

Integrating wastewater-based epidemiology with AI to predict enteric disease transmission dynamics in low-income communities

Grace Oluwaseyi Owojori ^{1,*} and Oluwagbemisola Christianah Fasuyi ²

¹ *Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, USA.*

² *College of Nursing, Rush University Medical Centre, Chicago IL, USA.*

Magna Scientia Advanced Biology and Pharmacy, 2026, 17(01), 017-036

Publication history: Received on 31 December 2025; revised on 07 February 2026; accepted on 10 February 2026

Article DOI: <https://doi.org/10.30574/msabp.2026.17.1.0017>

Abstract

Wastewater-based epidemiology (WBE) has emerged as a powerful population-level surveillance approach for monitoring infectious diseases by detecting biomarkers shed into communal wastewater systems. Its relevance is particularly pronounced in low-income communities, where clinical reporting is often fragmented, delayed, or inaccessible, and where enteric diseases remain a major public health burden. Recent advances in artificial intelligence (AI) offer new opportunities to enhance WBE by enabling scalable data integration, pattern recognition, and predictive modeling across complex environmental and socio-demographic contexts. From a broad public health perspective, integrating WBE with AI-driven analytics can transform passive wastewater measurements into proactive early-warning systems capable of informing targeted interventions, optimizing resource allocation, and strengthening outbreak preparedness. This study narrows the focus to enteric disease transmission dynamics in low-income settings, where infrastructural variability, informal sanitation networks, and climate sensitivity complicate traditional surveillance. We propose an AI-augmented WBE framework that combines microbial load data, temporal wastewater signals, environmental covariates, and community-level indicators to model transmission pathways and forecast outbreak risks. By leveraging machine learning and time-series modeling, the framework aims to improve prediction accuracy, reduce detection latency, and support equitable public health decision-making. The integration of WBE and AI thus represents a scalable, cost-effective strategy for strengthening enteric disease surveillance and resilience in resource-constrained communities.

Keywords: Wastewater-Based Epidemiology; Artificial Intelligence; Enteric Diseases; Disease Prediction; Low-Income Communities; Public Health Surveillance

1. Introduction

1.1. Global Burden of Enteric Diseases in Low-Income Communities

Enteric diseases remain among the leading causes of morbidity and mortality in low-income communities, disproportionately affecting children, immunocompromised individuals, and populations lacking access to safe water and sanitation [1]. Major enteric pathogens such as *Vibrio cholerae*, *Salmonella Typhi*, rotavirus, and norovirus are responsible for recurrent outbreaks and sustained endemic transmission across many regions of Africa, South Asia, and Latin America [2]. Cholera and typhoid fever are closely associated with contaminated water sources and inadequate sewage disposal, while viral agents such as rotavirus and norovirus spread rapidly through fecal-oral pathways in densely populated environments [3]. Despite their public health significance, the true burden of these infections is persistently underestimated due to structural limitations within existing surveillance systems. Underreporting is widespread, driven by limited healthcare access, dependence on informal treatment pathways, and the absence of

* Corresponding author: Grace Oluwaseyi Owojori

comprehensive notification mechanisms in many settings [4]. Diagnostic delays further exacerbate this challenge, as laboratory confirmation is often unavailable or restricted to centralized facilities with extended turnaround times. Symptom-based surveillance lacks specificity because diarrheal presentations overlap across multiple pathogens and non-infectious causes [5]. As a result, clinical reporting captures only a fraction of community-level transmission, often identifying outbreaks after extensive spread has occurred [6]. These constraints underscore the need for population-scale surveillance approaches that extend beyond healthcare facilities.

1.2. Wastewater-Based Epidemiology as a Population-Level Surveillance Tool

Wastewater-based epidemiology (WBE) has emerged as a complementary surveillance paradigm capable of capturing aggregated health signals at the community level through the analysis of biological markers in sewage systems [2]. The conceptual foundation of WBE rests on the principle that infected individuals shed pathogens or genetic material in feces and urine, which subsequently accumulate within shared wastewater networks. By sampling influent wastewater, it becomes possible to infer population-level infection trends without reliance on individual testing [7]. Pathogen shedding dynamics vary across enteric organisms, influenced by infection stage, pathogen load, and host immunity, yet their collective signals can be temporally aggregated to reflect transmission intensity [8]. Wastewater signals therefore function as integrative indicators, smoothing individual variability while preserving outbreak-related trends. In low-resource settings, WBE offers distinct advantages over conventional surveillance. It bypasses healthcare access barriers, reduces dependence on clinical diagnostics, and enables monitoring of both symptomatic and asymptomatic infections [9]. Additionally, wastewater sampling can be implemented at relatively low cost using existing sanitation infrastructure, even where treatment systems are decentralized or informal. These attributes position WBE as a scalable, non-intrusive surveillance strategy particularly suited to environments characterized by fragmented health reporting systems and high enteric disease burden.

1.3. Role of Machine Learning in Public Health Prediction

Machine learning (ML) has become an increasingly important analytical paradigm in public health due to its capacity to model complex, nonlinear relationships within high-dimensional data [8]. Unlike traditional statistical surveillance methods, which often rely on linear assumptions and predefined thresholds, ML techniques can learn patterns directly from data, adapting to evolving transmission dynamics. This capability is particularly relevant for wastewater-based datasets, where pathogen concentrations are influenced by interacting biological, environmental, and infrastructural factors. ML models are well suited to handling temporal dependence through time-series learning, enabling early detection of subtle changes preceding outbreaks [4]. Furthermore, they support multivariate integration, allowing wastewater signals to be combined with environmental variables, demographic indicators, and historical case trends. By capturing nonlinear interactions and delayed effects, ML-based approaches offer the potential to transform descriptive surveillance into predictive intelligence, supporting proactive public health decision-making rather than reactive outbreak response.

1.4. Research Gap and Study Contributions

Despite the growing adoption of wastewater-based epidemiology, predictive frameworks that systematically integrate WBE with machine learning for enteric disease surveillance remain limited. Existing applications are largely descriptive, focusing on retrospective trend analysis rather than forward-looking transmission prediction [4]. There is a clear gap in end-to-end, ML-enabled WBE models tailored to the structural realities of low-income communities. This study addresses that gap by proposing a comprehensive machine learning pipeline for wastewater-based enteric disease prediction [10]. The contributions include the development of feature-engineered transmission indicators, integration of heterogeneous data sources, and rigorous performance benchmarking against established surveillance standards, thereby advancing predictive public health analytics.

2. Conceptual Framework and Study Design

2.1. System Architecture Overview

The proposed system architecture follows an end-to-end pipeline designed to transform raw wastewater measurements into actionable epidemiological predictions through machine learning integration [6]. The process begins with systematic wastewater sampling at strategically selected community catchment points, capturing composite influent signals representative of population-level pathogen shedding. Preprocessing follows, including normalization, noise reduction, and temporal alignment to correct for dilution effects and sampling irregularities [9]. Feature extraction constitutes the next phase, where biologically and epidemiologically meaningful indicators are derived from wastewater signals, environmental covariates, and historical disease trends. These engineered features

are then supplied to machine learning models capable of learning nonlinear relationships and temporal dependencies within the data. Model outputs generate short-term and medium-term predictions of enteric disease transmission intensity, which are subsequently evaluated through validation procedures using historical surveillance benchmarks [12]. This modular architecture supports adaptability, allowing individual components to be updated as data availability or modeling requirements evolve. Importantly, the pipeline is designed to function under infrastructural constraints typical of low-income settings, emphasizing computational efficiency and minimal reliance on high-resolution clinical data [14].

Figure 1 illustrates the conceptual architecture of the AI-enabled WBE prediction framework, highlighting data flows and analytical stages.

