding also reveals the complex interactions between these variables. For instance, the SV model highlights the importance of the length of residences. Individuals who have lived in an area for a longer time are more likely to have strong social networks and support systems. These networks can provide critical assistance during and after a disaster. On the other hand, those with shorter lengths of stay may have weaker social ties, making them more vulnerable due to a lack of awareness of local risks and resources, limited access to informal support networks and difficulty navigating unfamiliar systems for assistance after a flood event. Other important factors identified in the model include language proficiency, education, and dwelling characteristics. These findings contribute to the broader understanding of vulnerability by recognizing the importance of both multidimensional factors that shape a community's vulnerability to flooding.

The value of this approach lies in its ability to complement and enhance existing social vulnerability models by ranking complex and non-linear vulnerability factors. By analysing predictor importance, we can identify the specific prominent factors that make certain groups more vulnerable to coastal flooding than others. This approach builds upon existing methods by detecting non-linear patterns and providing interpretable feature importance, making it adaptable to various contexts. Rather than replacing traditional methods, we view this approach as a powerful tool to augment and improve existing social vulnerability models. Furthermore, the machine learning approach refines the traditional method by more adeptly addressing the inherent uncertainties in social vulnerability datasets which results in more accurate predictions as shown by the model's evaluation metrics.

5.1. Limitations

The methodology used in this study has its limitations. We used purposive sampling, which could introduce bias in the selection of participating households, which may lead to over- or under-representation of certain neighbourhoods in the data. Furthermore, this study analysis relied on a relatively moderate sample size of 1334 households, which raises concerns, especially in the context of machine learning. In this study, we mitigate this risk, by employing techniques such as 10-fold cross-validation, hyperparameter tuning, and L2 regularization. Also, the SVR used in this study is well-known for performing effectively on limited data due to the principle of margin maximization[155]. Nevertheless, we acknowledge that the sample size may still impact the generalization of our findings, and future research with larger datasets is recommended to validate our results. Also, the five-point scaling system used to rate subcomponents and the narrative to the key indicator within the framework relied on the authors' subjective input. Future research can address this by adopting a more formal weighting system and incorporating one-hot encoding to quantify the influence of individual contributing factors. Finally, the analysis heavily relies on household-level data, which, while providing valuable insights, may not fully capture the influence of broader infrastructural factors on vulnerability. Future research could benefit from incorporating additional data sources, such as community-level infrastructure data or records of historical hazard events, to provide a more comprehensive understanding of vulnerability.

6. Conclusion

This article proposed a novel methodology for assessing social vulnerability profiles at a granular level through indicators using the Lekki Peninsula as a case study. In contrast to traditional geographically bound assessments, we employed a hybrid machine learning approach, combining non-spatial clustering using the EM algorithm, SVR predictions, and Permutation Importance for indicator ranking. This allowed us to identify seven distinct vulnerability clusters and rank the key indicators influencing them, including financial stability, social networks, insurance, residency length, health conditions, and language proficiency. Our finding revealed that vulnerability is on a moderate scale. Non-spatial clustering revealed hidden patterns, showing that wards contain heterogeneous mixes of vulnerability groups. This data-driven approach significantly reduces the bias inherent in traditional methods, leading to more accurate and equitable interventions.

In view of this finding discovered from this study, it is recommended that for policymakers to effectively reduce flood impacts should focus on preparedness by promoting the adoption of

insurance. Also, authorities should actively build and support community social networks. This can be achieved through targeted community-based programs that promote social interaction, information sharing, and mutual support this will strengthen these networks before a flood event and can significantly reduce community vulnerability and enhance recovery capabilities. In addition, authorities should prioritize the care of vulnerable populations during flood events (especially individuals with prolonged illness, children, the elderly, and the disabled). Efforts should be made to improve living conditions through inclusive housing policies and the development of social housing. This will alleviate pressure on informal settlements, which is a major contributor to social vulnerability. Long-term mitigation strategies through urban planning can promote equitable housing distribution within the community.

Future research should explore how this approach can be extrapolated to communities outside the location where primary data collection and investigate the long-term impacts of targeted interventions. This research paves the way for a more scientific and informed approach to flood vulnerability, aiming to improve the social conditions of vulnerable populations and build resilient communities.

CRediT authorship contribution statement

Adesola Akindejoye: Writing – original draft, Software, Data curation, Writing – review & editing, Visualization, Resources, Conceptualization. **Christophe Viavattene:** Writing – review & editing, Validation, Methodology, Conceptualization, Visualization, Supervision, Formal analysis. **Sally Priest:** Supervision, Writing – review & editing, Methodology. **David Windridge:** Writing – review & editing, Methodology, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijdrr.2025.105693.

Data availability

Data will be made available on request.

References

[1] I.O. Adelekan, A.P. Asiyanbi, Flood risk perception in flood-affected communities in Lagos, Nigeria, Nat. Hazards 80 (1) (2016) 445–469, https://doi.org/10.1007/s11069-015-1977-2.
[2] V. Hadipour, F. Vafaie, K. Deilami, Coastal flooding risk assessment using a GIS-based spatial multi-criteria decision analysis approach, Water (Switzerland) 12 (9) (2020), https://doi.org/10.3390/W12092379.
[3] E. Mavhura, T. Manyangadze, A comprehensive spatial analysis of social vulnerability to natural hazards in Zimbabwe: driving factors and policy implications, Int. J. Disaster Risk Reduct. 56 (2021) 102139. https://www.sciencedirect.com/science/article/pii/S2212420921001059.
[4] V. Hadipour, F. Vafaie, N. Kerle, An indicator-based approach to assess social vulnerability of coastal areas to sea-level rise and flooding: a case study of Bandar Abbas city, Iran, Ocean and Coastal Management Journal 188 (2020) 105077, https://doi.org/10.1016/j.ocecoaman.2019.105077.
[5] J.N. Obiefuna, A. Omojola, O. Adeaga, N. Uduma-Olugu, Groins or not: some environmental challenges to urban development on a Lagos coastal barrier island of Lekki Peninsula, J Construct Business Manag 1 (1) (2017) 14–28. http://journals.uct.ac.za/index.php/jcbm.

