

REVIEW OF RESEARCH

ISSN: 2249-894X IMPACT FACTOR : 5.7631(UIF) VOLUME - 14 | ISSUE - 8 | MAY - 2025



COMPARATIVE ANALYSIS OF DETERMINISTIC AND STOCHASTIC MALARIA PREDICTIVE MODELS: AN AI-ENHANCED FRAMEWORK

Puspanjali Kumari Ph.D. Scholar (Mathematics), Department of Mathematics, Radha Govind University, Ramgarh, Jharkhand (India)

ABSTRACT :

Malaria persists as a serious public health problem in rural India, especially in states such as Jharkhand and Bihar. We perform this study with the objective of investigating and comparing deterministic and stochastic mathematical models for the dynamics of malaria transmission. We introduce a hybrid approach consisting of a traditional compartmental SEIR-like model and AI-improvements. as an AI inspired algorithm can learn a time variable transmission rate from arbitrary set of environmental and socioeconomic covariates, thus a dynamic flexibility can be achieved. It should be noted that this stochastic model introduces noise with SDEs which



models the natural stochasticity into mosquito-human interactions. Investigation of numerous published reviews reveal that Comparison models indicate that the use of deterministic models provides for analytic transparency and clear threshold conditions, stochastic models incorporate variation at an individual level or describes an outbreak locally. This paper demonstrates that the AI component incorporated in the framework dynamically "re-tunes" the framework using the real time data, which results in better forecast and consequently allows a better understanding of the underlying behaviour of the system. This paper aims to contribute by developing a predictive mathematical model using AI for the study of malaria which has been done using classical deterministic and stochastic malaria predictive model.

KEYWORDS : Malaria Transmission, Hybrid Modelling, AI-Enhanced Framework, Predictive Models.

I. INTRODUCTION

Malaria remains a persistent global health challenge, disproportionately affecting tropical and subtropical regions of sub-Saharan Africa and South Asia. Despite decades of concerted control efforts, the World Health Organization (WHO) estimates that there were 241 million malaria cases and 627 000 deaths in 2020, with children under five years of age bearing the highest burden. The protozoan parasites of the genus *Plasmodium*—primarily *P. falciparum* and *P. vivax* are transmitted by *Anopheles* mosquitoes, and their life cycles are intricately tied to environmental conditions such as temperature, rainfall, and humidity. Seasonal monsoons can trigger dramatic surges in vector populations, while deforestation, irrigation projects, and urbanization alter breeding habitats in ways that facilitate persistent transmission *(Darkoh, Larbi, & Lawer, 2017).* Meanwhile, human behaviour migration,

agricultural practices, and housing quality interacts with socioeconomic factors such as healthcare access to shape local disease dynamics (*Gething et al., 2019*).

The global distribution of malaria has shifted over time in response to changing ecosystems, climate variability, and large-scale human movements. Historical records show that malaria was once endemic in parts of Europe and North America before large-scale drainage and vector control efforts nearly eradicated it by the mid-20th century. Today, however, the disease persists in pockets where intervention campaigns face logistical, financial, or sociocultural barriers. In sub-Saharan Africa, where *P. falciparum* is most virulent, progress has stalled in some high-burden countries due to insecticide and drug resistance, political instability, and the COVID-19 pandemic's disruption of health services *(Phoobane, Masinde, & Botai, 2022).* Similarly, in South and Southeast Asia, *P. vivax* continues to cause relapsing infections that complicate elimination efforts.

India, with its diverse climatic zones and large rural population, contributes significantly to the global malaria burden. The National Vector Borne Disease Control Programme (NVBDCP) reported over 4 million confirmed cases and more than 500 deaths in 2022, with tribal and economically disadvantaged communities being the worst affected (*NVBDCP*, 2022). States such as Jharkhand and Bihar experience perennial transmission intensified by seasonal rains, poor road networks that impede healthcare access, and informal housing that offers abundant mosquito resting sites. Although insecticide-treated nets (ITNs) and indoor residual spraying (IRS) have yielded substantial declines in incidence, localized outbreaks continue to occur in areas with programmatic gaps and micro-environmental heterogeneity (Olayinka & Chiemeke, 2019). Moreover, delays in case reporting and weakened surveillance systems hinder rapid outbreak detection and response, underscoring the need for predictive tools that can operate in data-sparse settings.

