Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 1, January 2026, pp: 395-408 **e-ISSN: 2656-8632**

RESEARCH ARTICLE

Dengue Risk Stratification in Semarang City Using a Gaussian Mixture Model Based on Multi-Dimensional Urban Indicators

Nabila Izzatil Ismah[1] and Amiq Fahmi[2]

[1] Computer Science Program, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
[2] Information Systems Program, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

Corresponding author: Amiq Fahmi (e-mail: amiq.fahmi@dsn.dinus.ac.id), **Author(s) Email**: Nabila Izzatil Ismah (e-mail: 111202213972@mhs.dinus.ac.id)

Abstract Dengue fever remains a pressing public health challenge in major Indonesian cities, including Semarang. The complex interplay of heterogeneous demographic structures and built-environment characteristics generates spatially uneven transmission risks, while conventional risk-mapping approaches often fail to capture the probabilistic nature of these risks at fine-scale administrative levels, limiting their utility for targeted interventions. This study aims to develop a robust, replicable framework for dengue risk stratification that more accurately identifies localized high-risk areas and supports evidence-based public health decision-making. The research introduces a probabilistic clustering approach using Gaussian Mixture Models (GMM) to move beyond rigid partitioning methods, while simultaneously integrating multi-year incidence data (2021–2024) with eighteen multidimensional urban indicators across 177 sub-districts (kelurahan). This combined contribution advances methodological rigor by accommodating overlapping data distributions and probabilistic cluster memberships, and provides a nuanced, evidence-driven tool for stratifying dengue risk and guiding hyper-local interventions. Several GMM configurations were evaluated using the Bayesian Information Criterion (BIC) to determine the optimal number of clusters. The BIC value declined markedly when the number of clusters increased from two to three, indicating a substantial improvement in model fit. Further increases yielded only marginal gains, and the lowest BIC was achieved at three clusters, representing the most parsimonious and effective solution. Internal validation confirmed that the cluster structure robustly captured epidemiological variance despite the inherent heterogeneity of urban spatial data. Cluster 2 emerged as a critical high-risk epicenter, geographically limited yet characterized by consistently elevated incidence, pronounced temporal variability, and extreme values. The proposed GMM-based framework demonstrates that dengue risk in Semarang is concentrated within localized foci of heightened vulnerability rather than uniformly distributed. Ultimately, the methodology is replicable in other complex tropical urban environments, thereby strengthening both academic rigor and practical public health decision-making.

Keywords Dengue risk stratification; Gaussian mixture model; Bayesian information criterion; Urban built-environment; Spatial epidemiology.

I. Introduction

Dengue fever remains a critical public health threat in Semarang, one of Indonesia's major metropolitan areas, where rapid urbanization has intensified the complex interplay between demographic factors and the built environment, leading to spatially uneven transmission risks [1]. Semarang City has consistently reported a substantial number of dengue cases over recent years, with marked variability across sub-districts, reflecting persistent spatial heterogeneity in transmission risk. In this diverse urban landscape, dengue transmission is often highly localized, with certain sub-districts experiencing persistently higher incidence rates than their neighbors due to varying socio-environmental vulnerabilities [2]. This spatial heterogeneity poses a significant challenge for surveillance and control, as city-wide data aggregation frequently masks critical high-risk zones that require prioritized intervention [3][4]. Consequently, fine-scale identification of dengue risk patterns at the sub-district level is essential for implementing targeted vector control and evidence-based public health strategies [5].

A wide range of spatial and spatiotemporal analytical approaches have been proposed to examine dengue risk patterns in complex urban environments. Geographic information system (GIS)-based analyses

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

and spatial statistical methods have been widely used to detect dengue clusters and hotspots, enabling visualization of high-risk areas [6]. Systematic reviews have further emphasized the integration of epidemiological data with demographic and environmental indicators to improve dengue risk mapping in complex urban settings [7]. In Southeast Asia, where rapid urbanization continues to reshape socio-environmental conditions, such integrative spatial approaches have become increasingly relevant [8][9].

Beyond traditional spatial analysis, machine learning techniques have recently been applied to dengue studies, primarily focusing on incidence forecasting and early warning systems [10]. Several studies reported promising predictive accuracy using supervised learning algorithms when sufficient labeled data were available [11][12]. However, these forecasting-oriented models are often restricted to temporal predictions and require predefined outcome labels, which limits their usefulness for exploratory spatial risk stratification where the goal is to identify latent groupings across heterogeneous units [13]. Furthermore, supervised models are less suitable for uncovering overlapping spatial risk profiles in multi-dimensional urban datasets, where transmission boundaries are rarely clearly defined [14][15].

Despite these advances, existing dengue risk studies still exhibit significant limitations. Many clustering-based approaches rely on distance-based algorithms that assume homogeneous cluster shapes and rigid group assignments [16]. Such fixed partitioning oversimplifies the transitional or overlapping risk patterns commonly observed in dense, highly mobile urban environments. Furthermore, standard clustering approaches often struggle to maintain robustness when faced with correlated noise and the complex data structures inherent in multi-dimensional spatial datasets [17]. Additionally, many existing studies rely on single-year incidence observations, which are highly sensitive to short-term outbreak dynamics and may fail to capture persistent risk trends. As a result, there remains a need for a flexible and probabilistic clustering framework that can capture multi-year dengue risk heterogeneity at fine administrative scales. These limitations underscore the need for a flexible, model-based clustering approach capable of accommodating heterogeneous cluster geometries, overlapping risk profiles, and interrelated multi-dimensional urban indicators.

To address this gap, this study proposes a dengue risk stratification framework based on the Gaussian Mixture Model (GMM). GMM is a model-based clustering approach that represents data as a mixture of probability distributions, allowing for clusters with different shapes, sizes, and orientations while assigning probabilistic membership rather than rigid class labels [18]. This characteristic makes GMM particularly well-suited for multi-dimensional urban dengue data, where risk patterns are influenced by interacting demographic and built-environment factors that frequently overlap across sub-district boundaries [19]. In addition, the Bayesian Information Criterion (BIC) is employed to objectively determine the optimal number of dengue risk clusters, effectively reducing the risk of overfitting and subjective selection bias [20].

Accordingly, this study focuses on stratifying dengue risk at the sub-district level in Semarang City through a probabilistic clustering framework based on the Gaussian Mixture Model. Multi-year dengue incidence rate data (IR-DD+DBD) from 2021 to 2024 are integrated with eighteen demographic and built-environment indicators to characterize persistent spatial heterogeneity across 177 urban sub-districts.

In this study, urban conditions are represented using a set of demographic indicators, built-environment characteristics, and the availability of public facilities at the urban sub-district level.

The contributions of this study are as follows:

1. integrating multi-year dengue incidence data with multi-dimensional demographic and built-environment indicators at the sub-district level;
2. applying a Gaussian Mixture Model to capture heterogeneous and overlapping dengue risk patterns in an urban environment;
3. employing the Bayesian Information Criterion to objectively determine the optimal number of dengue risk clusters; and
4. providing interpretable fine-scale dengue risk profiles to support targeted surveillance and vector control planning by local health authorities.