[6] D. Olayinka, H. Irivbogbe, Flood vulnerability mapping of Lagos Island and Eti-Osa local government areas using a multi-criteria decision making approach, Nigerian J Environ Sci Technol 1 (2) (2017) 244–255, https://doi.org/10.36263/nijest.2017.02.0011.
[7] A. Dewan, A.M. Dewan, Hazards, Risk, and Vulnerability, Springer, 2013.
[8] G. Di Baldassarre, D. Nohrstedt, J. Mård, S. Burchardt, C. Albin, S. Bondesson, K. Breinl, F.M. Deegan, D. Fuentes, M.G. Lopez, An integrative research framework to unravel the interplay of natural hazards and vulnerabilities, Earths Future 6 (3) (2018) 305–310.
[9] T.B. Paveglio, C.M. Edgeley, A.M. Stasiewicz, Assessing influences on social vulnerability to wildfire using surveys, spatial data and wildfire simulations, J. Environ. Manag. 213 (2018) 425–439.
[10] C. Armenakis, E.X. Du, S. Natesan, R.A. Persad, Y. Zhang, Flood risk assessment in urban areas based on spatial analytics and social factors, Geosciences 7 (4) (2017) 123.
[11] UNDRR, Understanding disaster risk. https://www.preventionweb.net/understanding-disaster-risk/terminology, 2017.
[12] D. Houston, A. Werritty, T. Ball, A. Black, Environmental vulnerability and resilience: social differentiation in short- and long-term flood impacts, Trans. Inst. Br. Geogr. 46 (1) (2021) 102–119, https://doi.org/10.1111/tran.12408.
[13] S.M. Tapsell, E.C. Penning-Rowsell, S.M. Tunstall, T.L. Wilson, Vulnerability to flooding: health and social dimensions, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 360 (1796) (2002) 1511–1525, https://doi.org/10.1098/rsta.2002.1013.
[14] N. Watson, E. Kashefi, W. Medd, G. Walker, S. Tapsell, C. Twigger-Ross, Institutional and social responses to flooding from a resilience perspective, in: Flood Risk Management: Research and Practice, CRC Press, 2008, p. 163.
[15] Dewan, Societal impacts and vulnerability to floods in Bangladesh and Nepal, Weather Clim. Extrem. 7 (2015) 36–42, https://doi.org/10.1016/j.wace.2014.11.001.
[16] S.F. Balica, N.G. Wright, F. van der Meulen, A flood vulnerability index for coastal cities and its use in assessing climate change impacts, Nat. Hazards 64 (1) (2012) 73–105, https://doi.org/10.1007/s11069-012-0234-1.
[17] C. Zarowsky, S. Haddad, V.-K. Nguyen, Beyond 'vulnerable groups': contexts and dynamics of vulnerability, Global Health Promotion 20 (1_suppl) (2013) 3–9.
[18] A. Deria, P. Ghannad, Y.C. Lee, Evaluating implications of flood vulnerability factors with respect to income levels for building long-term disaster resilience of low-income communities, Int. J. Disaster Risk Reduct. 48 (2020), https://doi.org/10.1016/j.ijdrr.2020.101608.
[19] I. Nur, K.K. Shrestha, An integrative perspective on community vulnerability to flooding in cities of developing countries, Procedia Eng. 198 (2017) 958–967.
[20] P. Orr, C. Twigger-Ross, K. Brooks, R. Sadauskis, 'Pieces of kit'are not enough: the role of infrastructure in community resilience, E3S Web of Conferences 7 (2016) 8009.
[21] G. Terti, I. Ruin, S. Anquetin, J.J. Gourley, Dynamic vulnerability factors for impact-based flash flood prediction, Nat. Hazards 79 (2015) 1481–1497.
[22] N. Guerrero, M. Contreras, A. Chamorro, C. Martínez, T. Echaveguren, Social vulnerability in Chile: challenges for multi-scale analysis and disaster risk reduction, Nat. Hazards 117 (3) (2023) 3067–3102.
[23] S. Percival, R. Teeuw, A methodology for urban micro-scale coastal flood vulnerability and risk assessment and mapping, Nat. Hazards 97 (1) (2019) 355–377.
[24] J.H. Masterson, W.G. Peacock, S.S. Van Zandt, H. Grover, L.F. Schwarz, J.T. Cooper, J.H. Masterson, W.G. Peacock, S.S. Van Zandt, H. Grover, Assessing social vulnerability, Planning for Community Resilience: A Handbook for Reducing Vulnerability to Disasters (2014) 97–112.
[25] T. Moroke, C. Schoeman, I. Schoeman, Neighbourhood sustainability assessment model for developing countries: a comprehensive approach to urban quality of life, Int. J. Sustain. Dev. Plann. 15 (1) (2020) 107–123.
[26] P. O'Keefe, K. Westgate, B. Wisner, P. O'keefe, K. Westgate, B. Wisner, No Title, vol. 260, 1976, pp. 566–567.
[27] B.E.O. Monte, J.A. Goldenfum, G.P. Michel, J.R. de A. Cavalcanti, J.R. de Albuquerque Cavalcanti, Terminology of natural hazards and disasters: a review and the case of Brazil, Int. J. Disaster Risk Reduct. 52 (2021) 101970. https://www.sciencedirect.com/science/article/pii/S2212420920314722.
[28] S.L. Cutter, B.J. Boruff, W.L. Shirley, Social vulnerability to environmental hazards, Soc. Sci. Q. 84 (2) (2003) 242–261.
[29] P.M. Kelly, W.N. Adger, Theory and practice in assessing vulnerability to climate change andFacilitating adaptation, Clim. Change 47 (4) (2000) 325–352.
[30] J.X. Kasperson, R.E. Kasperson, The social contours of risk: publics, Risk Communication and the Social Amplification of Risk 1 (2005).