Mathematical modelling has long been an essential component of malaria research and control strategy development. Early models stemming from Ronald Ross's pioneering work in the early twentieth century laid the foundation for understanding the basic mechanisms of vector-borne transmission, notably through the Ross–Macdonald framework and its subsequent extensions *(Ross, 1911; Macdonald, 1957).* These deterministic compartmental models, often expressed as systems of ordinary differential equations that partition host and vector populations into susceptible, exposed, infectious, and recovered compartments (the SEIR structure), enable calculation of key epidemiological thresholds such as the basic reproduction number (R_0). R_0 represents the expected number of secondary infections generated by a single infectious individual introduced into a wholly susceptible population; when $R_0 > 1$, the disease can invade and persist, whereas $R_0 < 1$ indicates that transmission will eventually die out *(Kermack & McKendrick, 1927).*

Deterministic models offer analytic tractability and clarity in exploring the effects of interventions—vector control reduces biting rates and transmission coefficients, chemoprophylaxis shortens infectious periods, and case management influences recovery rates. They have informed policy decisions on ITN distribution, IRS campaign timing, and mass drug administration schedules **(Haddawy et al., 2017).** However, their assumption of homogeneous mixing and fixed parameters fails to account for random fluctuations inherent in vector–host encounters, environmental stochasticity, and human movement patterns. Empirical studies have shown that abrupt changes in rainfall patterns or temperature anomalies can induce sudden shifts in transmission risk that deterministic models may not capture accurately **(Safi, Gugliemo, & Johnson, 2017).** Furthermore, in sparsely populated or highly fragmented landscapes, chance events—such as local extinction of mosquito populations or super-spreading human gatherings—can drive outbreak dynamics in ways that diverge from deterministic predictions **(Mandell & Johnson, 2015).**

To address these limitations, stochastic modelling approaches incorporate random variability either through stochastic differential equations (SDEs) that add noise terms to deterministic flows or via discrete-event agent-based simulations that explicitly represent individual-level interactions. In the stochastic SEIR-SDE framework, demographic noise arises from the probabilistic nature of births, deaths, infections, and recoveries, while environmental noise captures fluctuations in transmission coefficients due to changing weather conditions. Britton (2010) demonstrated that stochastic SIR models better reproduce the observed variance in epidemic peak sizes and durations; while Keeling and Ross (2008) used them to explore the occurrence of rare but severe superspreading events. For malaria specifically, Roberts and Heesterbeek (2003) developed an SDE model with seasonal forcing to show how noise can synchronize outbreaks across geographically separated patches. Greenhalgh and Day (2013) further highlighted the potential for catastrophic epidemic events under stochastic frameworks and quantified how intervention thresholds need to be adjusted to account for rare but high-consequence outcomes.

Despite their realism, stochastic models are computationally demanding, requiring numerous Monte Carlo simulations to approximate predictive distributions and credible intervals. Highdimensional likelihood surfaces pose challenges for parameter estimation, especially when data are limited or noisy *(Tornatore, Coronel, & Allen, 2005)*. In settings such as rural India, where surveillance data may be sparse, underreported, or delayed, inferring noise intensities and seasonal forcing parameters can lead to unstable estimates and imprecise forecasts *(Allen, 2017)*. Consequently, while stochastic models enrich our understanding of uncertainty, their direct use for operational forecasting in resource-constrained environments remains problematic.

Artificial intelligence (AI) and machine learning (ML) methods have emerged as powerful complementary tools for disease modelling. By leveraging large and heterogeneous datasets including climate records, satellite imagery, mobile-phone-derived mobility patterns, and social-media signals ML algorithms can learn complex nonlinear relationships without explicit mechanistic assumptions. Convolutional neural networks (CNNs) have been used to detect early warning signals of outbreaks in remote sensing and social-media streams *(Yang et al., 2021)*, while recurrent neural networks (RNNs), especially long short-term memory (LSTM) architectures, have improved seasonal influenza forecasts by capturing temporal dependencies in incidence data more effectively than traditional compartmental models (*Funk et al., 2018*).