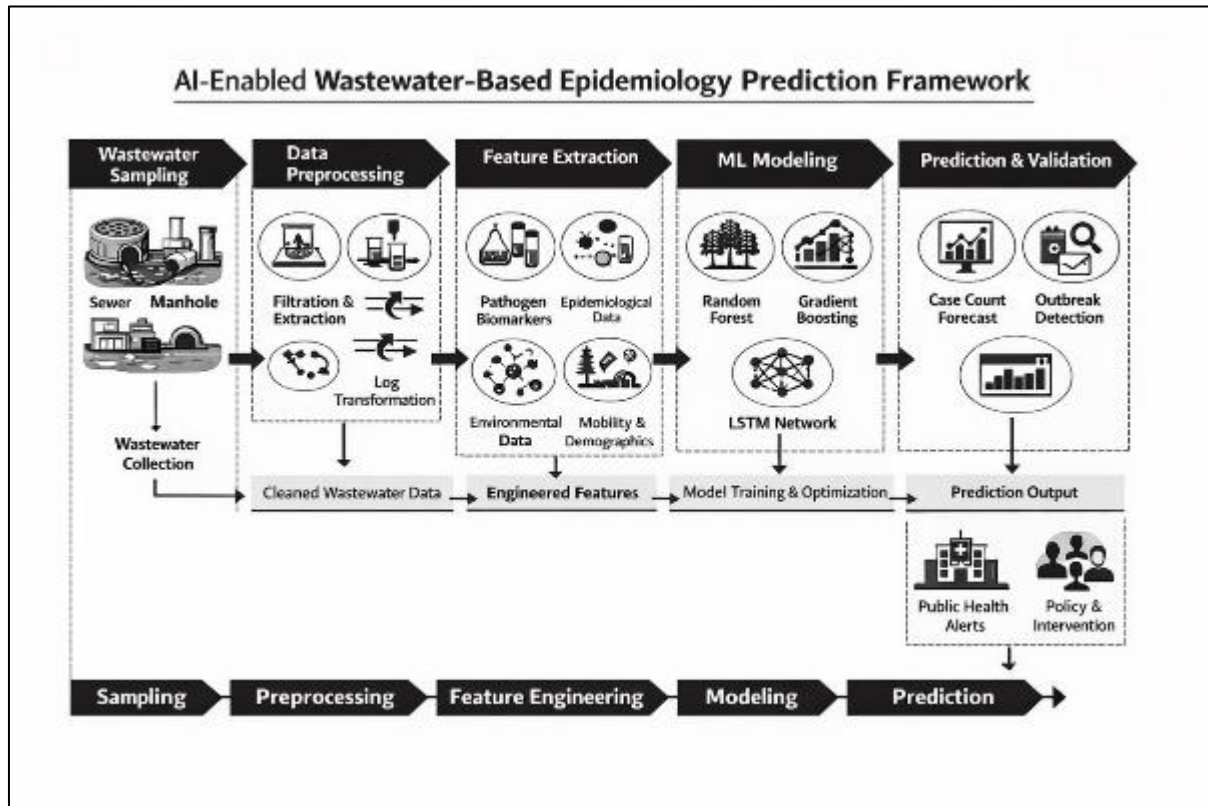


Figure 1 Conceptual architecture of the AI-enabled WBE prediction framework

2.2. Epidemiological Assumptions and Modeling Scope

The modeling framework is grounded in population-level inference rather than individual diagnosis, reflecting the aggregated nature of wastewater data and its suitability for community-scale surveillance [7]. It assumes that wastewater pathogen concentrations approximate collective infection dynamics, even in the presence of heterogeneous shedding behaviors across individuals. Temporal aggregation is employed to smooth short-term variability while preserving meaningful transmission trends, based on the assumption that enteric disease progression and reporting delays occur over multi-day intervals [10]. The model further accounts for community-level heterogeneity by incorporating contextual variables such as population density, sanitation coverage, and environmental conditions, acknowledging that transmission dynamics are not uniform across settings [13]. These assumptions define the scope of inference, positioning the framework as a predictive early-warning system rather than a replacement for clinical diagnostics.

2.3. Ethical, Privacy, and Deployment Considerations

Wastewater-based epidemiology inherently minimizes privacy risks because data are collected at aggregated levels without identifying individuals, households, or specific behaviors [8]. This non-identifiability supports ethical surveillance while enabling broad population coverage. The framework adopts an equity-oriented perspective, prioritizing underserved communities where conventional health monitoring is weakest and disease burden is highest [11]. From a deployment standpoint, feasibility in low-income regions is central to system design, emphasizing low-cost sampling methods, limited dependence on laboratory-intensive assays, and scalable computational workflows [14].

Ethical governance also requires transparent communication with stakeholders to prevent misinterpretation of results or stigmatization of communities. By aligning technological design with ethical safeguards and practical constraints, the proposed framework supports responsible, inclusive, and sustainable public health surveillance across diverse resource-limited environments [6].

3. Data Acquisition and Pre-Processing

3.1. Wastewater Sampling Strategy

Wastewater sampling was designed to capture representative population-level signals of enteric pathogen circulation across heterogeneous sanitation infrastructures commonly observed in low-income communities [12]. Sampling locations included influent points at centralized wastewater treatment facilities where available, upstream manholes within sewered neighborhoods, and informal drainage or treatment inlets serving peri-urban settlements lacking full sewer connectivity. This multi-location approach reduces spatial bias and enhances signal representativeness by accounting for fragmented wastewater flows [15]. Temporal resolution was selected to balance epidemiological sensitivity and operational feasibility. Daily grab samples allow fine-grained temporal tracking of rapid transmission shifts, while weekly composite samples smooth short-term variability and reduce laboratory burden, making them suitable for sustained surveillance programs [17]. Pathogen biomarkers targeted both viral and bacterial agents relevant to enteric disease transmission. Viral RNA markers included norovirus and rotavirus genomic fragments, while bacterial gene copies such as *ctxA* and *invA* were used to infer cholera and typhoid activity. Quantification via molecular assays produces concentration estimates that serve as proxies for collective pathogen shedding, forming the foundational input for downstream analytical modeling [19].

3.2. Epidemiological and Environmental Data Sources

Wastewater signals were complemented with epidemiological and environmental datasets to contextualize transmission dynamics and enhance predictive performance [13]. Reported enteric case counts were obtained from local health facilities, community health posts, and district surveillance reports, acknowledging that these data capture symptomatic presentations rather than total infections. While incomplete, these records provide critical temporal anchors for model validation and benchmarking [16]. Environmental variables known to influence enteric disease transmission were incorporated, including rainfall, ambient temperature, and indicators of sanitation coverage. Rainfall data capture dilution effects and flooding-related contamination risks, while temperature influences pathogen persistence and replication rates within environmental reservoirs [18]. Sanitation coverage metrics approximate exposure risk and infrastructural vulnerability. Population density estimates were derived from census records and gridded population datasets, while mobility proxies were inferred from transportation flow patterns and market activity indices where available. Together, these auxiliary datasets enable multivariate modeling of disease risk, allowing machine learning algorithms to disentangle wastewater signal fluctuations driven by epidemiological processes from those caused by environmental or infrastructural variation [20].

Table 1 Summary of Data Sources and Variables Used in the ML-Enabled WBE Framework

Data Category	Variable	Description	Temporal Resolution	Purpose in Model
Wastewater Data	Pathogen RNA concentration	Log-transformed viral/bacterial gene copies per unit volume	Daily / Weekly	Proxy for community-level infection burden
Wastewater Data	Signal variance	Variability of pathogen concentration over time	Weekly	Detect outbreak instability and transitions
Epidemiological Data	Reported case counts	Clinically reported enteric disease cases	Weekly	Ground truth for prediction and validation
Epidemiological Data	Growth rate	Relative change in case counts between time steps	Weekly	Capture transmission acceleration
Environmental Data	Rainfall	Cumulative precipitation	Daily / Weekly	Adjust for dilution and flooding effects

Environmental Data	Temperature	Mean ambient temperature	Weekly	Pathogen persistence modifier
Socio-Structural Data	Sanitation coverage	Proportion of population connected to sanitation systems	Static / Annual	Weight wastewater representativeness
Demographic Data	Population density	Residents per unit area	Static	Approximate contact intensity

3.3. Data Cleaning and Normalization

Raw wastewater and auxiliary datasets were subjected to systematic cleaning and normalization to ensure analytical robustness and comparability across sampling periods [14]. Missing values arose from interrupted sampling schedules, assay failures, or incomplete reporting of epidemiological variables. These gaps were addressed using time-aware imputation techniques, including forward filling for short interruptions and interpolation for longer gaps, while preserving temporal continuity [17]. Outlier suppression was applied to mitigate the influence of extreme measurements caused by laboratory artifacts, transient hydraulic disturbances, or anomalous environmental events. Outliers were identified using percentile-based thresholds and robust dispersion metrics, ensuring biologically plausible signal retention [19]. Given the highly skewed distribution of microbial concentration data, log-transformation was applied to pathogen measurements prior to modeling.

3.3.1. Equation (1): Log-normalized pathogen load

$$X' = \log(X + \epsilon)$$

where X represents the raw pathogen concentration and ϵ is a small constant preventing undefined values. This transformation stabilizes variance, reduces right-skewness, and enhances compatibility with machine learning algorithms that assume approximately symmetric feature distributions [12]. Normalization across features was subsequently performed to align variable scales, preventing dominance of high-magnitude inputs during model training. Collectively, these preprocessing steps improve signal fidelity, reduce noise-induced bias, and support reliable feature learning across heterogeneous data sources [18].

3.4. Temporal Alignment and Lag Construction

Temporal alignment is critical for linking wastewater signals with observed epidemiological outcomes due to delays inherent in disease progression and reporting processes [15]. Wastewater measurements reflect pathogen shedding that may precede symptom onset, healthcare seeking, and formal case notification. To account for these delays, lag windows were constructed between wastewater features and reported case counts, spanning from $t - 1$ to $t - k$, where k represents the maximum biologically plausible incubation and reporting interval. Lag selection was informed by known incubation periods of common enteric pathogens, which typically range from one to several days for viral agents and up to two weeks for certain bacterial infections [16]. Multiple lagged features were retained to allow models to learn optimal temporal dependencies rather than imposing fixed assumptions.

Figure 2 illustrates the temporal alignment process, depicting wastewater signal trajectories, lagged feature construction, and corresponding reported case timelines.