[31] O.D. Cardona, M.K. Van Aalst, J. Birkmann, M. Fordham, G. Mc Gregor, P. Rosa, R.S. Pulwarty, E.L.F. Schipper, B.T. Sinh, H. Décamps, Determinants of risk: exposure and vulnerability, in: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2012, pp. 65–108.
[32] S.L. Cutter, L. Barnes, M. Berry, C. Burton, E. Evans, E. Tate, J. Webb, A place-based model for understanding community resilience to natural disasters, Glob. Environ. Change 18 (4) (2008) 598–606, https://doi.org/10.1016/j.gloenvcha.2008.07.013.
[33] UNDRR, UN (2016). Report of the open-ended Intergovernmental Expert Working Group on Indicators and Terminology Relating to Disaster Risk Reduction, Prevention web, 2016.
[34] W.N. Adger, Vulnerability, Glob. Environ. Change 16 (3) (2006) 268–281, https://doi.org/10.1016/j.gloenvcha.2006.02.006.
[35] D. Alexander, D.E. Alexander, Confronting Catastrophe: New Perspectives on Natural Disasters, Oxford University Press, USA, 2000.
[36] J. Birkmann, O.D. Cardona, M.L. Carreño, A.H. Barbat, M. Pelling, S. Schneiderbauer, S. Kienberger, M. Keiler, D. Alexander, P. Zeil, Framing vulnerability, risk and societal responses: the MOVE framework, Nat. Hazards 67 (2013) 193–211.
[37] J. Birkmann, A. Jamshed, J.M. McMillan, D. Feldmeyer, E. Totin, W. Solecki, Z.Z. Ibrahim, D. Roberts, R.B. Kerr, H.-O.O. Poertner, M. Pelling, R. Djalante, M. Garschagen, W. Leal Filho, D. Guha-Sapir, A. Alegría, Understanding human vulnerability to climate change: a global perspective on index validation for adaptation planning, Sci. Total Environ. 803 (2022) 150065, https://doi.org/10.1016/j.scitotenv.2021.150065.
[38] F. Fatemi, A. Ardalan, B. Aguirre, N. Mansouri, I. Mohammadfam, Social vulnerability indicators in disasters: findings from a systematic review, Int. J. Disaster Risk Reduct. 22 (2017) 219–227, https://doi.org/10.1016/j.ijdrr.2016.09.006.
[39] G.C. Gallopín, Linkages between vulnerability, resilience, and adaptive capacity, Glob. Environ. Change 16 (3) (2006) 293–303.
[40] J. Birkmann, Data, indicators and criteria for measuring vulnerability: theoretical bases and requirements, Measuring Vulnerability to Natural Hazards: Towards Disaster Resilient Societies (2013) 80–106.
[41] W.N. Adger, M. Agnew, New Indicators of Vulnerability and Adaptive Capacity, vol 122, Tyndall Centre for Climate Change Research Norwich, 2004.
[42] M.M. de Brito, M. Evers, B. Höllermann, Prioritization of flood vulnerability, coping capacity and exposure indicators through the Delphi technique: a case study in Taquari-Antas basin, Brazil, Int. J. Disaster Risk Reduct. 24 (2017) 119–128.
[43] A. Jamshed, J. Birkmann, I. Ahmad Rana, D. Feldmeyer, The effect of spatial proximity to cities on rural vulnerability against flooding: an indicator based approach, Ecol. Indic. 118 (2020), https://doi.org/10.1016/j.ecolind.2020.106704.
[44] J. Birkmann, Indicators and criteria for measuring vulnerability: theoretical bases and requirements, Measuring Vulnerability to Natural Hazards: Towards Disaster Resilient Societies (2006) 55–77.
[45] G.C. Gallopin, Indicators and their use: information for decision-making, Scope-Scientific Committee on Problems of the Environment International Council of Scientific Unions 58 (1997) 13–27.
[46] S. Rufat, E. Tate, C.G. Burton, A.S. Maroof, Social vulnerability to floods: review of case studies and implications for measurement, Int. J. Disaster Risk Reduct. 14 (2015) 470–486. https://www.sciencedirect.com/science/article/pii/S2212420915300935.
[47] J. Shaji, Evaluating social vulnerability of people inhabiting a tropical coast in Kerala, south west coast of India, Int. J. Disaster Risk Reduct. 56 (2021) 102130. https://www.sciencedirect.com/science/article/pii/S2212420921000960.
[48] A. El-Zein, T. Ahmed, F. Tonmoy, Geophysical and social vulnerability to floods at municipal scale under climate change: the case of an inner-city suburb of Sydney, Ecol. Indic. 121 (2021) 106988. https://www.sciencedirect.com/science/article/pii/S1470160X20309274.
[49] S.Q. Ma, J.C. Weiss, Strong State or vulnerable homeland: how Chinese State media sought to combat democratic diffusion during the 2019 Hong Kong protests, J. Contemp. China 32 (139) (2023) 106–122, https://doi.org/10.1080/10670564.2022.2052442.
[50] Analytical Framework for Assessing the Social-Ecological System Trajectory Considering the Resilience-Vulnerability Dynamic Interaction in the Context of Disasters, 59 International Journal of Disaster Risk Reduction (2021) 102232. https://www.sciencedirect.com/science/article/pii/S2212420921001989.
[51] J. Birkmann, Risk and vulnerability indicators at different scales: applicability, usefulness and policy implications, Environ. Hazards 7 (1) (2007) 20–31.