The paper is organized as follows: first, the study area and methodology are described; then, clustering results are presented; discussion follows; and concluding remarks conclude.

II. Method

A. Study Area and Dataset

Semarang City is a major metropolitan area in Central Java, Indonesia, characterized by rapid urbanization and a heterogeneous population distribution that significantly influences dengue transmission dynamics. The city's diverse topography, spanning from low-lying coastal areas in the north to hilly terrain in the south, creates a complex urban environment with varying ecological conditions. Administratively, Semarang City consists of 177 urban sub-districts level, which serve as the primary unit of analysis in this study. This fine-

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

Table 1. Descriptive Statistics of Selected Demographic and Built - Environment Variables (n = 177 Sub - Districts)

Variable	Mean	SD	Min	Max
Area Size	25.68	18.44	5.17	58.27
Population Density	6719.92	3683.95	1211.94	12264.8
Total Population	9437.17	6450.75	584	37499
Number of DBD Patients (2024)	1.82	2.17	0	11
Number of DBD Deaths (2024)	0.03	0.18	0	1
Number of Dengue Patients (2024)	37.73	28.51	0	144
Larvae-Free Index (ABJ, %)	93.88	4.1	71.98	98.9
Number of Parks	17.53	16.43	1	49
Number of Neighborhood Units (RT)	51.42	39.71	4	292
Number of Community Units (RW)	9.3	9.66	1	89
Number of Family Heads	3002.3	1987.02	184	11224
Number of Primary & Secondary Students	13.59	7.78	2	44
Total Students	1496.4	1332.06	46	8281
Number of University Students	0.37	0.77	0	5
Number of Health Workers	3.25	3.09	0	20
Number of Shops/Markets/Stalls	110.14	86.18	6	426
Number of Public Facilities	7.19	20.08	0	254
Total Worship Facilities	35.08	21.58	5	166

scale administrative approach is essential for capturing the localized nature of dengue outbreaks in densely populated urban areas.

The dataset integrates multi-year dengue incidence rate data (IR-DD+DBD) for the period 2021–2024 with eighteen demographic and built-environment indicators representing population structure, residential density proxies, and public facilities at the sub-district level. By incorporating data across four years, the analysis effectively captures the temporal persistence of dengue risk patterns while minimizing the impact of short-term epidemic fluctuations [3][9]. The selection of these demographic and built-environment indicators follows established epidemiological research emphasizing the role of urbanization, population density, and environmental context in shaping dengue transmission risk [21][6][7].

B. Variable Selection and Description

The initial dataset comprised dengue incidence data from 177 urban sub-districts in Semarang City, combined with approximately 98 demographic, infrastructural, and public facility indicators obtained from secondary administrative sources. These indicators reflect various aspects of urban structure, population distribution, residential density, and public service availability, which are widely associated with dengue transmission dynamics in urban environments [6][7][9]. However, high-dimensional administrative datasets often contain significant redundancy and multicollinearity, which can destabilize unsupervised clustering models. Therefore, a rigorous variable screening process was conducted to identify a subset of indicators that are both epidemiologically meaningful and statistically appropriate for the Gaussian Mixture Model framework [11][22][23][24].

The variable selection process involved a multi-stage evaluation: descriptive statistical analysis to assess data completeness and variability; correlation-based evaluation to mitigate redundancy among highly correlated indicators; and domain relevance assessment based on established dengue risk factors. Specifically, indicators with extremely low variability or those exhibiting strong collinearity ($r > 0.8$) were scrutinized or consolidated to ensure analytical robustness [24][25]. In cases of strong collinearity, one of the correlated variables was retained based on epidemiological relevance and data completeness to

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

avoid redundancy while preserving interpretability. This approach balances model interpretability with the need to capture the multi-dimensional nature of the urban built environment. This combined statistical and domain-driven screening ensures that the retained variables are both epidemiologically meaningful and statistically suitable for stable Gaussian Mixture Model clustering.

As a result, eighteen demographic and built-environment variables were retained for further analysis. These variables represent population structure, residential density proxies (e.g., households and neighborhood units), public facilities (e.g., educational, religious, and commercial spaces), and environmental amenities (e.g., green spaces). These 18 indicators serve as multidimensional input for the GMM clustering, providing a comprehensive profile of the socio-environmental landscape across Semarang's urban sub-districts. Similar categories of variables have been reported as influential determinants of dengue risk in previous urban studies [6][7][12][25].

This variable selection stage focused on identifying relevant urban indicators, whereas the subsequent feature engineering stage focused on constructing temporal dengue incidence features used as inputs to the clustering model. Table 1 presents the descriptive statistics of the selected eighteen variables across 177 urban sub-districts. The selected indicators provide a comprehensive representation of urban heterogeneity while maintaining model stability and interpretability in the subsequent Gaussian Mixture Model clustering process [11][26].

C. Data Preprocessing

Prior to clustering analysis, a structured preprocessing pipeline was implemented to ensure the consistency, robustness, and comparability of all variables. The initial stage involved data cleaning and column harmonization to standardize variable names and numeric formats across the 177 sub-districts. Sub-district identifiers were utilized exclusively for labeling and spatial mapping; thus, they were excluded from the model-fitting process to prevent geographical bias in the unsupervised learning stage.

Missing values were handled using median imputation, a robust technique particularly suited for public health datasets where extreme values or outliers might otherwise distort the central tendency [27][28][29]. Median imputation was selected for its robustness to skewed distributions and the common presence of outliers in urban and epidemiological data. The proportion of missing values was relatively low across variables, ensuring that the imputation process did not substantially distort the original data structure. To account for the diverse numerical scales across the 18 indicators, ranging from population counts to percentages, all variables were normalized using Z-scores. This transformation ensures that each feature contributes equally to the Gaussian Mixture Model and prevents variables with larger magnitudes from dominating the cluster formation [30][31]. Finally, exploratory correlation analysis was performed to identify and mitigate high multicollinearity, ensuring a stable and interpretable clustering outcome that truly reflects the underlying spatial risk structures [32].

D. Feature Construction

After selecting urban indicators, a separate feature engineering step was performed to construct temporal dengue incidence features. While variable selection focused on identifying relevant urban characteristics, feature engineering was used to derive multi-year incidence rate variables that capture the temporal dynamics of dengue transmission. The final feature matrix for the clustering analysis comprises 22 variables, integrating 18 demographic and built-environment indicators with 4 annual dengue incidence rate variables (IR-DD+DBD) spanning 2021 to 2024. This multidimensional construction enables the model to characterize each sub-district not only by its current socio-environmental profile but also by its historical epidemiological trajectory. The incorporation of multi-year incidence data represents a deliberate strategy to strengthen the robustness of risk stratification. By accounting for transmission patterns over a four-year period, the framework can identify persistent risk hotspots while reducing sensitivity to episodic outbreak spikes or year-specific anomalies that often distort single-year observations [3][12]. This approach ensures that the resulting strata capture stable, long-term dengue transmission risks rather than transient fluctuations, thereby providing a more reliable foundation for sustainable public health planning [33][34].