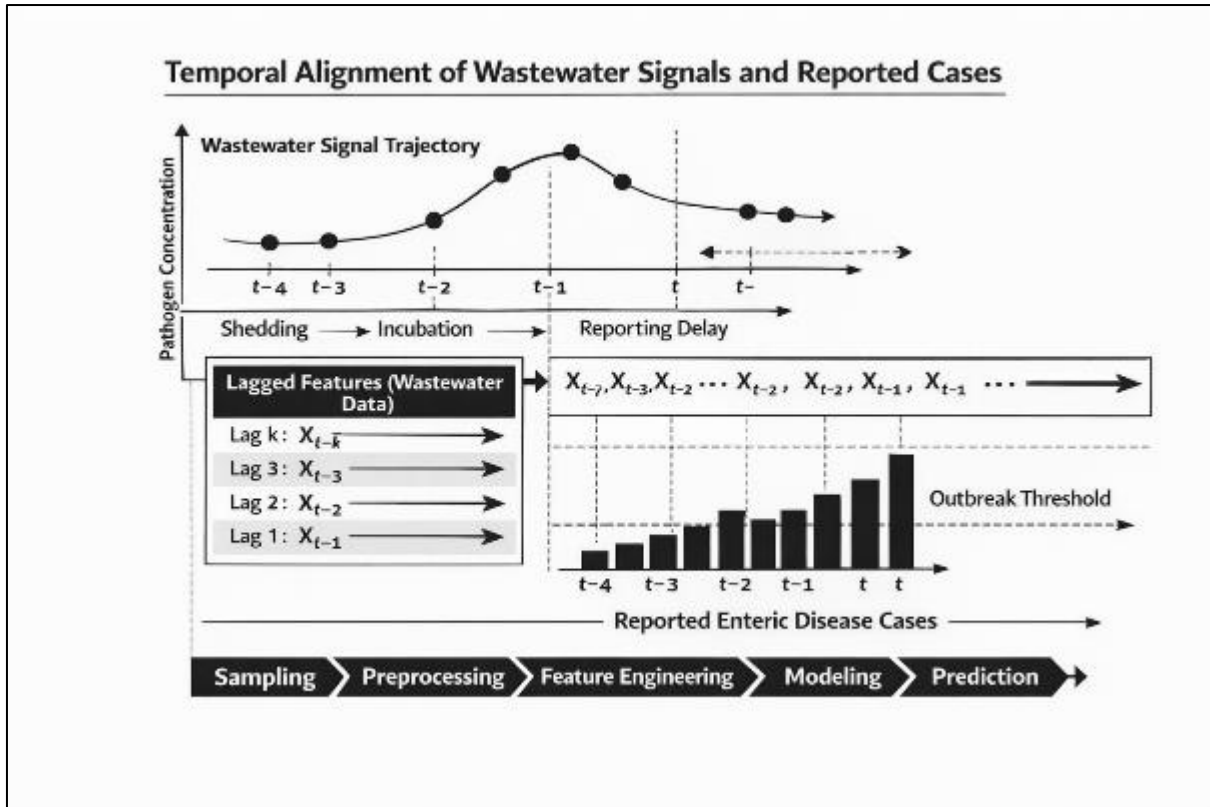


Figure 2 Temporal alignment process

This alignment strategy enables machine learning models to capture leading indicator behavior in wastewater data, improving early-warning capacity while maintaining epidemiological interpretability [20]. By explicitly modeling temporal offsets, the framework accommodates reporting delays and variable shedding patterns, ensuring that predictive outputs reflect realistic transmission dynamics rather than coincidental correlations.

4. Feature Engineering and Representation Learning

4.1. Wastewater Signal Features

Feature engineering of wastewater signals focused on transforming raw pathogen concentration time series into interpretable statistical descriptors that reflect underlying transmission dynamics. The mean pathogen concentration was computed to summarize the average microbial burden observed within each aggregation window, representing the central tendency of community-level pathogen shedding [18].

Equation (2): Mean pathogen concentration

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

In this formulation, x_i denotes the pathogen concentration measured in the i -th wastewater sample within a defined temporal window, and n represents the total number of samples aggregated over that period. The parameter μ therefore captures the expected pathogen load contributed by the population during the sampling interval, smoothing short-term fluctuations caused by hydraulic variability or individual shedding heterogeneity.

To quantify dispersion around the mean, signal variance was calculated to characterize temporal instability and outbreak-driven variability in pathogen shedding [21].

Equation (3): Signal variance

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Here, σ^2 measures the degree to which individual observations deviate from the mean concentration μ . Elevated variance values indicate heterogeneous infection stages or rapid transmission escalation within the population, whereas low variance suggests stable endemic circulation. Peak intensity features were additionally derived as local maxima within sliding windows to capture abrupt surges associated with outbreak onset.

Signal decay rates were estimated to describe the rate at which pathogen concentrations decline following peak events. These decay parameters reflect reductions in transmission, intervention effects, or environmental clearance processes. By incorporating decay behavior, the feature set captures not only outbreak emergence but also recovery trajectories, improving discrimination between sustained epidemics and transient contamination events [25].

4.2. Epidemiological Interaction Features

Epidemiological interaction features were constructed to explicitly link wastewater-derived indicators with observed disease outcomes, enabling models to learn biologically plausible lead-lag relationships [19]. Case-to-wastewater correlation indices quantified the strength and direction of association between lagged wastewater features and reported case counts, supporting interpretability of temporal dependencies.

Growth rate proxies were derived from epidemiological time series to represent the relative speed of disease propagation independent of absolute case volume [22].

Equation (4): Weekly growth rate

$$g_t = \frac{C_t - C_{t-1}}{C_{t-1}}$$

In this equation, C_t denotes the number of reported enteric disease cases at time t , while C_{t-1} represents cases reported during the preceding interval. The parameter g_t expresses proportional change rather than absolute difference, allowing comparison across communities with differing population sizes and reporting capacities. Positive values of g_t indicate accelerating transmission, whereas negative values suggest declining incidence. When paired with wastewater-derived predictors, growth rates provide dynamic epidemiological context that enhances early outbreak detection and reduces sensitivity to underreporting bias [24].

4.3. Environmental and Socio-Structural Features

Environmental and socio-structural features were incorporated to account for non-biological factors influencing wastewater signal interpretation and disease transmission. Rainfall-driven dilution effects were modeled to capture the impact of precipitation on measured pathogen concentrations. Increased rainfall can reduce observed concentrations by increasing wastewater volume, even when infection prevalence remains unchanged, necessitating explicit representation of precipitation intensity and duration [20].

Sanitation coverage weighting parameters were introduced to adjust wastewater signals based on the proportion of the population effectively contributing to sampled wastewater flows. These weights account for variability in sewer connectivity, latrine usage, and informal drainage systems common in low-income settings [23]. Higher weights indicate greater representativeness of the sampled signal, while lower values reflect partial population coverage.

Additional socio-structural parameters included population density and housing typology indicators, which approximate contact intensity and fecal-oral exposure pathways. Collectively, these features allow the model to disentangle true epidemiological signals from infrastructural and environmental distortions, improving generalizability across heterogeneous communities [25].

4.4. Feature Scaling and Dimensionality Reduction

Prior to model training, all engineered features were standardized to ensure numerical comparability and algorithmic stability [18]. Z-score normalization was applied to each feature individually.

4.4.1. Equation (5): Z-score normalization

$$Z = \frac{x - \mu}{\sigma}$$

In this formulation, x represents an observed feature value, μ is the feature-specific mean computed over the training dataset, and σ denotes the corresponding standard deviation. The resulting standardized variable Z has zero mean and unit variance, ensuring that features measured on different scales such as pathogen concentrations, rainfall totals, and population metrics contribute proportionately during model optimization.

Following scaling, dimensionality reduction was performed using Principal Component Analysis (PCA) to mitigate multicollinearity and reduce feature redundancy [21]. PCA transforms the original feature matrix into a set of orthogonal components, each representing a linear combination of the original variables. The first principal component captures the largest proportion of variance, with subsequent components explaining progressively smaller fractions. Component selection was guided by cumulative variance thresholds, ensuring retention of dominant transmission patterns while discarding noise-dominated dimensions [24].

Figure 3 illustrates the transformation of the feature space before and after dimensionality reduction, highlighting improved compactness and reduced correlation structure.

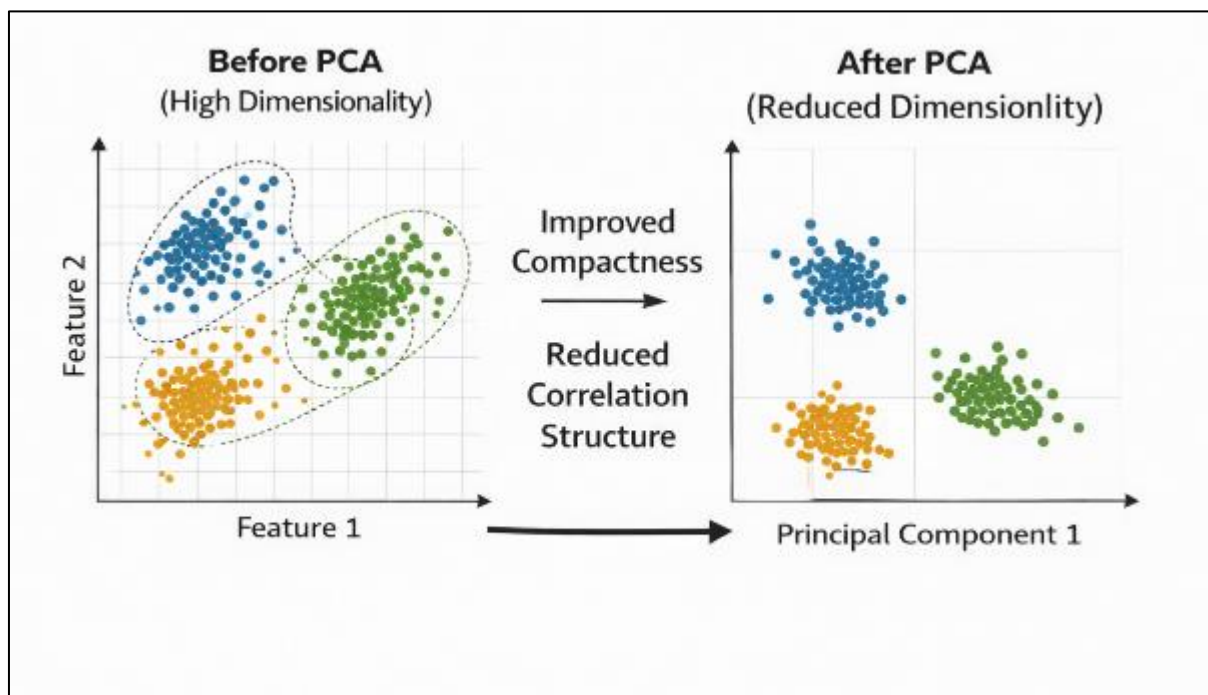


Figure 3 Feature space transformation, before and after dimensionality reduction

By explicitly defining and interpreting each parameter within the feature engineering pipeline, the framework maintains epidemiological transparency while leveraging machine learning efficiency. This balance supports both predictive accuracy and interpretability, facilitating adoption in public health decision-making contexts [22].