[52] M.C. de Ruiter, A.F. Van Loon, The challenges of dynamic vulnerability and how to assess it, iScience 25 (8) (2022).
[53] I. Jesna, C. Kurian, S.M.B.K.P. Sudheer, A New Framework Integrating Flood Inundation Modeling and Multicriteria decision-making for Enhanced Flood Risk Mapping, 2024.
[54] M. Englund, M.V. Passos, K. André, Å.G. Swartling, L. Segnestam, K. Barquet, Constructing a Social Vulnerability Index for Flooding : Insights from a Municipality in Sweden, 2022, 2009.
[55] S. Rufat, E. Tate, C.T. Emrich, F. Antolini, How valid are social vulnerability models? Ann. Assoc. Am. Geogr. 109 (4) (2019) 1131–1153, https://doi.org/10.1080/24694452.2018.1535887.
[56] R. de E.A. Revista de Estudios Andaluces, S. Gayen, L.I. Vallejo-Villalta, S. Mafizul Haque, Assessment of social vulnerability in Malaga Province, Spain: a comparison of indicator standardization techniques, Revista de Estudios Andaluces 41 (2021), https://doi.org/10.12795/rea.2021.i41, 0–1.
[57] H.B. Abbas, J.K. Routray, H. Bashier Abbas, J. K Routray, Vulnerability to flood-induced public health risks in Sudan, Disaster Prev. Manag. 23 (4) (2014) 395–419, https://doi.org/10.1108/DPM-07-2013-0112.
[58] A. Jamshed, I.A. Rana, U.M. Mirza, J. Birkmann, Assessing relationship between vulnerability and capacity: an empirical study on rural flooding in Pakistan, Int. J. Disaster Risk Reduct. 36 (2019) 101109.
[59] A.Y. Karunarathne, G. Lee, Developing a multi-facet social vulnerability measure for flood disasters at the micro-level assessment, Int. J. Disaster Risk Reduct. 49 (2020) 101679. https://www.sciencedirect.com/science/article/pii/S2212420920300613.
[60] R.K. Jha, H. Gundimeda, An integrated assessment of vulnerability to floods using composite index – a district level analysis for Bihar, India, Int. J. Disaster Risk Reduct. 35 (2019) 101074. https://www.sciencedirect.com/science/article/pii/S2212420918312627.
[61] S.A. Enderami, E. Sutley, Social vulnerability score: a scalable index for representing social vulnerability in virtual community resilience testbeds, Nat. Hazards (2024) 1–24.
[62] E.A. Goto, A.M. Suarez, H. Ye, Spatial analysis of social vulnerability in São Paulo city, Brazil, Appl. Geogr. 140 (2022) 102653. https://www.sciencedirect.com/science/article/pii/S0143622822000248.
[63] T. Bouramtane, I. Kacimi, K. Bouramtane, M. Aziz, S. Abraham, K. Omari, V. Valles, M. Leblanc, N. Kassou, O. El Beqqali, T. Bahaj, M. Morarech, S. Yameogo, L. Barbiero, Multivariate Analysis and Machine Learning Approach for Mapping the Variability and Vulnerability of Urban Moroccoflooding : the Case of Tangier City ,, 2021.
[64] M.C. Schmidtlein, R.C. Deutsch, W.W. Piegorsch, S.L. Cutter, A sensitivity analysis of the social vulnerability index, Risk Anal. 28 (4) (2008) 1099–1114, https://doi.org/10.1111/j.1539-6924.2008.01072.x.
[65] C.-H.H. Sung, S.-C.C. Liaw, A GIS-based approach for assessing social vulnerability to flood and debris flow hazards, Int. J. Disaster Risk Reduct. 46 (2020) 101531. https://www.sciencedirect.com/science/article/pii/S2212420919313792.
[66] A.N. Andrewin, J.M. Rodriguez-Llanes, D. Guha-Sapir, Determinants of the lethality of climate-related disasters in the Caribbean Community (CARICOM): a cross-country analysis, Sci. Rep. 5 (2015), https://doi.org/10.1038/srep11972.
[67] M. Madajewicz, Who is vulnerable and who is resilient to coastal flooding? Lessons from Hurricane Sandy in New York City, Clim. Change 163 (4) (2020) 2029–2053, https://doi.org/10.1007/s10584-020-02896-y.
[68] E.D. Mcconnell, E. Díaz McConnell, Rented, crowded, and unaffordable? Social vulnerabilities and the accumulation of precarious housing conditions in Los Angeles, Housing Policy Debate 27 (1) (2017) 60–79, https://doi.org/10.1080/10511482.2016.1164738.
[69] L. Chakraborty, H. Rus, D. Henstra, J. Thistlethwaite, A. Minano, D. Scott, Exploring spatial heterogeneity and environmental injustices in exposure to flood hazards using geographically weighted regression, Environ. Res. 210 (2022) 112982. https://www.sciencedirect.com/science/article/pii/S0013935122003097.
[70] S. Hinojos, L. McPhillips, P. Stempel, C. Grady, Social and environmental vulnerability to flooding: investigating cross-scale hypotheses, Appl. Geogr. 157 (2023) 103017.