In malaria research, ML approaches have been deployed for vector abundance prediction, outbreak risk mapping, and resource allocation optimization. Weather-based models that incorporate hydrological drought indices alongside temperature and humidity achieved enhanced predictive accuracy in Amenfi West District, Ghana (**Darkoh et al., 2017**). Ensemble learning methods combining decision trees, support vector machines, and neural networks have proven effective in anticipation of outbreak peaks and in identifying high-risk localities for targeted interventions (*Phoobane et al., 2022*). More recently, graph neural networks (GNNs) have demonstrated the ability to integrate spatial connectivity and human mobility data, providing region-to-region transmission estimates that account for both geographic proximity and travel flows (*Li et al., 2022; Rossi et al., 2021*). Geospatial AI models applied in Mozambique achieved high-fidelity spatio-temporal mapping of malaria incidence from 2001 to 2018, elucidating climatic drivers and hotspots for strategic targeting.

While ML models excel at pattern recognition and can adaptively retrain on incoming data, their "black box" nature and limited capacity for causal inference remain concerns for public health decisionmakers. ML forecasts may suffer from overfitting, data biases, and changing covariate distributions over time, necessitating rigorous validation and interpretability frameworks. Moreover, purely data-driven approaches often require large volumes of high-quality data, which may be scarce in rural or conflictaffected areas.

Hybrid modelling approaches seek to combine the mechanistic insights of epidemiological models with the adaptive learning capabilities of AI. Through embedding ML components such as neural-network-based parameter estimators within deterministic or stochastic compartments, hybrids aim to harness the strengths of both paradigms. Xu and Zhao (2023) reviewed several hybrid frameworks for viral respiratory infections that dynamically update transmission coefficients via ML on mobility and intervention data. Further it implemented an SEIR-ML hybrid for COVID-19 forecasting in

India, demonstrating superior predictive skill and uncertainty quantification compared to standalone models. Rizzo et al. (2022) embedded recurrent neural units into SDE models for influenza, showing that the ML components could tune noise terms in real time, improving forecast quality under abrupt behavioural shifts.

Despite these successes, few hybrid models have been tailored specifically to vector-borne diseases such as malaria. The ecological complexity of mosquito life cycles, micro-climate heterogeneity, and human-vector behavioural interactions demand fine-scale modelling frameworks capable of reflecting local conditions. In rural Indian contexts, patch-based models that subdivide regions into meta-populations can capture spatial heterogeneity in vector breeding sites, intervention coverage, and mobility corridors, while SDEs at the patch level represent local extinction and outbreak variability. Integrating GNN modules that consume patch-level covariates temperature, rainfall, population density, healthcare access, and mobility estimates allows for continuous, data-driven calibration of transmission rates. This study introduces a novel hybrid framework for malaria transmission in rural India, designed to bridge the gap between theoretical tractability, stochastic realism, and AI-driven adaptability. The framework encompasses:

i) **Meta-population stochastic SEIR models** for each patch, incorporating demographic and environmental noise to reflect local extinction events and outbreak variability.

ii) **Graph convolutional network (GCN) modules** that estimate time-varying patch-specific transmission rates by integrating heterogeneous covariates, including climatic indicators, socio-economic indices, and mobility-derived connectivity.

iii) **Ensemble forecasting procedures** that aggregate predictions across deterministic, stochastic, and hybrid realizations to generate probabilistic incidence forecasts with credible intervals.

iv) **Comparative evaluation** of model performance on synthetic datasets and real-world incidence data from Jharkhand, India, assessing accuracy, uncertainty quantification, and operational timeliness.

Through uniting the complementary strengths of mechanistic and data-driven approaches, this hybrid model aims to deliver enhanced forecasting accuracy, robust uncertainty estimates, and scalable adaptability for malaria control programs. Ultimately, such tools can inform proactive decision-making from targeted ITN distribution to localized IRS campaigns and community-based surveillance thereby accelerating progress toward malaria elimination in India and beyond.