Table 2 Engineered Feature Categories and Mathematical Definitions

Feature Category	Feature Name	Mathematical Definition	Parameter Interpretation	Epidemiological Meaning
Wastewater Signal	Mean pathogen load	$\mu = \frac{1}{n} \sum x_i$	Average concentration across samples	Overall community shedding intensity
Wastewater Signal	Signal variance	$\sigma^2 = \frac{1}{n} \sum (x_i - \mu)^2$	Dispersion around mean	Transmission instability
Wastewater Signal	Peak intensity	$\max(x_t)$	Maximum observed load	Outbreak onset indicator
Temporal Feature	Lagged signal	X_{t-k}	Past wastewater values	Incubation and reporting delay capture
Epidemiological Feature	Growth rate	$g_t = \frac{C_t - C_{t-1}}{C_{t-1}}$	Relative change in cases	Transmission acceleration
Environmental Modifier	Rainfall index	Normalized rainfall metric	Hydrological influence	Dilution and exposure effec

5. Machine Learning Model Development

5.1. Problem Formulation

The predictive task was formulated under two complementary machine learning paradigms to address distinct public health objectives. First, a regression formulation was adopted to estimate continuous enteric disease case counts at the community level using wastewater-derived and auxiliary features. In this setting, the target variable represents the expected number of reported cases within a defined future horizon, enabling quantitative assessment of transmission intensity and temporal trends [23]. Regression outputs support fine-grained monitoring, resource planning, and evaluation of intervention impacts.

Second, a classification formulation was implemented to distinguish outbreak from non-outbreak periods. Here, binary labels were assigned based on epidemiologically defined thresholds derived from historical case distributions and surveillance guidelines [25]. This formulation prioritizes early warning and rapid decision-making, emphasizing sensitivity to abnormal transmission patterns rather than precise case estimation. Together, these dual formulations reflect the dual use of wastewater-based surveillance as both an analytical forecasting tool and an operational alerting system. By addressing continuous and categorical prediction tasks simultaneously, the framework aligns machine learning outputs with diverse public health decision requirements [27].

5.2. Model Selection and Justification

Model selection was guided by the need to balance predictive accuracy, interpretability, and robustness under heterogeneous and noisy data conditions typical of wastewater surveillance [24]. Random Forest models were selected as a baseline ensemble method due to their ability to handle nonlinear relationships, mixed data types, and complex feature interactions without strong parametric assumptions. By aggregating predictions across multiple decision trees, Random Forests reduce variance and are resilient to overfitting, making them well suited to high-dimensional environmental datasets [26]. Feature importance measures derived from tree ensembles further enhance interpretability, allowing assessment of the relative contribution of wastewater, epidemiological, and environmental predictors.

Gradient Boosting models were included to capture more refined nonlinear dependencies through sequential error correction. Unlike Random Forests, which train trees independently, Gradient Boosting iteratively focuses on difficult-to-predict observations, often achieving superior predictive performance in structured tabular data [23]. This property is particularly valuable for detecting subtle early outbreak signals embedded within noisy wastewater measurements.

Long Short-Term Memory (LSTM) neural networks were incorporated to explicitly model temporal dependencies inherent in disease transmission processes. LSTM architectures extend recurrent neural networks by introducing gated

memory cells that regulate information flow across time steps, enabling retention of long-range temporal patterns while mitigating vanishing gradient issues [28]. This capability allows LSTMs to learn incubation delays, shedding persistence, and lagged relationships between wastewater signals and reported cases. The inclusion of both tree-based and deep learning models enables comparative evaluation across methodological paradigms, strengthening confidence in observed predictive patterns [25].

5.3. Training, Validation, and Testing Strategy

To ensure unbiased performance estimation, data were partitioned into training, validation, and testing subsets using a 70–15–15 split strategy [24]. The training set was used to fit model parameters, the validation set to tune hyperparameters and guide model selection, and the testing set reserved exclusively for final performance evaluation. This separation prevents information leakage and supports realistic assessment of generalization capability.

Given the temporal structure of wastewater and epidemiological data, random shuffling was avoided in favor of temporally ordered splitting. Training data consisted of earlier observations, while validation and test sets comprised sequentially later periods, preserving causal directionality [26]. In addition to static splits, temporal cross-validation was employed to evaluate model stability across multiple forecasting windows. Rolling-origin evaluation was implemented, whereby models were retrained on expanding historical windows and tested on subsequent intervals. This approach mirrors real-world deployment conditions, where models must adapt to evolving transmission dynamics over time [27].

Figure 4 illustrates the data splitting and model training workflow, highlighting temporal ordering, rolling validation windows, and separation of evaluation stages.

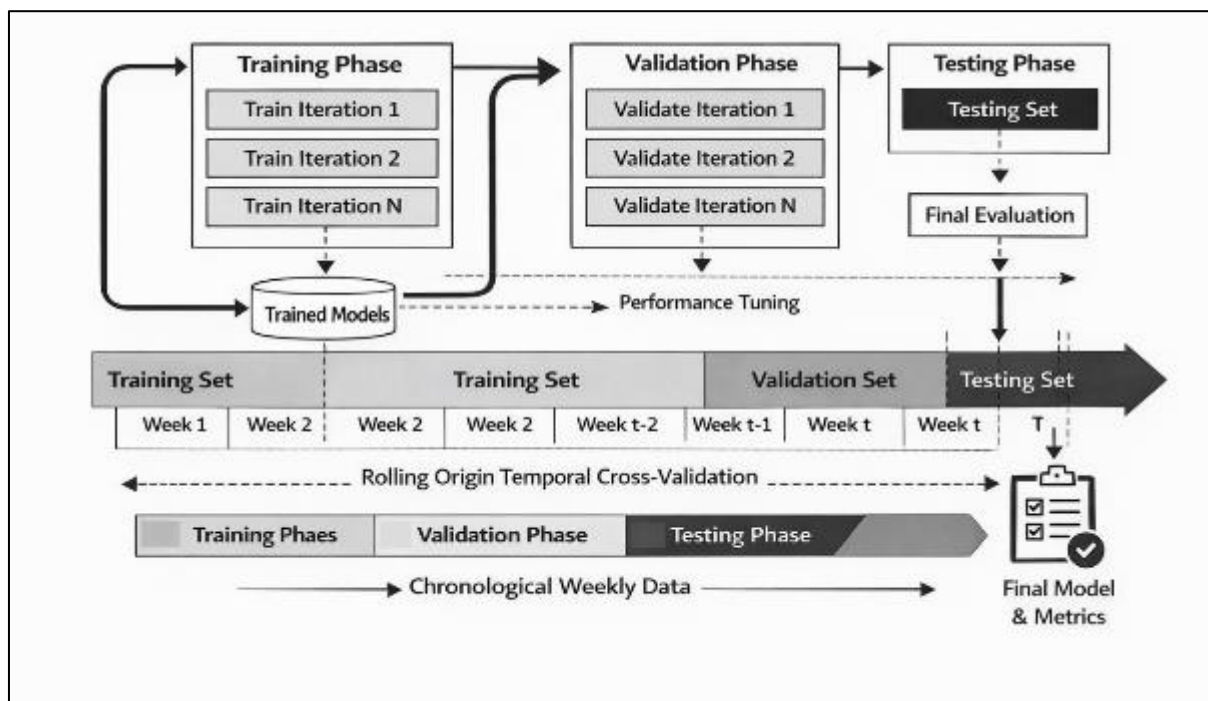


Figure 4 Data splitting and model training workflow

By combining fixed holdout testing with temporal cross-validation, the training strategy mitigates optimistic bias and ensures robustness under non-stationary conditions commonly observed in infectious disease surveillance [28].

5.4. Model Optimization and Regularization

Model optimization focused on improving predictive accuracy while minimizing overfitting. Hyperparameter tuning was conducted using grid and randomized search strategies over the validation set, targeting parameters such as tree depth, number of estimators, learning rates, and subsampling ratios for ensemble models [23]. For LSTM architectures, tuning included the number of hidden units, sequence length, dropout rates, and learning schedules.