[71] D. Morrison, L. Beevers, G. Wright, M.D. Stewart, The impact of data spatial resolution on flood vulnerability assessment. https://doi.org/10.1080/17477891.2021.1912694, 2022.
[72] E. Mavhura, Applying a systems-thinking approach to community resilience analysis using rural livelihoods: the case of Muzarabani district, Zimbabwe, Int. J. Disaster Risk Reduct. 25 (2017) 248–258.
[73] E. Mavhura, Analysing drivers of vulnerability to flooding: a systems approach, S. Afr. Geogr. J. 101 (1) (2019) 72–90, https://doi.org/10.1080/03736245.2018.1541020.
[74] S. Uddin, S. Ong, H. Lu, Machine learning in project analytics: a data-driven framework and case study, Sci. Rep. 12 (1) (2022) 15252.
[75] T. Gudiyangada Nachappa, S. Tavakkoli Piralilou, K. Gholamnia, O. Ghorbanzadeh, O. Rahmati, T. Blaschke, Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory, J. Hydrol. 590 (July) (2020) 125275, https://doi.org/10.1016/j.jhydrol.2020.125275.
[76] A.R.M.T. Islam, S. Talukdar, S. Mahato, S. Kundu, K.U. Eibek, Q.B. Pham, A. Kuriqi, N.T.T. Linh, Flood susceptibility modelling using advanced ensemble machine learning models, Geosci. Front. 12 (3) (2021) 101075.
[77] P. Deroliya, M. Ghosh, M.P. Mohanty, S. Ghosh, K.H.V.D. Rao, S. Karmakar, A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions, Sci. Total Environ. 851 (2022) 158002.
[78] H. Mojaddadi, B. Pradhan, H. Nampak, N. Ahmad, A. H. bin Ghazali, Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS, Geomat. Nat. Hazards Risk 8 (2) (2017) 1080–1102.
[79] N. Razali, S. Ismail, A. Mustapha, Machine learning approach for flood risks prediction, IAES Int. J. Artif. Intell. 9 (1) (2020) 73.
[80] S. Lee, M. Panahi, H.R. Pourghasemi, H. Shahabi, M. Alizadeh, A. Shirzadi, K. Khosravi, A.M. Melesse, M. Yekrangnia, F. Rezaie, Sevucas: a novel gis-based machine learning software for seismic vulnerability assessment, Appl. Sci. 9 (17) (2019) 3495.
[81] H. Hou, S. Yu, H. Wang, Y. Huang, H. Wu, Y. Xu, X. Li, H. Geng, Risk assessment and its visualization of power tower under typhoon disaster based on machine learning algorithms, Energies 12 (2) (2019) 205.
[82] A.A. Adepelumi, B.D. Ako, T.R. Ajayi, O. Afolabi, E.J. Omotoso, Delineation of saltwater intrusion into the freshwater aquifer of Lekki Peninsula, Lagos, Nigeria, Environ. Geol. 56 (5) (1999) 927–933, https://doi.org/10.1007/s00254-008-1194-3.
[83] I. Chukwuemeka Osmond, D. Olufemi Daniel, A. Caleb Abiodun, O. Mayowa Fadeke, Journal of sustainable development studies comparative study of rental values of two gated estates in Lekki peninsula Lagos 5 (2) (2014) 218–235.
[84] Lagos State Government, Lekki Peninsula Development Scheme Report, 1980, p. 111.
[85] S.D. Brody, W.E. Highfield, H.-C. Ryu, L. Spanel-Weber, Examining the relationship between wetland alteration and watershed flooding in Texas and Florida, Nat. Hazards 40 (2007) 413–428.
[86] J. Obiefuna, O. Adeaga, A. Omojola, A. Atagbaza, C. Okolie, Flood risks to urban development on a coastal barrier landscape of Lekki Peninsula in Lagos, Nigeria, Sci. Afr. 12 (2021) e00787, https://doi.org/10.1016/j.sciaf.2021.e00787.
[87] M.A.A. Abegunde, Shoreline erosion and land use management on the active sand barrier beaches around Lagos: a new focus in environmental management in Sada, PO & FO Odemerho, Environmental Issues and Management in Nigerian Development. Evans Brothers, Ibadan (1988) 231–238.
[88] R. Almar, E. Kestenare, J. Reyns, J. Jouanno, E.J. Anthony, R. Laibi, M. Hemer, Y. Du Penhoat, R. Ranasinghe, Response of the Bight of Benin (Gulf of Guinea, West Africa) coastline to anthropogenic and natural forcing, Part1: wave climate variability and impacts on the longshore sediment transport, Cont. Shelf Res. 110 (2015) 48–59.
[89] E.J. Anthony, R. Almar, M. Besset, J. Reyns, R. Laibi, R. Ranasinghe, G.A. Ondoa, M. Vacchi, Response of the Bight of Benin (Gulf of Guinea, West Africa) coastline to anthropogenic and natural forcing, part 2: sources and patterns of sediment supply, sediment cells, and recent shoreline change, Cont. Shelf Res. 173 (2019) 93–103.
[90] A.J. Osanyintuyi, Y.H. Wang, N.A.H. Mokhtar, Nearly five decades of changing shoreline mobility along the densely developed Lagos barrier-lagoon coast of Nigeria: a remote sensing approach, J. Afr. Earth Sci. 194 (2022) 104628, https://doi.org/10.1016/j.jafrearsci.2022.104628.