E. Gaussian Mixture Model Clustering

Dengue risk stratification in this study is conducted using the Gaussian Mixture Model (GMM), a probabilistic, model-based clustering technique that represents observations as a finite mixture of multivariate Gaussian distributions. In contrast to conventional distance-based algorithms such as K-Means, which enforce rigid and often spherical cluster boundaries, GMM provides greater flexibility by accommodating clusters with heterogeneous shapes, scales, and orientations. This capability is particularly critical in urban epidemiological contexts, where dengue risk profiles frequently overlap across administrative units and exhibit substantial spatial and socioeconomic heterogeneity [11][14][35].

Model-based clustering approaches, including GMM, have been widely reported as effective in uncovering latent risk structures in public health and spatial epidemiology analyses, especially when underlying class memberships are not sharply separable [36][37].

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeemi.v8i1.1426

These characteristics make GMM an appropriate methodological choice for dengue risk stratification at the sub-district level. This study employs the Gaussian Mixture Model (GMM) as a probabilistic framework to stratify urban sub-districts based on multidimensional dengue risk profiles. The GMM assumes that the observed data $X = \{x_1, x_2, \dots, x_n\}$ are generated from a finite mixture of K multivariate Gaussian distributions, where each $x_i \in R^D$ represents the standardized feature vector of a sub-district [11][37]. Within this framework, the probability density function of an observation x_i is formulated as a weighted sum of Gaussian components, as shown in Eq. (1)[38]:

$$p(xi) = \sum_{k=1}^{K} \pi k N(xi|\mu k, \sum k) \tag{1}$$

where π_k denotes the mixing proportion of the k-th component, satisfying $\sum$_{$k = 1$}^{K} $\pi_\mathrm{k} = 1$, while μ_k and Σ_k represent the mean vector and covariance matrix of the corresponding Gaussian distribution [11][39]. The multivariate Gaussian component $\mathcal{N}(x_\mathrm{i} \mid \mu_\mathrm{k}, \Sigma_\mathrm{k})$ is defined in Eq. (2) [40] as follows:

$$N(xi|\mu k, \sum k) = \frac{1}{(2\pi)^{\frac{D}{2}} \left|\sum k\right|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(xi - \mu k)^T \sum k^{-1}(xi - \mu k)\right] \tag{2}$$

The set of model parameters $\Theta = \{\pi_\mathrm{k}, \mu_\mathrm{k}, \Sigma_\mathrm{k}\}$ is estimated using the Expectation–Maximization (EM) algorithm, which iteratively maximizes the log-likelihood of the observed data [11][37], as shown in **Eq. (3)**[38]:

$$\ell(\ominus) = \sum_{i=1}^{n} \log[\sum_{k=1}^{K} \pi k N(xi|\sum k)] \tag{3}$$

Unlike deterministic clustering methods such as K-Means, the GMM provides a soft clustering mechanism that allows each sub-district to be associated with probabilistic cluster memberships rather than rigid class assignments [39][41]. This property is particularly relevant in Semarang's urban context, where demographic and environmental risk factors often overlap across administrative boundaries. By explicitly modeling uncertainty and transitional risk patterns, the probabilistic formulation of GMM offers a more flexible and realistic representation of dengue transmission dynamics at the sub-district level. The optimal number of mixture components K is subsequently determined using the Bayesian Information Criterion (BIC), as described in the following section. This approach not only ensures statistical rigor but also reflects a thoughtful balance between model complexity and interpretability, acknowledging the human need for clarity in understanding data-driven insights.

F. Determination of Optimal Number of Clusters

Model selection was guided by the Bayesian Information Criterion (BIC), which evaluates the trade-off between model goodness-of-fit and model complexity in a statistically objective manner, and helps prevent overfitting in model-based clustering [39][11]. Several GMM configurations with different numbers of components were evaluated, and the model with the lowest BIC value was selected as the optimal solution. The formula can be written as Eq. (4)[42].

$$BIC = -2\boldsymbol{ln}(L^{\wedge}) + p\boldsymbol{ln}(n) \tag{4}$$

Where L denotes the maximum likelihood of the fitted GMM, p represents the number of estimated parameters, and $n$is the number of urban sub-districts. Information-theoretic criteria such as BIC are widely used in model-based clustering to ensure parsimonious model selection and to avoid unnecessary model complexity [10][37][41].

G. Cluster Labeling and Risk Interpretation

Following clustering, the cluster labels were reordered by the mean 2024 dengue incidence rate to ensure consistent interpretation of dengue risk levels. Clusters were subsequently categorized from low to high risk. Cluster profiling was performed by calculating mean values of demographic, built-environment, and dengue incidence variables within each cluster. This step facilitates interpretation of characteristic urban features associated with different dengue risk levels and supports targeted public health interventions [6][19]. Temporal patterns of dengue incidence across clusters were further examined to assess the persistence and evolution of dengue risk during the 2021–2024 period [3][12].

H. Methodological Workflow

The overall methodological workflow of this study is illustrated in Fig. 1, which summarizes the sequential analytical steps implemented to achieve dengue risk stratification at the sub-district level. The workflow begins with multi-source data collection, including dengue incidence rate data (IR-DD+DBD) for the period 2021–2024 and selected demographic and built-environment indicators for all urban sub-districts in Semarang City. Following data collection, a data preprocessing stage was conducted to ensure consistency and reliability of the dataset. This stage includes data cleaning and harmonization, handling missing values using median imputation, feature standardization through z-score normalization, and exploratory correlation analysis to reduce redundancy among highly correlated variables. These preprocessing steps are essential to ensure balanced feature contribution and stable clustering performance. Subsequently, probabilistic clustering was performed using the Gaussian Mixture Model (GMM) to capture heterogeneous and overlapping dengue risk patterns across urban sub-districts. To objectively determine the optimal number of clusters, the Bayesian Information Criterion (BIC) was used to evaluate multiple GMM configurations with varying numbers of components.

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

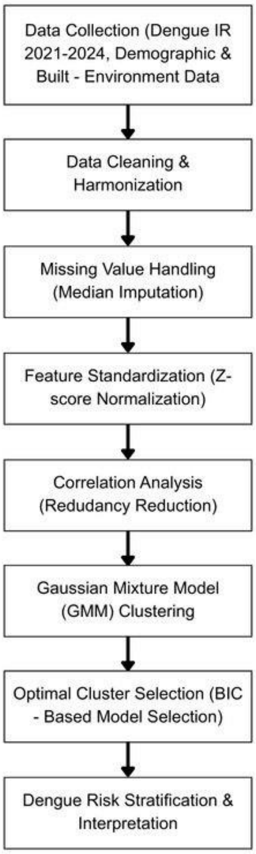

Fig. 1. Methodological workflow for dengue risk stratification using Gaussian Mixture Model and Bayesian Information Criterion.