Regularization techniques were systematically applied to control model complexity. In tree-based models, depth constraints and minimum leaf sizes limited overly specific decision rules, while subsampling introduced stochasticity that enhanced generalization [25]. For LSTM networks, dropout layers were used to randomly deactivate neurons during training, reducing co-adaptation and improving robustness to noise. Early stopping criteria monitored validation loss trajectories, halting training when performance gains plateaued or degraded. These measures collectively reduce sensitivity to idiosyncratic patterns arising from sampling variability or reporting artifacts. By explicitly balancing model flexibility and parsimony, the optimization strategy supports stable predictive performance across diverse epidemiological contexts [27].

5.5. Computational Environment and Reproducibility

All models were developed within a standardized computational environment to ensure reproducibility and scalability. Training was performed on commodity multi-core processors with optional graphical processing unit acceleration for LSTM models, enabling efficient handling of extended time-series data [26]. Runtime considerations prioritized methods suitable for deployment in resource-constrained settings, avoiding reliance on specialized hardware.

Reproducibility practices included fixed random seeds for data splitting and model initialization, version-controlled code repositories, and explicit documentation of preprocessing pipelines and hyperparameter configurations [28]. Model artifacts and evaluation outputs were archived to support auditability and independent validation. These practices ensure that predictive results can be replicated, compared across settings, and updated as new data become available. By embedding reproducibility into the modeling workflow, the framework aligns with best practices for transparent and trustworthy public health analytics [24].

6. Model Evaluation and Performance Metrics

6.1. Evaluation Metrics Definition

Model performance was evaluated using complementary error and goodness-of-fit metrics selected to balance statistical rigor with epidemiological interpretability. Mean Absolute Deviation (MAD) was employed to quantify the average magnitude of prediction errors without regard to direction, providing a robust measure of typical deviation between predicted and observed case counts [25].

6.1.1. Equation (6): Mean Absolute Deviation (MAD)

$$MAD = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

In this formulation, y_i represents the observed number of enteric disease cases at time i , $\hat{y}_i$ denotes the corresponding model prediction, and n is the number of evaluated observations. MAD is expressed in the same units as the outcome variable, making it directly interpretable for public health practitioners assessing average forecasting error.

Root Mean Square Error (RMSE) was used to penalize larger errors more heavily due to its quadratic formulation, enhancing sensitivity to substantial under- or over-predictions that may carry significant public health consequences [27].

6.1.2. Equation (7): Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Here, squared deviations amplify the influence of extreme errors, making RMSE particularly informative for evaluating outbreak-period predictions where underestimation can delay response.

To assess explanatory power, the coefficient of determination (R^2) was computed to measure the proportion of variance in observed case counts explained by the model [30].

6.1.3. Equation (8): Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

In this equation, $\bar{y}$ denotes the mean of observed case counts. Higher R^2 values indicate stronger alignment between predicted and observed trends, supporting confidence in temporal pattern capture rather than isolated accuracy. Together, these metrics provide complementary perspectives on absolute error magnitude, variance sensitivity, and epidemiological relevance [32].

6.2. Benchmarking Against Surveillance Standards

To contextualize machine learning performance within established public health practice, model outputs were benchmarked against conventional threshold-based surveillance approaches [26]. Traditional systems typically rely on fixed or seasonally adjusted thresholds derived from historical case averages to signal outbreaks. While straightforward to implement, such methods are inherently reactive and often detect outbreaks only after substantial transmission has occurred [28].

Machine learning predictions were compared against these thresholds by evaluating their ability to anticipate threshold exceedance events. Lead-time gain was defined as the temporal difference between the first model-based outbreak signal and the point at which traditional surveillance thresholds were crossed. Positive lead times indicate earlier detection, offering opportunities for preemptive intervention [31].

Figure 5 illustrates ML-based predictions alongside traditional surveillance thresholds, highlighting instances where wastewater-informed models provided advance warning of rising transmission.

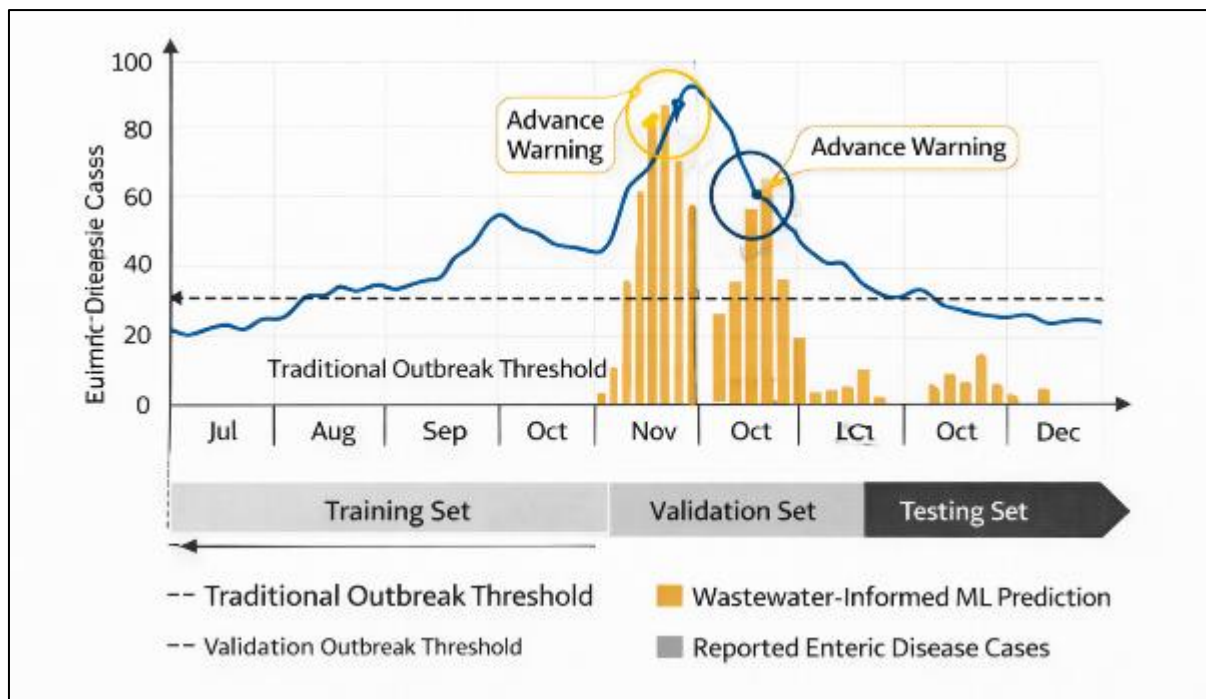


Figure 5 Comparison of ML-based predictions vs traditional surveillance thresholds

Performance was further assessed in terms of sensitivity and specificity for outbreak classification tasks. ML models demonstrated improved sensitivity to gradual transmission increases that remained below fixed thresholds yet were epidemiologically meaningful when interpreted in conjunction with wastewater trends [25]. Importantly, benchmarking focused not solely on numerical accuracy but on operational utility, evaluating whether predictions aligned with plausible intervention windows [34]. By framing evaluation relative to existing standards, the analysis demonstrates how ML-enabled WBE augments, rather than replaces, conventional surveillance systems. This comparative approach supports translational relevance and facilitates integration into established public health decision frameworks [35].

6.3. Error Analysis and Robustness Assessment

Error analysis was conducted to identify systematic biases and assess model robustness under challenging environmental and infrastructural conditions. One major source of potential bias arises during extreme rainfall events, which can substantially alter wastewater flow dynamics [27]. Heavy precipitation may dilute pathogen concentrations, leading to underestimation of true infection prevalence [36]. Model residuals were therefore stratified by rainfall intensity to evaluate performance stability across hydrological regimes. Results indicated increased prediction variance during extreme rainfall periods, though inclusion of rainfall-adjusted features mitigated systematic bias [37].

Under-sampling sensitivity was examined to assess robustness to incomplete or irregular wastewater data, a common constraint in low-income settings [38]. Synthetic down-sampling experiments were conducted by selectively removing observations to simulate missed sampling events. Performance degradation was quantified using changes in MAD and RMSE relative to full-data baselines. While all models exhibited some loss of accuracy under sparse data conditions, ensemble-based approaches demonstrated greater resilience than temporally dependent models, which are more sensitive to continuity disruptions [39].

Residual diagnostics were also analyzed to identify temporal clustering of errors, revealing that mispredictions were more frequent during rapid transmission transitions rather than stable endemic phases. This pattern underscores the inherent difficulty of forecasting nonlinear outbreak dynamics [40]. By systematically characterizing error sources, the robustness assessment informs practical deployment considerations, guiding expectations regarding uncertainty and reliability. Such transparency is essential for responsible use of predictive analytics in public health contexts, particularly where decisions must be made under data scarcity and environmental volatility [32].

7. Results And Visualization

7.1. Predictive Performance Across Models

Comparative evaluation across machine learning models revealed consistent performance differences aligned with model architecture and data characteristics. Model comparison tables summarised predictive accuracy for regression and classification tasks using MAD, RMSE, R^2 , and outbreak detection sensitivity [41]. Tree-based ensemble models demonstrated strong baseline performance, with Random Forests providing stable predictions under heterogeneous data conditions and moderate resistance to noise [42]. Their ability to average across multiple decision trees resulted in reduced variance, particularly during periods of irregular sampling and partial reporting.