[91] E. Olaniyan, E.A. Afiesimama, On marine winds, waves and swells over West African coast for effective coastal management-a case study of Victoria Island beach in Nigeria, OCEANS'02 MTS/IEEE 1 (2002) 561–568.
[92] L.F. Awosika, C.O. Dublin-Green, R. Folorunsho, E.A. Adekoya, M.A. Adekanmbi, J. Jim-Saiki, Study of Main Drainage Channels of Victoria and Ikoyi Islands in Lagos Nigeria and their Response to Tidal and Sea Level Changes, CSI-UNESCO Special Report, 2000.
[93] P.C. Nwilo, Managing the impacts of storm surges on Victoria Island, Lagos, Nigeria, IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences 239 (239) (1997) 325–330.
[94] M. Simire, The ocean can consume us, yet we can't help ourselves. https://www.Environewsnigeria.Com/the-Ocean-Can-Consume-Us-yet-We-Cant-Help-Ourselves/, 2014.
[95] I.O. Adelekan, Vulnerability of poor urban coastal communities to flooding in Lagos, Nigeria, Environ. Urbanization 22 (2) (2010) 433–450, https://doi.org/10.1177/0956247810380141.
[96] O. Ekanade, A. Ayanlade, O.O.I. Orimoogunje, Climate change impacts on coastal urban settlements in Nigeria, Interdiscipl. Environ. Rev. 12 (1) (2011) 48–62.
[97] A.E. Ihenyen, Recent sedimentology and ocean dynamics of the Western Nigerian continental shelf and coastline, J. Afr. Earth Sci. 36 (3) (2003) 233–244.
[98] U. Ajaero, Eko Resilience (Re) Designing Residential Communities that Are Resilient for the Urban Poor in Lagos, the Coastal City, University of Maryland, College Park, 2024.
[99] A. Soneye, An overview of humanitarian relief supply chains for victims of perennial flood disasters in Lagos, Nigeria (2010-2012), J. Humanit. Logist. Supply Chain Manag. 4 (2) (2014) 179–197, https://doi.org/10.1108/JHLSCM-01-2014-0004.
[100] I.O. Ogbonna, O. Umeh, Flood-risk insurance in flood-prone areas of Nigerian cities: a case study of Agiliti, Ketu, Lagos, African J Housing Sustain Development 4 (2) (2023).
[101] U.C. Nkwunonwo, M. Whitworth, B. Baily, Urban flood risk management in the Lagos region of Nigeria, Nat. Hazards Earth Syst. Sci. 16 (2016) 349–369.
[102] Y. Wang, Near-optimal quantum kernel principal component analysis, Quantum Sci. Technol. 10 (1) (2024) 15034.
[103] L. Plonsky, Advancing quantitative methods in second language research, in: Advancing Quantitative Methods in Second Language Research, 2015, https://doi.org/10.4324/9781315870908 (Issue August).
[104] A.A. Neath, J.E. Cavanaugh, The Bayesian information criterion: background, derivation, and applications, Wiley Interdisciplinary Reviews: Comput. Stat. 4 (2) (2012) 199–203.
[105] R.H. Jones, Bayesian information criterion for longitudinal and clustered data, Stat. Med. 30 (25) (2011) 3050–3056.
[106] J. Lorah, A. Womack, Value of sample size for computation of the Bayesian information criterion (BIC) in multilevel modeling, Behav. Res. Methods 51 (2019) 440–450.
[107] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Stat. Soc. B 39 (1) (1977) 1–22.
[108] G. Kwon, D. Basukala, S. Lee, K.H. Lee, M. Kang, Brain image segmentation using a combination of expectation-maximization algorithm and watershed transform, Int. J. Imag. Syst. Technol. 26 (3) (2016) 225–232.
[109] H. Deng, J. Han, Probabilistic models for clustering, in: Data Clustering, Chapman and Hall/CRC, 2018, pp. 61–86.
[110] C.L. Byrne, REPAIRING THE EM ALGORITHM: PART TWO-THE MISSING-DATA MODEL, 2024.
[111] D. Johnson, P. Blackett, A.E.F. Allison, A.M. Broadbent, Measuring social vulnerability to climate change at the coast: embracing complexity and context for more accurate and equitable analysis, Water 15 (19) (2023) 3408.
[112] A. Acconcia, M. Carannante, M. Misuraca, G. Scepi, Measuring vulnerability to poverty with latent transition analysis, Soc. Indic. Res. 151 (2020) 1–31.
[113] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, John Wiley & Sons, 2008.
[114] M. Awad, R. Khanna, Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, Springer nature, 2015.
[115] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (2) (1998) 121–167.
[116] Y. Liu, X. Zhang, W. Liu, Y. Lin, F. Su, J. Cui, B. Wei, Seismic vulnerability and risk assessment at the urban scale using support vector machine and GIScience technology : a case study of the Lixia District in Jinan City , China, Geomat. Nat. Hazards Risk 14 (1) (2023), https://doi.org/10.1080/19475705.2023.2173663.
[117] A.M. Youssef, H. Reza, P. Ali, M.M. Saleh, Flood vulnerability mapping and urban sprawl suitability using FR , LR , and SVM models, Environ. Sci. Pollut. Control Ser. (2023) 16081–16105, https://doi.org/10.1007/s11356-022-23140-3.
[118] R. Hamad, An assessment of artificial neural networks, support vector machines and decision trees for land cover classification using sentinel-2A data, Sciences 8 (6) (2020) 459–464.
[119] Y. Wang, Y. Fan, P. Bhatt, C. Davatzikos, High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables, Neuroimage 50 (4) (2010) 1519–1535.
[120] W. Dhhan, H. Midi, T. Alameer, Robust support vector regression model in the presence of outliers and leverage points, Mod. Appl. Sci. 11 (8) (2017) 1–92.
[121] J. Valyon, G. Horváth, Extended least squares LS–SVM, Training 500 (2007) 1.
[122] B. Frénay, M. Verleysen, Parameter-insensitive kernel in extreme learning for non-linear support vector regression, Neurocomputing 74 (16) (2011) 2526–2531.
[123] A.M. Carrington, P.W. Fieguth, H.H. Chen, A new mercer sigmoid kernel for clinical data classification, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, pp. 6397–6401.
[124] Y. Hur, S. Lim, Customer churning prediction using support vector machines in online auto insurance service, International Symposium on Neural Networks (2005) 928–933.
[125] Y.-J. Oyang, S.-C. Hwang, Y.-Y. Ou, C.-Y. Chen, Z.-W. Chen, Data classification with radial basis function networks based on a novel kernel density estimation algorithm, IEEE Trans. Neural Network. 16 (1) (2005) 225–236.
[126] H.-T. Lin, C.-J. Lin, A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods, Neural Comput. 3 (1–32) (2003) 16.
[127] S. Ghosh, A. Dasgupta, A. Swetapadma, A study on support vector machine based linear and non-linear pattern classification, in: 2019 International Conference on Intelligent Sustainable Systems (ICISS), 2019, pp. 24–28.
[128] A.O. Kuyoro, S. Alimi, O. Awodele, Comparative analysis of the performance of various support vector machine kernels, in: 2022 5th Information Technology for Education and Development (ITED), 2022, pp. 1–7, https://doi.org/10.1109/ITED56637.2022.10051564.
[129] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[130] A.J. Smola, B.S.C.H. Scholkopf, A Tutorial on Support Vector Regression, 2004, pp. 199–222.
[131] I.S. Mangkunegara, P. Purwono, Analysis of DNA sequence classification using SVM model with hyperparameter tuning grid search CV, in: 2022 IEEE International Conference on Cybernetics and Computational Intelligence (Cyberneticscom), 2022, pp. 427–432.
[132] J. Wu, H. Liu, G. Wei, T. Song, C. Zhang, H. Zhou, Flash Flood Forecasting Using Support Vector Regression Model in a Small Mountainous Catchment, 2019.
[133] D. Tien Bui, B. Pradhan, O. Lofman, I. Revhaug, Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naive Bayes Models, Math. Probl Eng. 2012 (1) (2012) 974638.
[134] M. Loki, A. Mindila, W. Cheruiyot, Hyper-parameter optimization of kernel functions on multi-class text categorization: a comparative evaluation, Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 15 (1) (2025) e1572.
[135] F. Cheng, G. Ou, M. Wang, C. Liu, Remote sensing estimation of forest carbon stock based on machine learning algorithms, Forests 15 (4) (2024) 681.
[136] G.-H. Kwak, N.-W. Park, Impact of texture information on crop classification with machine learning and UAV images, Appl. Sci. 9 (4) (2019) 643.
[137] D. Liu, C.Q. Wang, Y. Ji, Q. Fu, M. Li, S. Ali, T.X. Li, S. Cui, Measurement and analysis of regional flood disaster resilience based on a support vector regression model refined by the selfish herd optimizer with elite opposition-based learning, J. Environ. Manag. 300 (2021), https://doi.org/10.1016/j.jenvman.2021.113764.
[138] F. Dottori, R. Figueiredo, M.L.V. Martina, D. Molinari, A.R. Scorzini, INSYDE: a synthetic, probabilistic flood damage model based on explicit cost analysis, Nat. Hazards Earth Syst. Sci. 16 (12) (2016) 2577–2591.