The model with the lowest BIC value was selected as the optimal clustering solution. Finally, the resulting clusters were interpreted to generate dengue risk stratification profiles, enabling the identification of persistent spatial risk patterns across sub-districts. The complete workflow presented in Fig. 1 ensures methodological transparency and provides a structured framework for integrating multi-year epidemiological data with probabilistic clustering techniques to support fine-scale dengue risk assessment.

III. Result

A. Determination of Optimal Number of Clusters

The optimal number of clusters for dengue risk stratification was determined using the Bayesian Information Criterion (BIC). Several Gaussian Mixture Model (GMM) configurations with different numbers of components were evaluated to identify the model that provides the best balance between model fit and complexity.

As shown in Fig. 2, the BIC value decreases substantially when the number of clusters increases from two to three, indicating a notable improvement in model fit. However, further increasing the number of clusters beyond three results in only marginal reductions in BIC, suggesting diminishing returns in model performance. The minimum BIC value is achieved at three clusters, indicating that this configuration represents the most parsimonious and optimal clustering solution for the dataset.

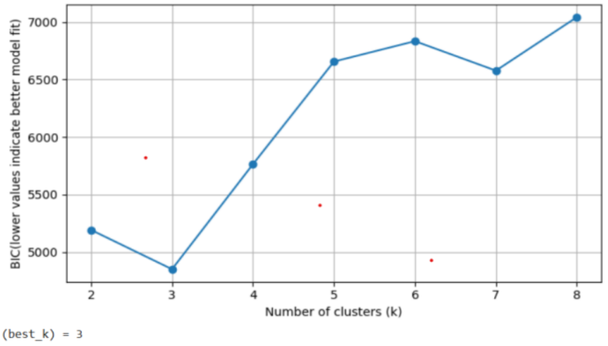

Fig. 2. Bayesian Information Criterion values obtained from Gaussian Mixture Model clustering for different numbers of clusters

B. Cluster Validation Using Silhouette Coefficient

To further ensure that the identified clusters exhibit measurable internal structure, an internal cluster validation was conducted using the silhouette coefficient. This metric assesses the degree of similarity among observations within the same cluster relative to those in neighboring clusters. The resulting silhouette score of 0.234 indicates a moderate but acceptable level of cluster separation, currently commonly observed in unsupervised clustering applied to heterogeneous urban epidemiological data. This result suggests that, despite partial overlap between clusters, the overall clustering structure captures meaningful distinctions in dengue risk characteristics across sub-districts.

Table 2. Distribution of urban sub-districts across dengue risk clusters.

Cluster	Number of Sub-District
0	92
1	77
2	8

C. Distribution of Dengue Risk Clusters

Based on the optimal three-component GMM configuration, the 177 urban sub-districts in Semarang City were stratified into three distinct dengue risk

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

clusters. The resulting distribution of sub-districts across these clusters is summarized in Table 2.

Cluster 0 contains 92 sub-districts, Cluster 1 includes 77 sub-districts, and Cluster 2 comprises 8 sub-districts. The unequal distribution across clusters indicates substantial spatial heterogeneity in dengue risk levels among urban sub-districts. The relatively small size of Cluster 2 suggests that high dengue risk is concentrated in a limited number of localized sub-districts rather than being uniformly distributed across the city. The spatial distribution of dengue risk clusters further indicates that the high-risk cluster (Cluster 2) is geographically limited and concentrated within a small number of urban sub-districts in Semarang City. These sub-districts are predominantly located in densely built-up urban areas characterized by high population concentration and intensive human activity. In contrast, low-risk sub-districts (Cluster 0) are more widely distributed across the city, including areas with lower neighborhood density and more residential land-use patterns. This spatial concentration suggests that dengue risk in Semarang City is not uniformly distributed but rather clustered in specific urban pockets that exhibit heightened vulnerability.

Fig. 3 illustrates the spatial distribution pattern of dengue risk clusters derived from the Gaussian Mixture Model. The visualization employs an abstract spatial projection to highlight the heterogeneity and relative concentration of clusters rather than precise administrative boundaries. The high-risk cluster (Cluster 2) appears spatially limited, concentrated within a small number of urban sub-districts, while low-risk (Cluster 0) and intermediate-risk (Cluster 1) clusters are more widely distributed. This pattern confirms that dengue risk in Semarang City is spatially heterogeneous and localized within specific urban pockets. Although this visualization does not represent exact administrative boundaries, it provides a clear spatial context for identifying localized high-risk sub-districts relevant for targeted public health interventions. Table 3 presents the mean values of demographic, built-environment, and multi-year dengue incidence indicators for each identified cluster. Substantial structural differences are evident across the three dengue risk strata. Cluster 0 (low risk) is characterized by lower population size, fewer neighborhood units (RT/RW), and lower concentrations of commercial and public facilities, indicating relatively stable residential environments. In contrast, Cluster 2 (high risk) exhibits markedly higher mean values across population-related indicators, neighborhood density proxies, and commercial facilities, reflecting intense human mobility and interaction. Cluster 1 demonstrates intermediate characteristics, suggesting a transitional urban structure between low- and high-risk settings.

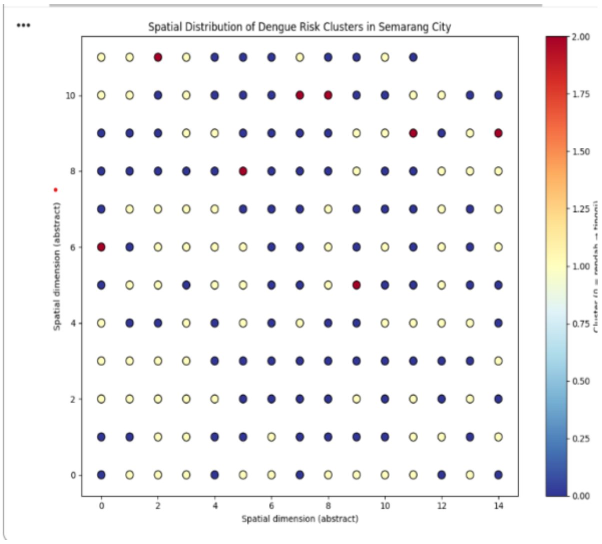

Fig. 3. Spatial distribution pattern of dengue risk clusters based on GMM clustering

D. Temporal Characteristics of Dengue Incidence (2021–2024)

To examine temporal differences in dengue dynamics among the identified clusters, average dengue incidence rates from 2021 to 2024 were analyzed for each cluster. The temporal trends are illustrated in Fig. 4. , Cluster 2 consistently exhibits the highest average dengue incidence rates across all 4 years, indicating a persistent high-risk pattern. In contrast, Cluster 0 demonstrates relatively low and stable incidence levels throughout the observation period. Cluster 1 shows intermediate incidence values with greater variability, reflecting a transitional risk profile. These results

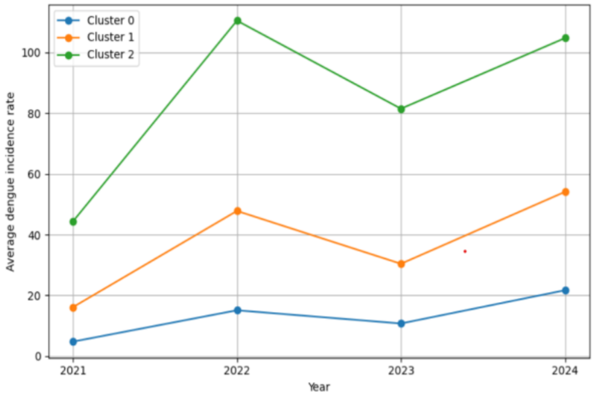

Fig. 4. Average dengue incidence rate trends for each identified cluster during the 2021–2024

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

Table 3. Mean Values of Demographic, Built-Environment, and Dengue Incidence Variables by Cluster.