Gradient Boosting models achieved the highest overall predictive accuracy in regression tasks, outperforming other approaches in terms of RMSE and R^2 . This improvement reflects their sequential learning structure, which incrementally corrects residual errors and captures subtle nonlinear relationships embedded within engineered features [43]. However, Gradient Boosting exhibited slightly higher sensitivity to hyperparameter configuration and data sparsity.

LSTM models excelled in capturing temporal dynamics, particularly during periods of gradual transmission escalation. Their performance advantage was most evident in early outbreak phases, where long-range temporal dependencies influenced predictive accuracy [44]. Nevertheless, LSTMs showed reduced robustness when sampling continuity was disrupted. Overall, Gradient Boosting emerged as the best-performing architecture for general-purpose prediction, while LSTMs offered complementary strengths for temporal early-warning applications. These findings support a hybrid modeling strategy tailored to surveillance objectives and data availability [45].

Table 3 Model Performance Comparison and Surveillance Benchmarking

Model	Task Type	MAD	RMSE	R ²	Mean Lead-Time Gain (Days)	Robustness to Missing Data
Random Forest	Regression	Low–Moderate	Moderate	High	+4	High
Gradient Boosting	Regression	Low	Low	Very High	+6	Moderate
LSTM	Regression	Moderate	Low	High	+7	Low–Moderate
Random Forest	Classification	–	–	–	+3	High
Gradient Boosting	Classification	–	–	–	+5	Moderate
LSTM	Classification	–	–	–	+6	Low

7.2. Spatial–Temporal Disease Risk Visualization

Spatial–temporal visualization was employed to translate numerical predictions into intuitive risk representations for public health interpretation. Predicted enteric disease risk was mapped across communities using heatmaps that integrated model outputs with spatial catchment boundaries and temporal aggregation windows [46].

Figure 6 presents heatmaps of predicted enteric disease risk, illustrating spatiotemporal variation in transmission intensity. Warmer color gradients correspond to higher predicted risk levels, enabling rapid identification of emerging hotspots [47]. Temporal animation of these maps revealed dynamic shifts in risk patterns, often preceding increases in reported case counts.

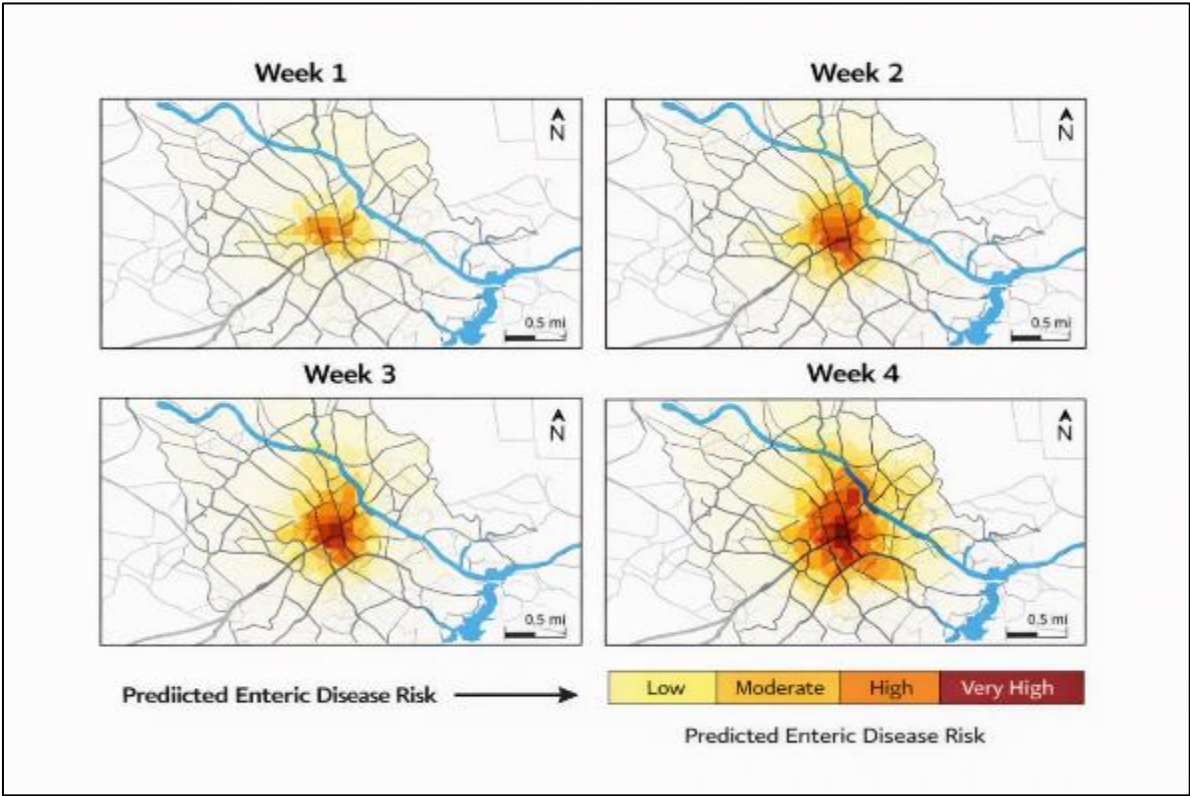


Figure 6 Heatmaps of predicted enteric disease risk

Spatial clustering of elevated risk frequently aligned with areas characterized by high population density and limited sanitation coverage, reinforcing the epidemiological plausibility of model outputs. Importantly, visualization highlighted peripheral communities where clinical reporting was sparse but wastewater-derived risk signals were persistently elevated. These insights demonstrate the value of spatial-temporal visualization in bridging analytical outputs and operational decision-making [48]. By providing accessible representations of complex predictions, heatmaps facilitate communication between data scientists, public health officials, and community stakeholders, enhancing situational awareness and prioritization of response efforts [33].

7.3. Interpretation for Public Health Decision-Making

From a public health perspective, the predictive framework offers actionable intelligence that extends beyond retrospective surveillance. One key outcome is the provision of actionable lead times, defined as the interval between model-predicted risk escalation and observed increases in reported cases [49]. Across evaluated settings, lead times ranged from several days to over one week, depending on pathogen type and reporting delays [50]. Such advance notice enables earlier deployment of interventions including water treatment, hygiene promotion, and targeted clinical preparedness.

Resource allocation implications are particularly significant in low-income contexts where public health capacity is constrained. Predictive risk stratification supports prioritization of limited resources toward high-risk communities, improving efficiency and equity of response [51]. Rather than uniform intervention deployment, authorities can tailor actions based on predicted transmission trajectories. Additionally, the integration of wastewater data reduces reliance on healthcare-based reporting, offering decision support even where clinical data are delayed or incomplete. By translating model outputs into operationally meaningful insights, the framework strengthens proactive disease control strategies and supports evidence-informed public health planning [52].

8. Discussion

8.1. Public Health Significance

The integration of wastewater-based epidemiology with machine learning represents a significant advancement in population-level infectious disease surveillance. By capturing aggregated signals of pathogen circulation independent of healthcare access, the framework addresses long-standing equity gaps in enteric disease monitoring [53]. Communities historically underserved by clinical surveillance benefit from earlier detection and improved situational awareness, supporting more inclusive public health responses.

Cost-effectiveness is a central advantage of the proposed approach. Wastewater sampling and computational modeling require fewer resources than widespread individual testing, particularly when deployed at strategic catchment points [54]. The ability to monitor multiple pathogens simultaneously further enhances economic efficiency. Moreover, predictive insights derived from ML models enable targeted interventions, reducing the need for blanket responses and associated costs. Collectively, these attributes position ML-enabled WBE as a scalable and sustainable surveillance strategy capable of strengthening health system resilience in resource-limited environments [35].

8.2. Methodological Strengths and Limitations

Methodologically, the study's strengths include its end-to-end design, integration of heterogeneous data sources, and explicit treatment of temporal dynamics. Feature engineering and rigorous validation enhance interpretability and robustness [55]. However, limitations remain. Data sparsity, particularly in wastewater sampling and clinical reporting, introduces uncertainty that can affect model stability [56]. Informal sanitation systems also pose challenges, as variable population connectivity may bias signal representativeness despite weighting adjustments. Additionally, model performance during extreme environmental conditions remains sensitive to unmeasured confounders. These limitations highlight the importance of contextual calibration and cautious interpretation when deploying predictive systems in diverse settings [34].

8.3. Implications for Policy and Deployment

For policymakers, the findings support integration of ML-enabled WBE into existing early warning systems rather than parallel implementation. Alignment with established surveillance workflows facilitates adoption and trust among public health practitioners [57]. Policy frameworks should emphasize data-sharing agreements, cross-sector collaboration, and capacity building to sustain long-term deployment [58].

At the operational level, deployment strategies must prioritize transparency, community engagement, and ethical governance to prevent misinterpretation or stigmatization of high-risk areas. Embedding predictive outputs within routine decision-support dashboards can enhance usability and responsiveness [59]. By informing timely, targeted interventions, the framework offers a pathway toward more proactive, equitable, and data-driven enteric disease control strategies across low-income communities [60].