[139] S. Sulong, N.S. Romali, The role of socio-economic and property variables in the establishment of flood depth-damage curve for the data-scarce area in Malaysia, Urban Water J. 19 (8) (2022) 859–878, https://doi.org/10.1080/1573062X.2022.2099292.

[140] S. Win, W.W. Zin, A. Kawasaki, Z.M.L.T. San, Establishment of flood damage function models: a case study in the Bago River Basin, Myanmar, Int. J. Disaster Risk Reduct. 28 (2018) 688–700.

[141] C. Molnar, Interpretable Machine Learning, Lulu. com, 2020.

[142] C. Molnar, A guide for making black box models explainable. https://Christophm.Github.Io/Interpretable-Ml-Book, 2018, 2(3), 10.

[143] B.E. Flanagan, E.W. Gregory, E.J. Hallisey, J.L. Heitgerd, B. Lewis, A social vulnerability index for disaster management, J. Homel. Secur. Emerg. Manag. 8 (1) (2011) 0000102202154773551792, https://doi.org/10.2202/1547-7355.1792.

[144] S. Lee, K. Vink, No title, Water Policy 17 (6) (2015) 1045–1061, https://doi.org/10.2166/wp.2015.089.

[145] M.C. Montgomery, J. Chakraborty, Social vulnerability to coastal and inland flood hazards: a comparison of GIS-based spatial interpolation methods, Int. J. Appl. Geospatial Res. (IJAGR) 4 (3) (2013) 58–79, https://doi.org/10.4018/jagr.2013070104.

[146] C. Wang, B. Yarnal, The vulnerability of the elderly to hurricane hazards in Sarasota, Florida, Nat. Hazards 63 (2) (2012) 349–373, https://doi.org/10.1007/s11069-012-0151-3.

[147] M. M. N. de Andrade, C.F. Szlafsztein, Vulnerability assessment including tangible and intangible components in the index composition: an Amazon case study of flooding and flash flooding, Sci. Total Environ. 630 (2018) 903–912, https://doi.org/10.1016/j.scitotenv.2018.02.271.