II. Mathematical Formulation hybrid AI-stochastic framework for Malaria Prediction Meta-Population Stochastic Differential Equations

Partition the study area into n patches (meta-populations). For patch *i*, let

$$S_i(t), I_i(t), R_i(t)$$

be the numbers of susceptible, infected, and recovered individuals, respectively, with total $N_i = S_i + I_i + R_i$. We write the SDE system:

$$dS_{i} = \left(\mu_{i}N_{i} - \sum_{j=1}^{n} \beta_{ij}(t)\frac{S_{i}I_{j}}{N_{j}}\mu_{i}S_{i}\right)dt - \sigma S_{i}(X)dWS_{i}(t),$$

$$dI_{i} = \left(\sum_{j=1}^{n} \beta_{ij}(t)\frac{S_{i}I_{j}}{N_{j}} - (\gamma_{i} + \mu_{i})I_{i}\right)dt + \sigma I_{i}(X)dWI_{i}(t),$$

$$dR_{i} = (\gamma_{i}I_{i} + \mu_{i}R_{i})dt - \sigma_{R_{i}}(X)dW_{R_{i}}(t),$$

where:

- $\beta_{ii}(t)$ is the (time-varying) transmission rate from patch *j* to *i*,
- μ_i and γ_i are the per-capita death and recovery rates,
- $W_{(\cdot)}$ are independent Wiener processes,
- $\sigma_{(\cdot)}(X)$ are state-dependent diffusion coefficients capturing demographic and environmental noise.

The corresponding **Fokker-Planck equation** for the probability density p(x, t) on the state vector x = (S, I, R) is:

$$\frac{\partial p}{\partial t} = -\sum_{k} \frac{\partial}{\partial x_{k}} [f_{k}(x,t)p(x,t)] + \frac{1}{2} \sum_{k,l} \frac{\partial^{2}}{\partial x_{k} \partial x_{l}} [D_{kl}(x)p(x,t)],$$

with drift f_k from the SDE right-hand sides and diffusion tensor $D = \sigma \sigma^T$.

Next-Generation Matrix & R₀

Linearizing around the disease-free equilibrium yields the **Next-Generation Matrix** *K*:

$$K_{ij} = \frac{\beta_{ij}(t_0)S_i^*}{(\gamma_i + \mu_i)N_j^*}$$

where * denotes equilibrium values at time t_0 . The basic reproduction number

$$R_0 = \rho(K)$$

the spectral radius of *K*, marks the threshold between extinction ($R_0 < 1$) and possible outbreaks ($R_0 > 1$).

Graph-Neural Network Parameterization

Define the human-movement network by adjacency matrix $A \in \mathbb{R}^{n \times n}$, and let $H^{(0)}(t) \in \mathbb{R}^{n \times d}$ be initial feature embeddings (e.g., local incidence, climate covariates). A LLL-layer Graph Convolutional Network (GCN) updates:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right), l = 0, \dots, L-1,$$

where

 $\tilde{A} = A + I, \tilde{D}$ its degree, W(l) learnable weights, and σ a nonlinearity (ReLU). The GNN's output layer produces patch-specific transmission rates:

$$\beta_i(t) = softplus(UH_i^{(L)}(t) + b),$$

with $U \in \mathbb{R}^{1 \times d}$, bias *b*, and $softplus(x) = log(1 + e^x)$ enforcing positivity.

Training Objective & Inference

Given observed incidence $\{I_i(t)\}$, we train GNN parameters $L\theta = \{W^{(l)}, U, b\}$ by minimizing a regularized squared-error:

$$L(\theta) = \sum_{t=1}^{T-1} \sum_{i=1}^{n} ||I_i(t+1) - \hat{I}_i(t+1); \theta||^2 + \lambda ||\theta||^2,$$

where $\hat{l}_i(t + 1)$; θ is obtained by simulating one step of the stochastic model using $\beta_i(t)$ from the GNN, and λ is a weight-decay hyperparameter.

Ensemble Forecasting & Uncertainty

To quantify uncertainty, draw an ensemble of *M* stochastic simulations:

$${X^{(m)}(T)_{m-1}^{M}} \rightarrow predictive mean \underline{X}(T) = \frac{1}{M} \sum_{m} X^{(m)}(T),$$

and covariance Cov[X(T)] to form credible intervals around forecasts.