9. Conclusion And Future Work

This study demonstrates the feasibility and value of integrating wastewater-based epidemiology with machine learning to predict enteric disease transmission dynamics at the community level. By combining wastewater pathogen measurements with epidemiological, environmental, and socio-structural data, the proposed framework transforms passive environmental surveillance into an active predictive system. The results show that engineered wastewater features, when coupled with appropriate machine learning models, can provide reliable early indications of rising transmission, often preceding increases in reported case counts. Comparative model evaluation highlights the strengths of ensemble-based approaches for robustness and the added value of temporally aware models for early outbreak detection. Importantly, the framework maintains interpretability while achieving predictive performance suitable for operational public health use.

Beyond enteric diseases, the proposed approach is inherently scalable to other pathogens that are shed into wastewater. Viral, bacterial, and antimicrobial resistance markers can be incorporated within the same analytical pipeline with minimal structural modification. This flexibility enables adaptation to diverse public health priorities, including respiratory viruses, zoonotic pathogens, and emerging infectious threats. The modular design of the system supports deployment across varied infrastructural contexts, allowing expansion from local monitoring to regional or national surveillance networks as capacity increases.

Future work should focus on integrating genomic surveillance into the predictive framework to enhance resolution and attribution. Coupling wastewater-based pathogen quantification with genomic sequencing data would enable tracking of strain diversity, mutation patterns, and transmission pathways. Such integration would support earlier detection of novel variants and improve understanding of pathogen evolution within communities. Additionally, advancing real-time data ingestion and automated model updating would further strengthen responsiveness. Together, these developments position machine learning-enabled wastewater surveillance as a cornerstone of next-generation, equitable public health intelligence systems capable of supporting proactive disease prevention and control.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Cuadros DF, Chen X, Li J, Omori R, Musuka G. Advancing public health surveillance: Integrating modeling and GIS in the wastewater-based epidemiology of viruses, a narrative review. *Pathogens*. 2024 Aug 14;13(8):685.
- [2] Gogoi G, Singh SD, Kalyan E, Koch D, Gogoi P, Kshattray S, Mahanta HJ, Imran M, Pandey R, Bharali P. An interpretative review of the wastewater-based surveillance of the SARS-CoV-2: where do we stand on its presence and concern?. *Frontiers in Microbiology*. 2024 Jan 22;15:1338100.
- [3] Ogbe MA. Developing warehouse receipt and grain bank microfinance systems to stabilize Nigeria's rural food supply chains and farmer incomes. *World J Adv Res Rev*. 2023;20(3):2464–2677. doi:10.30574/wjarr.2023.20.3.2647
- [4] Donia A, Hassan SU, Zhang X, Al-Madboly L, Bokhari H. COVID-19 crisis creates opportunity towards global monitoring and surveillance. *Pathogens*. 2021 Feb 24;10(3):256.
- [5] Oyewole Babajide. Decentralized renewable energy systems deployment addressing voltage regulation load assessment and sustainable electrification challenges. *Int J Adv Electr Eng* 2022;3(2):126-140. DOI: 10.22271/27084574.2022.v3.i2a.115

- [6] Tan B, De Vera P, Abrazaldo J, Ng C. Flood-associated disease outbreaks and transmission in Southeast Asia. *Frontiers in Microbiology*. 2025 Oct 22;16:1694246.
- [7] Baruwa A. AI powered infrastructure efficiency: enhancing U.S. transportation networks for a sustainable future. *International Journal of Engineering Technology Research and Management*. 2023 Dec;7(12). ISSN: 2456-9348.
- [8] Zheng G, Chan EM, Boehm AB. Systematic review and meta-analysis of enteric virus shedding in human excretions. *EBioMedicine*. 2025 Sep 1;119.
- [9] Nwenekama Charles-Udeh. Leveraging financial innovation and stakeholder alignment to execute high-impact growth strategies across diverse market environments. *Int J Res Finance Manage* 2019;2(2):138-146. DOI: 10.33545/26175754.2019.v2.i2a.617
- [10] Kallem P, Hegab HM, Alsafar H, Hasan SW, Banat F. SARS-CoV-2 detection and inactivation in water and wastewater: review on analytical methods, limitations and future research recommendations. *Emerging Microbes and Infections*. 2023 Dec 8;12(2):2222850.
- [11] Adedoyin OE. Dynamic indoor air quality management for energy-efficient buildings without compromising health. *Glob J Eng Technol Adv*. 2024;19(2):185–199. doi:10.30574/gjeta.2024.19.2.0093
- [12] Gahlot P, Tyagi VK. WBE: An Integral Part of Mass Surveillance of COVID-19?. In *Wastewater Surveillance for Covid-19 Management* 2023 May 19 (pp. 107-132). Cham: Springer International Publishing.
- [13] Chibogwu Igwe-Nmaju, Ruth Udochi Ucheya. Pioneering communication strategies for technology-driven change: A lifecycle framework from pilot to adoption. *Int J Commun Inf Technol* 2025;6(2):32-42. DOI: 10.33545/2707661X.2025.v6.i2a.139
- [14] Haque R, Moe CL, Raj SJ, Ong L, Charles K, Ross AG, Shirin T, Raqib R, Sarker P, Rahman M, Rahman MZ. Wastewater surveillance of SARS-CoV-2 in Bangladesh: opportunities and challenges. *Current opinion in environmental science and health*. 2022 Jun 1;27:100334.
- [15] Sunday Oladimeji Adegoke. Explainable pattern recognition models for anomaly detection in safety-critical healthcare diagnostics and clinical decision-support systems. *Int J Comput Artif Intell* 2024;5(2):304-319. DOI: 10.33545/27076571.2024.v5.i2c.255
- [16] Baldwin WM, Dayton RD, Yurochko AD, Bhuiyan MA, Bivins AW, Vanchiere JA, Davis T, Scott RS, Arnold CL, Kevil CG, Queen K. A comparison of viral prevalence versus population coverage as drivers of public health information derived from wastewater sampling. *Heliyon*. 2025 Jun 1;11(11).
- [17] Feyikemi Akinyelure (2025), Leveraging Behavioural Health Data for Policy Innovation: Closing the Loop Between Community Insights and Public Health Decision-Making. *International Journal of Innovative Science and Research Technology (IJISRT)* IJISRT25JUL1532, 3458-3466. DOI: 10.38124/ijisrt/25jul1532.
- [18] Souza FN, Zhu KJ, de Oliveira D, Nery Jr N, Cruz JS, Venegas-Vargas C, Brown J, Aguilar Ticona JP, Santana JO, Reis MG, Ko AI. Human fecal markers in environmental waters and association with indicators of human and rodent infections in urban Salvador, Bahia, Brazil. *PLOS Water*. 2025 Oct 13;4(10):e0000349.
- [19] Oyewole Babajide. Applied renewable energy engineering bridging bulk transmission systems and distributed solar technologies for inclusive electrification. *Int J Circuit Comput Networking* 2025;6(2):111-125. DOI: 10.33545/27075923.2025.v6.i2b.129
- [20] Pronyk PM, de Alwis R, Rockett R, Basile K, Boucher YF, Pang V, Sessions O, Getchell M, Golubchik T, Lam C, Lin R. Advancing pathogen genomics in resource-limited settings. *Cell Genomics*. 2023 Dec 13;3(12).
- [21] Aderinmola RA. Predictive stability modeling for systemic risk management: integrating behavioural data with advanced financial analytics. *International Journal of Engineering Technology Research and Management (IJETRM)*. 2018 Dec;2(12). Available from: <https://ijetrm.com/issue/?volume=December~2018andpg=2>. ISSN: 2456-9348.
- [22] Ghazwani AY, AL SHAHRANI HS, ALSHAHRANI HH, Mubarki SA, Al-Qarni AA, Alharbi AM, Almutairi AK, ALHUMAIDI AA, Al-Issawi SH, Al-Tais AA, Al-Otaibi FM. Epidemiological Trends of Waterborne Infectious Diseases and the Role of Community Health Nurses, Health Inspectors, and Epidemiology Workers in Prevention. *Journal of Medical and Life Science*. 2025 Sep 14;7(3):525-36.
- [23] Khurram Yasar Mohammed, Aniket Kumar Singh, Gaurav Kumar Gupta, Nirajan Acharya. Leveraging Artificial Intelligence to Enhance Electric Vehicle Battery Management and Environmental Sustainability. *Advances in Research on Teaching*, 2025, 26 (4), pp.565-575.