[148] G. Delrieu, J. Douvinet, E. Gaume, E. Gruntfest, J. Naulin, O. Payrastre, O. Vannier, I. Ruin, C. Line Lutoff, B. Boudevillain, J. Creutin, S. Anquetin, M. Bertran Rojo, L. Boissier, L. Bonnifait, M. Borga, L. Colbeau-justin, L. Creton-cazanave, American Meteorological Society Social and Hydrological Responses to Extreme Precipitations:: An Interdisciplinary Strategy for Postflood Investigation Social and Hydrological Responses to Extreme Precipitations: An Interdisciplinary Strategy for Postfloo. And Society 6 (1) (2014) 135–153, https://doi.org/10.2307/24907319.

[149] A. Fekete, Validation of a social vulnerability index in context to river-floods in Germany, Nat. Hazards Earth Syst. Sci. 9 (2) (2009) 393–403, https://doi.org/10.5194/nhess-9-393-2009.

[150] P.M. Kelly, W.N. Adger, Theory and Practice in Assessing Vulnerability to Climate Change and Facilitating Adaptation. Climatic Change 47 (325 – 352–47) (2000) 325–352.

[151] E. Tate, M.A. Rahman, C.T. Emrich, C.C. Sampson, Flood exposure and social vulnerability in the United States, Nat. Hazards 106 (1) (2021) 435–457, https://doi.org/10.1007/s11069-020-04470-2.

[152] K. Garbutt, C. Ellul, T. Fujiyama, Mapping social vulnerability to flood hazard in Norfolk, England, Environ. Hazards 14 (2) (2015) 156–186, https://doi.org/10.1080/17477891.2015.1028018.

[153] Y. Han, X. Ye, C. Zhu, The unequal impact of disasters: assessing the interplay between social vulnerability, public assistance, flood insurance, and migration in the US, Urban Informatics 3 (1) (2024) 1–12.

[154] P. Ireland, F. Thomalla, The role of collective action in enhancing communities' adaptive capacity to environmental risk: an exploration of two case studies from Asia, PLoS Currents 3 (2011).

[155] A.P. Santosa, M. Akrom, Quantum support vector regression for predicting corrosion inhibition of drugs, J Multiscale Mater Informatics 1 (2) (2024) 30–34.

[156] L. Chakraborty, H. Rus, D. Henstra, J. Thistlethwaite, D. Scott, A place-based socioeconomic status index: measuring social vulnerability to flood hazards in the context of environmental justice, Int. J. Disaster Risk Reduct. 43 (2020) 101394, https://doi.org/10.1016/j.ijdrr.2019.101394.

[157] O. Lawal, S.B. Arokoyu, Modelling social vulnerability in sub-saharan West Africa using a geographical information system, Jamba -J Disaster Risk Studies 7 (1) (2015), https://doi.org/10.4102/jamba.v7i1.155.

[158] B.A. Harahap, Relevant indicators of social vulnerability during response phase context of flooding inTebingTinggi City, North Sumatra Indonesia, IOP Conf. Ser. Earth Environ. Sci. 921 (1) (2021), https://doi.org/10.1088/1755-1315/921/1/012015.

[159] G. Oulahen, L. Mortsch, K. Tang, D. Harford, Unequal vulnerability to flood hazards: "ground Truthing" a social vulnerability index of five municipalities in Metro Vancouver, Canada, Ann. Assoc. Am. Geogr. 105 (3) (2015) 473–495, https://doi.org/10.1080/00045608.2015.1012634.

[160] B.B. Cutts, A.J. Greenlee, N.K. Prochaska, C.V. Chantrill, A.B. Contractor, J.M. Wilhoit, N. Abts, K. Hornik, Is a clean river fun for all? Recognizing social vulnerability in watershed planning, PLoS One 13 (5) (2018) e0196416, https://doi.org/10.1371/journal.pone.0196416.

[161] A. Styles, C. Viavattene, S. Priest, Resilience-increasing strategies for coasts-toolkit library of coastal vulnerability indicators guidance document*. www.risckit.eu, 2014.

[162] HR Wallingford, Flood risks to people (Phase 2) FD2321/TR1. *Flood risks to people phase 2: FD2321 technical report 1*, 117. http://randd.defra.gov.uk/Document.aspx?Document=FD2317_1060_TRP.pdf, 2006.

[163] S. Priest, T. Wilson, S. Tapsell, E. Penning-Rowsell, C. Viavattene, A. Fernández-Bilbao, Building a Model to Estimate Risk to Life for European Flood Events, Milestone Report T1-07-10 for Floodsite, 2007. November.

[164] U.P. Guragain, P. Doneys, Social, economic, environmental, and physical vulnerability assessment: an index-based gender analysis of flood prone areas of Koshi River Basin in Nepal, Sustainability Article 14 (16) (2022) 10423, https://doi.org/10.3390/su141610423.

[165] C.T. Nguyen, B. Van Nguyen, Application of flood vulnerability index in flood vulnerability assessment: a case study in Mai Hoa Commune, Tuyen Hoa District, Quang Binh Province, Sustain Water Resou Manag 5 (4) (2019) 1917–1927, https://doi.org/10.1007/s40899-019-00337-y.

[166] S. Bera, A. Das, T. Mazumder, A multi-objective framework for multidimensional vulnerability assessment – case of a coastal district of West Bengal, India, J. Environ. Manag. 249 (2019) 109411, https://doi.org/10.1016/j.jenvman.2019.109411.

[167] C. Finch, C.T. Emrich, S.L. Cutter, Disaster disparities and differential recovery in New Orleans, Popul. Environ. 31 (4) (2010) 179–202, https://doi.org/10.1007/s11111-009-0099-8.

[168] R.B. Munyai, N.S. Nethengwe, A. Musyoki, An assessment of flood vulnerability and adaptation: a case study of Hamutsha-Muungamunwe village, Makhado municipality, Jàmbá: J Disaster Risk Studies 11 (2) (2019) 1–8.