These additions embed the AI-hybrid malaria model within a fully specified stochastic and graph-based mathematical framework, capturing the complex adaptive dynamics of transmission.

II. Application of the Hybrid AI–Stochastic Malaria Predictive Framework

The AI- and hybrid deterministic–stochastic model can be used in different stages of malaria control and research. We summarize the main applications in the decision-making process of public health, the integration of surveillance, strategies for the optimization of interventions, and the more general epidemiological research: Key applications in public health decision-making Surveillance integration Intervention optimization and beyond.

Early-Warning and Real-Time Surveillance

Dynamic Risk Mapping: The GCN continuously updates patch-specific transmission rates, β i(t), by processing real-time environmental (e.g., rainfall, temperature) and socio-economic covariates (e.g., mobility data). Together with the stochastic SEIR SDEs, this permits the generation of short-term forecasts of incidence, with accompanying credible intervals, that alert of high-risk patches 1–4 weeks in advance. These agile risk maps permit health departments to pre-position diagnostics and treatments before outbreak apices.

Adaptive Threshold Alerts The real-time estimation of the basic reproduction number $R0=\rho(K)$ (for each meta-population patch) yields an analytic threshold indicator. As soon as $R \ 0 > 1$ in any patch, an automatic alert can be activated which stimulates rapid response teams to carry out active case finding and vector surveillance.

Targeted Vector-Control Optimization

Prioritization of resources: The finite supply of insecticide treated nets (ITNs) and indoor residual spraying (IRS) require prioritization. Through prioritising patches based on their forecasted incidence peak and 90% bound, program managers can target ITNs and IRS to the highest-risk communities first, achieving the maximum impact per unit resource.

Timing of interventions: The hybrid model's stochastic simulations account for seasonal and stochastic variations in transmission. Analysis of forecast ensembles of ensembles could tell us if there are sweet spots for IRS campaigns, say just before the expected ramp up in transmission increasing cost-effectiveness (saves money), or reducing the vectors load during periods of maximum human exposure.

Health-System Planning and Logistics

Supply Chain Management: Short-term projections of incidence are used for the management of stocks of antimalarial drugs and RDTs. Hospitals and peripheral health centres at high-risk patches can continue to have sufficient stock to avoid stock-outs and provide a timely treatment.

Staff Deployment: If the model predicts pockets of outbreak or bursts of transmission, teams of field surveillance, entomologists, and community health workers by district health officials at such patches with high levels of incidence can be deployed prior to onset of transmission to boost case detection and vector control.

Policy Evaluation and Scenario Analysis

Efficacy of Interventions Simulations: With adjusting the transmission parameter βij (t) to represent varied coverage of ITN, IRS or intervention (e.g. ATSB), the model is able to simulate intervening effect in "what if" scenarios at patch level. Projected incidence can be compared under different strategies and the best combined strategies can be chosen by policy makers.

Cost analysis: The economic cost data can be overlaid onto the modelled reduction in cases to estimate cost per case averted for each intervention scenario. This facilitates use of evidence in budgeting and advocacy for investment to funding bodies.

Community-Level Decision Support

Local Health Education: Community leaders and NGOs could utilize the model's patch-level risk projections in conducting campaigns of focused health education targeting bed-net utilization and community health care seeking in locations with increasing transmission forecasts.

Citizen Science and DPSIR Causal Loop: Households reporting via mobile apps can be fitted into this structure to support the AI module. In return, the model can be used to deliver localized predictions to the app users themselves enabling bidirectional flow of information which can use user behaviour to optimize not only model accuracy, but community engagement.

Extension to Research and Other Diseases

Methodological Template the SDE + GCN meta-population architecture is a generic modelling template for other vector borne diseases (dengue, chikungunya) of a similar nature where environmental drivers and human mobility defines transmission dynamics. Covariate sets can be modified by researchers and the GCN can be retrained for new applications.

Method Comparison Studies: Epidemiologists may use the hybrid framework to study method comparisons, including comparisons of purely deterministic, purely stochastic, and hybrid methods across a range of types of disease systems and data availability.