Variable	Cluster 0 (Low Risk)	Cluster 1 (Intermediate Risk)	Cluster 2 (High Risk)
Total Population	5331.21	12421.48	27931.75
Female Population	2718.60	6262.12	14152.00
Male Population	2633.98	6168.69	13941.63
Number of Family Heads	1751.89	3905.71	8686.63
Number of Neighborhood Units (RT)	34.25	63.42	133.50
Number of Community Units (RW)	5.63	12.97	16.13
Number of Early Childhood Schools (PAUD)	4.96	9.81	22.88
Total Worship Facilities	22.88	43.83	91.13
Number of Mosque	4.15	8.56	15.75
Number of Langgar	12.55	23.44	49.88
Primary & Secondary Education Facilities	9.62	16.23	33.75
Number of Elementary Schools	2.34	3.26	6.13
Number of Shops/ Market/ Stalls	73.04	139.99	249.38
Area of Green/ Open Space	75464.59	29868.53	24271.62
Grocery Stores	35.86	76.88	119.50
Mini Markets	2.68	4.19	9.13
IR DD+DBD 2021	4.78	16.14	44.38
IR DD+DBD 2022	15.11	47.83	110.50
IR DD+DBD 2023	10.76	30.43	81.50
IR DD+DBD 2024	21.71	54.12	104.75

indicate that the GMM-based clustering captures not only spatial differentiation but also meaningful temporal patterns of dengue incidence.

E. Spatial Heterogeneity and Cluster Profiling

Further analysis was undertaken to examine the distributional characteristics of dengue incidence within each identified cluster. The temporal distribution of dengue incidence rates across clusters for each observation year is illustrated in Fig. 5. As shown, Cluster 2 consistently exhibits the highest median incidence rates and the widest interquartile ranges across all years, indicating persistently elevated dengue risk and substantial intra-cluster heterogeneity. In contrast, Cluster 0 displays tightly concentrated distributions with low median values, reflecting relatively stable and consistently low dengue transmission levels. Cluster 1 occupies an intermediate risk stratum, characterized by moderate median incidence rates and comparatively broader dispersion, suggesting transitional risk dynamics. The presence of extreme values within Cluster 2 further highlights localized sub-districts experiencing exceptionally high dengue incidence, underscoring the existence of concentrated hotspots that deviate markedly from the broader epidemiological patterns observed across the city. Beyond differences in median values, the high-risk cluster (Cluster 2) exhibits pronounced extreme dengue incidence levels. In the 2024 observation year, the mean dengue incidence rate (IR DD+DBD) in cluster 2 reached 104.75, which is nearly five times higher than that observed in the low-risk cluster. Moreover, a small number of sub-districts within cluster 2 recorded substantially higher incidence values, as reflected by the wide interquartile range and the presence of upper-end outliers in Fig. 5. These extreme values indicate localized hyper-endemic conditions within a small number of urban sub-districts, highlighting that dengue

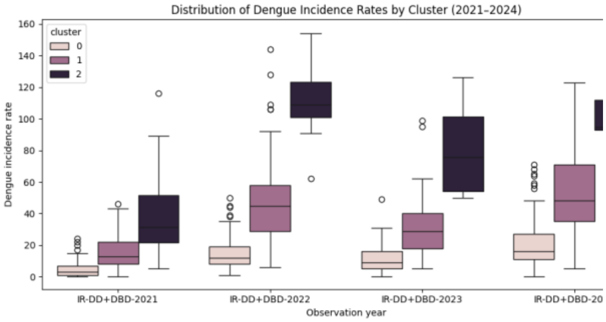

Fig. 5. Distribution of dengue incidence rates across clusters for each observation year from 2021 to 2024

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

risk in Semarang City is not only elevated in Cluster 2 but also highly heterogeneous in magnitude.

IV. Discussion

A. Interpretation of Dengue Risk Clusters

A Gaussian Mixture Model (GMM) was employed to stratify dengue risk in Semarang City, with model selection optimized based on the Bayesian Information Criterion (BIC). As illustrated in Fig. 2, the most significant reduction in the BIC value occurred at the transition from two to three clusters k = 3. This result suggests that a three-cluster configuration provides the most parsimonious and representative framework for capturing the urban risk heterogeneity in the study area without introducing unnecessary model complexity.

The spatial distribution of clusters shown in Fig. 3 and summarized in Table 2 indicates that the dengue burden is not uniformly distributed across the city but is heavily concentrated in eight specific sub-districts classified as Cluster 2. Although this cluster comprises only a small fraction of the total areas, its Incidence Rate (IR) remained consistently high throughout the 2021-2024 period. As shown in Fig. 4, Cluster 2 reached a peak IR of 104.75 in 2024, approximately 4.8 times higher than the IR of Cluster 0 (21.71). This persistent temporal pattern confirms that these eight sub-districts function as the primary epicenters of dengue transmission in Semarang City, rather than experiencing transient case surges.

The demographic and built-environment indicators presented in Table 3 provide a deeper explanation for the elevated risk in Cluster 2. This cluster exhibits a high average population of 27931.75 residents per sub-district, substantially exceeding Cluster 0's average of 5331.21. Furthermore, Cluster 2 is characterized by intensive economic activity, with an average of 249.38 shops and markets, alongside dense residential structures averaging 133.50 neighborhood units (RT) per sub-district. This convergence of high population density, commercial activity, and compact settlements likely increases contact between humans and Aedes aegypti mosquitoes, thereby accelerating local virus transmission [6][7].

From a public health operational perspective, the substantial disparities in IR values between clusters, as reflected in the data distribution in Fig. 5, suggest that a uniform, city-wide intervention approach is no longer effective. The persistently high incidence in Cluster 2, despite the implementation of routine control programs, points to the presence of structural risk factors. Therefore, dengue control efforts in these high-risk areas require more targeted strategies, including improvements to environmental infrastructure and residential area planning, to complement standard interventions such as fogging and larval source reduction.