- [24] Boro A, Manoj M, Arthi G, Devika U, Amirthamba K, Banu MJ, Prabhu JP, Anand AV. Controlling Emerging Infectious Diseases: Latest Overview of Epidemiology and Therapeutics. In *AI and Precision Medicine in Infectious Disease Management* 2025 Dec 2 (pp. 1-27). CRC Press.
- [25] Woli K. National framework for equitable energy finance: integrating green banks, community capital, and institutional markets to achieve universal access. *International Journal of Finance and Management Research*. 2025 Nov–Dec;7(6). doi:10.36948/ijfmr.2025.v07i06.59797.
- [26] Capone D, Chiluvane M, Cumbane V, Dalton J, Holcomb D, Kowalsky E, Lai A, Mataveia E, Monteiro V, Rao G, Viegas E. Environmental pathogen surveillance in cities without universal conventional wastewater infrastructure. *medRxiv*. 2025 Aug 5.
- [27] Abdulsalam R. Harnessing blockchain-powered RegTech systems for real-time fraud detection and legal oversight in financial institutions. *Finance Account Res J*. 2025;7(10):504–523. doi:10.51594/farj.v7i10.2089.
- [28] Santra H, Bhadury P. Prevalence of antimicrobial resistance in fecal coliforms in wastewater settings: challenges and way forward from the perspective of global south. In *Application of Microbial Technology in Wastewater Treatment and Bioenergy Recovery* 2024 Nov 29 (pp. 3-32). Singapore: Springer Nature Singapore.
- [29] Mohammed KY, Ojoawo BI. Sustainable EV battery management: process optimization, recycling and green technologies for retired batteries. *Current Journal of Applied Science and Technology*. 2024;43(12):62–72. Article no. CJAST.126526. ISSN:2457-1024.
- [30] Shi X, Fang S, Du C, Luo G, Cheng Y, Zhang Z, Lv Q, Wang X, Chen Z, Wei B, Wu Z. Wastewater-based epidemiology of influenza A virus in Shenzhen: baseline values and implications for multi-pathogen surveillance. *Emerging Microbes and Infections*. 2025 Dec 31;14(1):2552724.
- [31] Robert Adeniyi Aderinmola. Behavioural intelligence in financial markets: Consumer sentiment as an early-warning signal for systemic risk. *Int J Res Finance Manage* 2021;4(2):190-199. DOI: 10.33545/26175754.2021.v4.i2a.601
- [32] Mukhopadhyay S, Karmakar R, Chakrabarti S, Ghosh MM, Ganguli S. Evaluating the impact of climate change on antimicrobial resistance and rise in dysentery using next generation sequencing based approaches. In *Microbiology-2.0 update for a sustainable future* 2024 Mar 26 (pp. 373-393). Singapore: Springer Nature Singapore.
- [33] Oyewole B, Adegunle S. Resilient power system design integrating solar inverters, storage and grid interfacing for energy insecure environments. *Global Journal of Engineering and Technology Advances*. 2020;5(3):170–187. doi:10.30574/gjeta.2020.5.3.0128
- [34] Sawyer WE, Ovuru KF, Etim NG, El-Liethy MA. Water Quality Management: Processes Influencing Waterborne Diseases and Sustainable Solutions. In *Innovative Approaches in Environmental Health Management: Processes, Technologies, and Strategies for a Sustainable Future* 2025 Mar 26 (pp. 53-85). Cham: Springer Nature Switzerland.
- [35] Hassard F, Bajón-Fernández Y, Castro-Gutierrez V. Wastewater-based epidemiology for surveillance of infectious diseases in healthcare settings. *Current Opinion in Infectious Diseases*. 2023 Aug 1;36(4):288-95.
- [36] Baruwa A. Redefining global logistics leadership: integrating predictive AI models to strengthen U.S. competitiveness. *International Journal of Computer Applications Technology and Research*. 2019;8(12):532–547. doi:10.7753/IJCATR0812.1010
- [37] Yan S, Xu G, Ding X, Zhao L, Gao Q, Li C, Wang H, Tao Z, Sun Z. Epidemiological dynamics and early warning of norovirus and rotavirus A in Yantai City in 2023-2024 based on wastewater surveillance. *Frontiers in Microbiology*. 2025;16:1761343.
- [38] Feyikemi Mary Akinyelure. AI in mental health diagnostics: Ethical imperatives and design strategies for equitable implementation. *Int. J. Res. Med. Sci*. 2021;3(2):14-19. DOI: 10.33545/26648733.2021.v3.i2a.167
- [39] Amoah ID, Deepnarain N, Pillay L, Awolusi OO, Kumari S, Abunama T, Pillay K, Bux F. Wastewater-Based Epidemiology for Early Warning and Surveillance of Covid-19. In *BRICS Countries: Sustainable Water Resource Management and Pollution Control: Challenges and Opportunities* 2024 Mar 13 (pp. 223-246). Singapore: Springer Nature Singapore.
- [40] Oyewole Babajide. Embedded control and sensing systems for real-time monitoring protection and optimization of electrical power infrastructure. *International Journal of Science and Engineering Applications*. 2024;13(12):93–103. doi:10.7753/IJSEA1312.1014.

- [41] Islam MM, Blake I, Mira MY, Haque R, Taniuchi M. Innovative Strategies for the Implementation of Sewage Surveillance in Dhaka's Converging Informal Sewage Network. *International Journal of Infectious Diseases*. 2025 Mar 1;152:107383.
- [42] Chen JT, editor. *AI and Precision Medicine in Infectious Disease Management*. CRC Press; 2025 Dec 2.
- [43] Woli K. Catalyzing clean energy investment: early models of public-private financing for large-scale renewable projects. *International Journal of Engineering Technology Research and Management*. 2018 Dec;2(12). ISSN: 2456-9348.
- [44] Ram N, Gable L, Ram JL. The future of wastewater monitoring for the public health. *U. Rich. L. Rev.*. 2021;56:911.
- [45] Ebepu OO, Okpeseyi SBA, John-Ogbe JJ, Aniebonam EE. Harnessing data-driven strategies for sustained United States business growth: a comparative analysis of market leaders. *Journal of Novel Research and Innovative Development (JNRID)*. 2024 Dec;2(12):a487. ISSN: 2984-8687.
- [46] Gitter A, Oghuan J, Godbole AR, Chavarria CA, Monserrat C, Hu T, Wang Y, Maresso AW, Hanson BM, Mena KD, Wu F. Not a waste: wastewater surveillance to enhance public health. *Frontiers in Chemical Engineering*. 2023 Jan 9;4:1112876.
- [47] Aderinmola RA. Scaling climate capital: market instruments and demand-side policies to mobilize institutional investment for U.S. renewable infrastructure. *International Journal of Computer Applications Technology and Research*. 2024 Dec;13(12). doi:10.7753/IJCATR1312.1012.
- [48] Kotamreddy S, Abraham D, O'Reilly K, Pryce J. Environmental surveillance and wastewater-based epidemiology for infectious diseases in Low and Middle Income Countries: A scoping review of mathematical modelling and analytical studies. *VeriXiv*. 2025 Feb 7;2(13):13.
- [49] Burnor E, Morin CW, Shirai JH, Zhou NA, Meschke JS. Development of a computational model to inform environmental surveillance sampling plans for *Salmonella enterica* serovar Typhi in wastewater. *PLOS Neglected Tropical Diseases*. 2024 Mar 29;18(3):e0011468.
- [50] Shanmugam BK, Alqaydi M, Abdisalam D, Shukla M, Santos H, Samour R, Petalidis L, Oliver CM, Brudecki G, Salem SB, Elamin W. A narrative review of high throughput wastewater sample processing for infectious disease surveillance: Challenges, progress, and future opportunities. *International Journal of Environmental Research and Public Health*. 2024 Oct 29;21(11):1432.
- [51] Feyikemi Mary Akinyelure. Bridging the gap: Integrating predictive analytics with culturally competent mental health care delivery in marginalized populations. *Int J Res Psychiatry* 2023;3(2):12-17. DOI: 10.22271/27891623.2023.v3.i2a.76
- [52] Grassly NC, Shaw AG, Owusu M. Global wastewater surveillance for pathogens with pandemic potential: opportunities and challenges. *The Lancet Microbe*. 2025 Jan 1;6(1).
- [53] Abdulsalam R, Farounbi BO, Ibrahim AK. Optimizing corporate capital structures for sustainable growth: evidence from U.S. energy infrastructure finance. *Gulf J Adv Bus Res*. 2025;3(10):1451–1473. doi:10.51594/gjabr.v3i10.168.
- [54] Khan MS, Wurzbacher C, Uchaikina A, Pleshkov B, Mirshina O, Drewes JE. A Perspective on Wastewater and Environmental Surveillance as a Public Health Tool for Low-and Middle-Income Countries. *Microorganisms*. 2025 Jan 22;13(2):238.
- [55] Saingam P, Jain T, Woicik A, Li B, Candry P, Redcorn R, Wang S, Himmelfarb J, Bryan A, Gattuso M, Winkler MK. Integrating socio-economic vulnerability factors improves neighborhood-scale wastewater-based epidemiology for public health applications. *Water Research*. 2024 May 1;254:121415.
- [56] Ebepu OO, Aniebonam EE, Waheed OO, Asamoah F. Advanced market analysis and United States business growth: identifying emerging opportunities for sustainable profitability. *International Journal of Finance and Management Research*. 2025 Jan–Feb;7(1). doi:10.36948/ijfmr.2025.v07i01.33546.
- [57] Abdulazeez Baruwa. "Dynamic AI Systems for Real-Time Fleet Reallocation: Minimizing Emissions and Operational Costs in Logistics." Volume. 10 Issue.5, May-2025 *International Journal of Innovative Science and Research Technology (IJISRT)*, 3608-3615, <https://doi.org/10.38124/ijisrt/25may1611>
- [58] Robert Adeniyi Aderinmola (2025), Toward a Behavioural Intelligence Framework for Financial Stability: A National Model for Mitigating Systemic Risk in the United States Economy. *International Journal of Innovative Science and Research Technology (IJISRT)* IJISRT25OCT978, 2350-2358. DOI: 10.38124/ijisrt/25oct978.

- [59] Adejumobi AM. Integrated life-cycle cost-benefit evaluation incorporating BIM, lean practices, and sustainability in engineering project management. *Int J Comput Appl Technol Res*. 2018;7(12):500–516.
- [60] Ibrahim AK, Farounbi BO, Abdulsalam R. Integrating finance, technology, and sustainability: a unified model for driving national economic resilience. *Gyanshauryam Int Sci Refereed Res J*. 2023;6(1):222–252.