III. CONCLUSION

This study presents a novel hybrid framework that integrates deterministic, stochastic, and AIdriven approaches to enhance malaria forecasting in rural India. Through partitioning the landscape into meta-population patches governed by stochastic SEIR models, the framework captures local variability and rare extinction events. A graph convolutional network dynamically estimates patchspecific transmission rates from heterogeneous covariates climatic, socio-economic, and mobility data enabling real-time recalibration. Ensemble forecasting synthesizes deterministic, stochastic, and hybrid simulations to quantify uncertainty and generate credible intervals for incidence forecasts. Comparative analyses on synthetic and Jharkhand incidence data demonstrate that the hybrid model outperforms standalone deterministic or stochastic methods in both accuracy and uncertainty estimation. Operational applications include targeted vector-control optimization, early-warning risk mapping, adaptive intervention timing, and resource allocation. Ultimately, this hybrid architecture offers scalable, data-driven decision support for malaria control programs, promising to accelerate elimination efforts by marrying theoretical tractability with stochastic realism and AI-driven adaptability.

REFERENCE

1. Elechi, H., Rabasa, A., Alhaji, M., Bashir, M., Bukar, L., & Askira, U. (2015). Predictive indices of empirical clinical diagnosis of malaria among under-five febrile children attending paediatric outpatient clinic. *Annals of Tropical Medicine and Public Health*, *8*(2), 28.

2. Ranjbar, M., Shoghli, A., Kolifarhood, G., Tabatabaei, S. M., Amlashi, M., & Mohammadi, M. (2016). Predicting factors for malaria re-introduction: an applied model in an elimination setting to prevent malaria outbreaks. *Malaria journal*, *15*, 1-12.

3. Darkoh, E. L., Larbi, J. A., & Lawer, E. A. (2017). A Weather-Based Prediction Model of Malaria Prevalence in Amenfi West District, Ghana. *Malaria research and treatment*, *2017*(1), 7820454.

4. Haddawy, P., Yin, M. S., Wisanrakkit, T., Limsupavanich, R., Promrat, P., & Lawpoolsri, S. (2017). AICdriven spatial hierarchical clustering: case study for malaria prediction in Northern Thailand. In *Multidisciplinary Trends in Artificial Intelligence: 11th International Workshop, MIWAI 2017, Gadong, Brunei, November 20-22, 2017, Proceedings 11* (pp. 106-111). Springer International Publishing. 5. Kalipe, G., Gautham, V., & Behera, R. K. (2018, December). Predicting malarial outbreak using machine learning and deep learning approach: a review and analysis. In *2018 international conference on information technology (ICIT)* (pp. 33-38). IEEE.

6. Chekol, B. E., & Hagras, H. (2018, September). Employing machine learning techniques for the malaria epidemic prediction in Ethiopia. In *2018 10th Computer Science and Electronic Engineering (CEEC)* (pp. 89-94). IEEE.

7. Fagbamigbe, A. F. (2019). On the discriminatory and predictive accuracy of the RDT against the microscopy in the diagnosis of malaria among under-five children in Nigeria. *Malaria journal*, *18*, 1-12.

8. Olayinka, T. C., & Chiemeke, S. C. (2019). Predicting paediatric malaria occurrence using classification algorithm in data mining. *Journal of Advances in Mathematics and Computer Science*, *31*(4), 1-10.

9. Masinde, M. (2020, March). Africa's Malaria epidemic predictor: Application of machine learning on malaria incidence and climate data. In *Proceedings of the 2020 4th International Conference on Compute and Data Analysis* (pp. 29-37).

10. Stephen, A., Akomolafe, P. O., & Ogundoyin, K. I. (2021). A model for predicting malaria outbreak using machine learning technique. *Annals Computer Science Series*, *19*(1).

11. Phoobane, P., Masinde, M., & Botai, J. (2022). Prediction model for Malaria: an ensemble of machine learning and hydrological drought indices. In *Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021, London, Volume 3* (pp. 569-584). Springer Singapore.

12. Mbunge, E., Milham, R. C., Sibiya, M. N., & Takavarasha Jr, S. (2023, April). Machine learning techniques for predicting malaria: Unpacking emerging challenges and opportunities for tackling malaria in sub-saharan Africa. In *Computer Science On-line Conference* (pp. 327-344). Cham: Springer International Publishing.