B. Temporal Stability and Transitional Risk Dynamics

In terms of methodological effectiveness, the use of the Gaussian Mixture Model (GMM) in this study provides a more flexible clustering approach than conventional rigid clustering methods. Unlike algorithms that force each region into a specific cluster (hard clustering), GMM allows each sub-district to have a membership probability. This enables the model to better represent heterogeneous urban areas where risk boundaries often overlap. Based on the evaluation results, the GMM model applied to Semarang City data produced a Silhouette Score of 0.234. Although urban epidemiological data are highly complex, this value indicates a sufficient level of cluster separation. Compared with the hierarchical approach used by Lowe et al. [1], which applies strict risk classification, the probabilistic GMM approach is more adaptable in handling areas with transitional risk characteristics (Cluster 1), thereby reducing bias in defining priority zones. Comparison with previous studies summarized in Table 4 highlights the novelty of this research. The study by Feliciano et al. [9] employed spatial scan statistics to identify significant dengue risk clusters, with a primary focus on hotspot detection. In contrast, the GMM model in this study enables more detailed risk stratification at the sub-district administrative level by simultaneously considering 18 multidimensional indicators.

Furthermore, unlike the study by Yin et al. [6], which emphasized meteorological variables, the present study demonstrates that integrating built environment characteristics and settlement density contributes to more stable dengue risk patterns for long-term planning. The probabilistic nature of the GMM approach also helps reduce potential classification errors caused by rigid cluster boundaries. This characteristic has important practical implications for public health planning, as it supports a more precise and targeted allocation of dengue control resources in Semarang City based on the degree of risk probability in each region.

C. Comparison with Previous Dengue Risk Studies

A methodological comparison between the proposed GMM framework and previous studies is summarized in Table 4. Unlike the Bayesian Hierarchical Model approach used by Lowe et al. [1] and do Carmo et al. [9], which focused on relative risk (RR) and correlations between social determinants in Brazil, this study applies a Gaussian Mixture Model (GMM) to capture the probabilistic nature of dengue risk. The use of a GMM allows for the identification of transition areas characterized by overlapping clusters, an aspect often

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

Table 4. Comparison of Dengue Risk Modeling Approaches Used in Previous Studies and the Proposed Method.

Study	Method	Input Data & Scale	Limitations
Lowe et al. [1]	Bayesian hierarchical models & DLNM (via INLA).	Monthly data (2001-2019), 558 Brazilian microregions. Inputs: PDSI, temperature, & urbanization.	Depends on the accuracy of national secondary data; the issue of multicollinearity between water infrastructure and urbanization.
Yin et al. [6]	Systematic Review (PRISMA).	28 articles (2014-2022); building scale, grid, to district.	Variation in model performance across regions; lack of global standardization of input data.
Feliciano et al. [9]	Local Empirical Bayesian & Spatial Scan Statistics.	Annual data (2014-2017), 1,794 municipalities in Northeastern Brazil. Input: Dengue cases & social indicators.	Potential underreporting of cases; ecological designs cannot demonstrate individual cause-and-effect relationships.
Liu et al. [13]	ML-based modeling (XGBoost, SVR, MLP) & SARIMA.	Monthly data (2022-2023), 5 divisions in Bangladesh.	Limited to short-term data (2 years); accuracy depends on the quality of surveillance reporting.
This Study	Gaussian Mixture Model (GMM) - Probabilistic Clustering	Annual data (2021-2024), 177 urban villages in Semarang. 18 multidimensional urban indicators.	Relies on secondary administrative data; does not directly integrate meteorological/climatic variables.

overlooked in rigid partitioning models. From a spatial perspective, this study aligns with the systematic review conducted by Yin et al. [6], which employed multidimensional urban indicators to capture risk heterogeneity within complex built environments. In contrast, while the Bangladesh study by Liu et al. [12] prioritized the accuracy of temporal predictions (forecasting) through machine learning approaches such as XGBoost, the present Semarang study places greater emphasis on spatial risk stratification at the micro-scale (sub-district level). A key distinction also emerges in the selection of input variables. Whereas most reference studies [1][12] rely predominantly on meteorological parameters such as temperature and rainfall, the proposed model demonstrates that dengue risk stability in a large city like Semarang can be robustly identified through 18 structural urban indicators. This finding underscores that, in high-density urban areas, built environment factors are no less critical than climatic conditions in shaping long-term risk patterns, thereby reinforcing the need for interdisciplinary approaches in urban health research.

D. Methodological Advantages of the Gaussian Mixture Model

A key strength of the proposed approach lies in the probabilistic formulation of GMM-based clustering. Unlike distance-based methods (e.g., k-means), GMM accommodates clusters with heterogeneous shapes, sizes, and covariance structures, enabling correlated demographic, built-environment, and incidence variables to be jointly modeled [11]. The optimal number of clusters is objectively supported by the Bayesian Information Criterion (BIC). The substantial reduction in BIC from 5192.91 (K=2) to 4851.36 (K=3) indicates a meaningful improvement in model fit, while additional clusters introduce unnecessary complexity with diminishing explanatory gain. Probabilistic memberships further imply overlapping risk distributions, particularly for Cluster 1, supporting the suitability of probabilistic clustering for urban dengue risk analysis.

The findings of this study have important implications for dengue control in urban environments with heterogeneous risk profiles, such as Semarang City. The identification of distinct dengue risk clusters indicates that transmission is spatially concentrated and strongly influenced by localized urban and environmental characteristics rather than being uniformly distributed across the city [6]. From a public health policy perspective, these results suggest that dengue control strategies should shift from broad, citywide interventions to more targeted, location-specific approaches. The identification of Cluster 2 as a persistent high-risk epicenter indicates that these urban villages require intensive, structural interventions, including environmental management and infrastructure improvement, rather than relying solely on reactive measures such as chemical fogging.

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

This perspective aligns with previous studies emphasizing the critical role of built-environment conditions in sustaining vector breeding and transmission risk [43][44][45].

Operationally, the probabilistic risk stratification produced by the GMM framework provides a practical decision-support tool for prioritizing dengue interventions with limited public health resources. By distinguishing persistent high-risk zones from transitional areas, health authorities can allocate surveillance efforts, personnel, and prevention programs more efficiently, while maintaining vigilance in areas with emerging risk profiles. More broadly, integrating epidemiological data with multidimensional urban indicators underscores the importance of interdisciplinary, data-driven approaches in infectious disease management. This framework offers a scalable model that can support evidence-based dengue control planning in other dengue-endemic cities experiencing similar urban complexity.

While the proposed GMM framework provides an objective approach for dengue risk stratification, several limitations should be acknowledged. First, this study relies on secondary administrative data for urban indicators and dengue case reporting. Such data sources may be subject to underreporting or inaccuracies at the local level, which could affect cluster precision and, consequently, the operational reliability of the model, particularly for city-level early warning applications [46]. Second, the current framework does not explicitly incorporate meteorological or climatic variables, such as rainfall, humidity, or short-term temperature variations, into the clustering process. These factors are known to influence the population dynamics of Aedes aegypti mosquitoes and short-term transmission patterns [1][12]. In addition, the use of annual aggregated data at the village level may obscure finer-scale heterogeneity within sub-districts, including localized sanitation conditions at the neighborhood unit (RT/RW) level.

Finally, the proposed clustering approach is unsupervised and descriptive in nature, and therefore cannot be used to infer causal relationships or generate real-time outbreak predictions. Future research is encouraged to integrate dynamic environmental data, such as satellite-derived indicators or real-time weather observations, and to combine probabilistic clustering with predictive modeling frameworks to enhance the operational value of dengue early warning systems.

V. Conclusion

This study developed a Gaussian Mixture Model (GMM)-based dengue risk stratification framework by integrating 18 multidimensional urban indicators in Semarang City. The analysis results show that this probabilistic clustering method objectively identifies three risk levels (Low, Intermediate, and High Risk). Key findings indicate a consistently extreme caseload in Cluster 2, with an average Incidence Rate (IR) of 104.75, encompassing eight urban villages as the main epicenters of transmission during the 2021–2024 period. This study applies a GMM approach to capture data uncertainty and identify risk transition areas, which are not captured by rigid partitioning methods. Practically, these stratification results allow Semarang City health authorities to shift their dengue control strategy from a uniform approach to area-based interventions, prioritizing high-risk urban villages as epicenters of transmission. However, this model still has limitations, including reliance on secondary data and the absence of daily climate variables. Further research is expected to integrate dynamic environmental data and real-time prediction models to strengthen the early warning system for dengue outbreaks in urban environments.

Acknowledgment

The authors would like to express their sincere gratitude to the Health Office of Semarang City for providing access to the dengue incidence data used in this study. Appreciation is also extended to Universitas Dian Nuswantoro for the academic support and research facilities that enabled the completion of this work. The authors acknowledge the valuable feedback and guidance provided during the research and manuscript preparation process.

Funding

This research received no external funding

Data Availability

The datasets used in this study were obtained from secondary administrative sources, including dengue surveillance records from the Semarang City Health Office and publicly available urban demographic data. The processed data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to data-sharing regulations.

Author Contribution

Nabila Izzatil Ismah conceptualized the study, conducted data analysis, implemented the methodology, and prepared the original manuscript draft. Amiq Fahmi supervised the research process, provided methodological guidance, reviewed the analysis, and contributed to manuscript revision. All authors have read and approved the final manuscript.

Declarations

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

Ethical Approval
This study utilized aggregated secondary data at the sub-district level, without involving any individual or identifiable personal information. Therefore, ethical approval was not required.

Consent for Publication Participants.
Not applicable.

Competing Interests
The authors declare no competing interests.

References

[1] R. Lowe *et al.*, "Articles Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil : a spatiotemporal modelling study," *Lancet Planet. Heal.*, vol. 5, no. 4, pp. e209–e219, 2021, doi: 10.1016/S2542-5196(20)30292-8.

[2] I. G. Nyoman, M. Jaya, Y. Andriyana, B. Tantular, and S. S. Pangastuti, "Spatiotemporal Dengue Forecasting for Sustainable Public Health in Bandung , Indonesia : A Comparative Study of Classical , Machine Learning , and Bayesian Models," pp. 1–26, 2025.

[3] A. Oliveira, L. W. Lopes, and R. M. De Moraes, "Evaluation of spatial cluster detection methods for dengue fever in the state of Paraiba , Brazil," vol. 20, pp. 393–400, 2025, doi: 10.4081/gh.2025.1393.

[4] S. Hossain, M. M. Safa, N. F. Juthi, and N. Tasnia, "Bayesian hybrid statistical and machine learning models for dengue forecasting in Bangladesh: Temporal and spatial analysis for an early warning system," 2025.

[5] C. E. Sekarrini, Sumarmi, S. Bachri, D. Taryana, and E. A. Giofandi, "The application of geographic information system for dengue epidemic in Southeast Asia: A review on trends and opportunity," *J. Public health Res.*, vol. 11, no. 3, Jul. 2022, doi: 10.1177/22799036221104170.

[6] S. Yin, C. Ren, Y. Shi, J. Hua, H. Yuan, and L. Tian, "A Systematic Review on Modeling Methods and Influential Factors for Mapping Dengue-Related Risk in Urban Settings," 2022.

[7] A. Lim *et al.*, "A systematic review of the data , methods and environmental covariates used to map Aedes -borne arbovirus transmission risk," pp. 1–18, 2023.

[8] P. Kumar *et al.*, "Envisioning urban environments resilient to vector-borne diseases : a protocol to study dengue in Vietnam," pp. 17–27, doi: 10.37349/edht.2023.00004.

[9] R. Feliciano, J. Valter, J. Silva, and A. F. Pastor, "Spatiotemporal dynamics , risk areas and social determinants of dengue in Northeastern Brazil , 2014 – 2017 : an ecological study," *Infect. Dis. Poverty*, pp. 1–16, 2020, doi: 10.1186/s40249-020-00772-6.

[10] M. K. Gurram, S. Gwee, Y. Wang, and J. Pang, "Spatiotemporal distribution of sustained dengue hotspots associated with climate and urbanisation in Singapore," pp. 1–11, 2025.

[11] I. C. Gormley, T. B. Murphy, and A. E. Raftery, "Model-Based Clustering," pp. 573–595, 2023.

[12] Y. Koesmaryono, A. Sopaheluwakan, R. Hidayati, and B. D. Dasanto, "Spatiotemporal Characterization of Dengue Incidence and Its," 2024.

[13] B. Liu, F. Hossain, and S. Hossain, "OPEN A comparative evaluation of multiple machine learning approaches for forecasting dengue outbreaks in Bangladesh," 2025.

[14] B. Chassagnol *et al.*, "Gaussian Mixture Models in R," vol. 15, no. June, pp. 56–76, 2023.

[15] M. L. B. Lopes, "Unsupervised Learning Applied to the Stratification of Preterm Birth Risk in Brazil with Socioeconomic Data," 2022.

[16] H. Fidan and M. Erkan, "A comparative study for determining Covid-19 risk levels by unsupervised machine learning methods," *Expert Syst. Appl.*, vol. 190, no. August 2021, p. 116243, 2022, doi: 10.1016/j.eswa.2021.116243.

[17] M. L. Aug, "Robust Unsupervised Multi-task and Transfer Learning on Gaussian Mixture Models".

[18] M. E. Mar, "Confidence set for mixture order selection," pp. 1–14.

[19] M. Bai, "Ecohydrological zoning based on the Gaussian mixture model clustering method," vol. 27, no. 4, pp. 771–786, 2025, doi: 10.2166/hydro.2025.305.

[20] M. Science, E. Division, and S. Arabia, "Bayesian spatial functional data clustering: applications in disease surveillance b," pp. 1–19, 2019.

[21] C. E. Sekarrini, Sumarmi, S. Bachri, D. Taryana, and E. A. Giofandi, "The application of geographic information system for dengue epidemic in Southeast Asia: A review on trends and opportunity," *J. Public health Res.*, vol. 11, no. 3, 2022, doi: 10.1177/22799036221104170.

[22] O. Mendoza-cano, R. Danis-lozano, X. Trujillo, and M. Huerta, "Spatial patterns and clustering of dengue incidence in Mexico : Analysis of Moran ' s index across 2 , 471 municipalities from 2022 to 2024," vol. 2024, pp. 1–14, 2025, doi: 10.1371/journal.pone.0324754.

[23] C. Rotejanaprasert, K. Chinpong, A. B. Lawson, and P. Chienwichai, "Evaluation and comparison of spatial cluster detection methods for improved decision making of disease surveillance : a case

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

study of national dengue surveillance in Thailand,” *BMC Med. Res. Methodol.*, pp. 1–13, 2024, doi: 10.1186/s12874-023-02135-9.

[24] A. L. Asyary, “Spatial Analysis on Dengue Fever Vulnerability in the Provinces of South Sulawesi and East Nusa Tenggara in Indonesia,” vol. 91, no. 1, pp. 1–13, 2025, doi: 10.5334/aogh.4915.

[25] Y. Wang and X. Chen, “A Review of Bayesian Spatiotemporal Models in Spatial Epidemiology,” 2024.

[26] F. P. Rocha and M. Giesbrecht, “Machine learning algorithms for dengue risk assessment : a case study for São Luís do Maranhão,” *Comput. Appl. Math.*, vol. 41, no. 8, pp. 1–27, 2022, doi: 10.1007/s40314-022-02101-z.

[27] C. Corvalan *et al.*, “Towards Climate Resilient and Environmentally Sustainable Health Care Facilities,” 2020.

[28] M. Mahdavian, K. Yin, and M. Chen, “Robust Visual Teach and Repeat for UGVs Using 3D Semantic Maps”.

[29] C. Inference and H. Sciences, *Causal Inference and Machine Learning*.

[30] V. Gómez-rubio, “Journal of Statistical Software,” vol. 100, no. 1, pp. 1–7, 2021, doi: 10.18637/jss.v100.i01.

[31] E. M. Dillon *et al.*, “Dermal denticle assemblages in coral reef sediments correlate with conventional shark surveys,” vol. 2020, no. September 2019, pp. 362–375, 2020, doi: 10.1111/2041-210X.13346.

[32] Y. Cao *et al.*, “Epidemiological Trends and Age – Period – Cohort Effects on Dengue Incidence Across High-Risk Regions from 1992 to 2021,” pp. 1–17, 2025.

[33] J. Briseno-ramirez *et al.*, “Annals of Epidemiology Trends and cyclical patterns of dengue disease in Mexico : A 40-year time series analysis,” vol. 112, no. October, pp. 53–63, 2025, doi: 10.1016/j.annepidem.2025.10.022.

[34] X. Shou, G. Mavroudeas, A. New, and K. Arhin, “Supervised Mixture Models for Population Health,” 2016.

[35] Z. R. Mccaw, H. Julienne, and H. Aschard, “MGMM : An R Package for fitting Gaussian Mixture Models on Incomplete Data,” pp. 0–46, 2020.

[36] P. A. J and E. P. T, “Spatial Autoregressive Model with Mixture of Gaussian Distribution for the Random Effect : Formulation , Estimation and Application,” vol. 12, no. 4, pp. 314–323, 2024, doi: 10.13189/ms.2024.120402.

[37] U. Padjadjaran, “Clustering of Childhood Diarrhea Diseases Using Gaussian Mixture Model,” pp. 1–17, 2024.

[38] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, 2006.

[39] S. O. Eui *et al.*, “Accessing the fine temporal scale of EUV brightenings and their quasi-periodic pulsations : 1 second cadence observations by,” vol. 2, 2025.

[40] P. Economou, “Research in Statistics A clustering algorithm for overlapping Gaussian mixtures,” *Res. Stat.*, vol. 00, no. 00, p. 0, 2023, doi: 10.1080/27684520.2023.2242337.

[41] S. Morrissette, S. Muthukumarana, and M. Turgeon, “Machine Learning with Applications Parsimonious Bayesian model-based clustering with dissimilarities,” *Mach. Learn. with Appl.*, vol. 15, no. January, p. 100528, 2024, doi: 10.1016/j.mlwa.2024.100528.

[42] M. Optimization and S. Energies, “Multi-Objective Optimization and Performance Assessments of an Integrated Energy System Based on Fuel , Wind and,” 2021.

[43] M. Amierul *et al.*, “The application of environmental management methods in combating dengue : a systematic review,” *Int. J. Environ. Health Res.*, vol. 33, no. 11, pp. 1148–1167, 2023, doi: 10.1080/09603123.2022.2076815.

[44] M. Rahmilah, “The Impact of Fogging Activities on Dengue Hemorrhagic Fever (DHF) Prevention in Balocci District,” no. August, pp. 590–594, 2025.

[45] I. Arfan, L. Sulistyorini, M. Sulistyowati, A. Rizky, S. Mohd, and S. Asia, “Dialogues in Health Prevention and control of dengue and Aedes mosquitoes in South and Southeast Asia : Interventions, challenges, and future recommendations,” *Dialogues Heal.*, vol. 7, no. August, p. 100253, 2025, doi: 10.1016/j.dialog.2025.100253.

[46] T. Salam, M. N. Begum, and M. Rahman, “Evaluating surveillance systems to understand dengue burden in Bangladesh : a narrative review,” 2026.

Author Biography

Nabila Izzatil Ismah is an undergraduate student in Informatics Engineering at Universitas Dian Nuswantoro, Semarang, Indonesia. She is currently completing her final year of study with a research focus on data science, machine learning, and spatial data analytics. Her academic interests include unsupervised learning, probabilistic modeling, and the application of computational methods to public health problems, particularly dengue risk analysis in urban environments.

She has experience working with multi-year epidemiological datasets, demographic indicators, and

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426

built-environment variables using data-driven analytical approaches. Her research emphasizes methodological robustness, model interpretability, and objective model selection techniques. In addition to data analysis, she has experience in developing web-based information systems and data visualization tools to support analytical reporting. Through her academic work, she aims to contribute to interdisciplinary research that integrates informatics, public health, and urban analytics.

Amiq Fahmi earned his Bachelor’s degree in Informatics Management from STMIK Dian Nuswantoro, Semarang, in 1998 and subsequently joined the faculty as a lecturer in the undergraduate program. He completed his Master of Science in Informatics Engineering at the Faculty of Computer Science, Universitas Dian Nuswantoro, in 2004, and in 2025 was awarded the Doctor of Philosophy (Ph.D.) in Electrical Engineering from Institut Teknologi Sepuluh Nopember, Surabaya. He is currently engaged in teaching, research, and community service at the Department of Information Systems, Faculty of Computer Science, Universitas Dian Nuswantoro. His scholarly interests focus on intelligent information systems and artificial intelligence applications in e-Health, with particular emphasis on predictive modeling and decision support for infectious diseases such as dengue fever.

Manuscript received 12 November 2025; Revised 08 January 2026; Accepted 18 January 2026; Available online 27 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1426